Capturing Unobserved Heterogeneity in the Austrian Labor Market Using Finite Mixtures of Markov Chain Models

Sylvia **Frühwirth-Schnatter** and Christoph **Pamminger**

Department of Applied Statistics and Econometrics
Johannes Kepler University Linz, Austria

Collaboration with Rudolf **Winter-Ebmer**,
Department of Economics, Johannes Kepler University Linz

Supported by the **Austrian Science Foundation** (FWF)
under grant P 17 959 ("Gibbs Sampling for Discrete Data")

Outline

Clustering

Motivating Example

- Research Question
- Data Description

Markov Chain Model

Dirichlet Multinomial Model

- Bayesian Analysis
- MCMC-Estimation

Estimation Results

Clustering

Clustering is a widely used statistical tool to determine subsets

Frequently used clustering methods are based on **distance-measures**

However, distance-measures are **difficult** to define for more **complex** data (e.g. time series)

⇒ **Model-based clustering methods** (mixture models)

We present an approach for **model-based clustering of discrete-valued time series data** following ideas discussed in Frühwirth-Schnatter and Kaufmann (2004)

Motivating Example

Wage Mobility in the Austrian labor market

Describes **chances** but also **risks** of an individual **to move** between wage categories

Assumption of **different** career progressions or income careers of employees

Task: Find **groups** of employees with **similar** behavior in terms of transition probabilities (focus on one-year transitions)

Data provided by the Austrian **social security authority**

Data Description

Time series for $N = 9,809$ individuals (only men, because of data inconsistencies with e.g. female part-time workers)

Gross **monthly wage** at May of successive years (with individual length T_i) divided into **6 categories** corresponding to quintiles of the particular income distribution (1-5) and zero-income (0) according to Weber (2002)

$\rightarrow \mathbf{y}_i = (y_{i0}, y_{i1}, y_{i2}, \ldots, y_{it}, \ldots, y_{i,T_i}),\ i = 1, \ldots, N$

Income careers of the first four employees in the data set

```
[1] 4 4 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5
[2] 1 1 0 0 0 0 0 2 0 0 0 0 0 0 0 0 0 0 0 0 0 4 4 4 4 4
[3] 4 0 0 1 0 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 1 0 5
[4] 3 2 3 5 4 4 4 4 5 5 2 3 3 2 3 3 3 4 4 4 4 4 4 4 4 4 4
```


Illustration

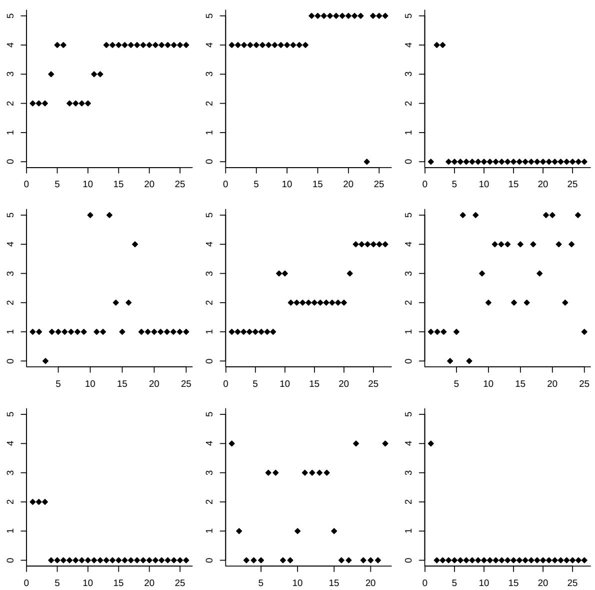

Figure 1: Individual wage mobility time series of nine selected employees.

Markov Chain Model

$y_{it} = k$ if subject $i \in \{1, \ldots, N\}$ belongs to wage category $k \in \{0, 1, \ldots, K\}$ in year $t \in \{0, \ldots, T_i\}$

Markov chain $\mathbf{y}_i$ is modeled with a (time-homogeneous) **Markov process** with unknown **transition matrix** $\boldsymbol{\xi}$, where

$$\xi_{jk} = \mathsf{P}\{y_{it} = k | y_{i,t-1} = j\} \qquad \text{and} \qquad \sum_{k=0}^{K} \xi_{jk} = 1$$

$$\boldsymbol{\xi} = \begin{pmatrix} \boldsymbol{\xi}_{0\cdot} \\ \boldsymbol{\xi}_{1\cdot} \\ \vdots \\ \boldsymbol{\xi}_{K\cdot} \end{pmatrix} = \begin{pmatrix} \xi_{00} & \xi_{01} & \cdots & \xi_{0K} \\ \xi_{10} & \xi_{11} & \cdots & \xi_{1K} \\ \vdots & & \ddots & \vdots \\ \xi_{K0} & \xi_{K1} & \cdots & \xi_{KK} \end{pmatrix}$$

Bayesian Analysis

Prior-distribution of $\boldsymbol{\xi}_{j\cdot}$, $j = 0, \ldots, K$:

$$\boldsymbol{\xi}_{j\cdot} \sim \mathcal{D}(e_{0,j0}, \ldots, e_{0,jK}).$$

Posterior-distribution of $\boldsymbol{\xi}_{j\cdot}$:

$$\boldsymbol{\xi}_{j\cdot} \sim \mathcal{D}(e_{N,j0}, \ldots, e_{N,jK}) \quad \text{with} \quad e_{N,jk} = e_{0,jk} + N_{jk},$$

where $N_{jk} = \#\{y_{it} = k, y_{i,t-1} = j\}$ is the number of transitions from state j to state k over **all** subjects $i = 1, \ldots, N$

$\Rightarrow$ $\boldsymbol{\xi}$ $\sim$ product of $(K+1$ indep.$)$ $\mathcal{D}$irichlet-distributions

Modeling Hidden Groups

Assumptions and **notations**

- H hidden groups with **group-specific** transition matrices $\boldsymbol{\xi}_h$, $h = 1, \ldots, H$
- **Individual** transition matrices $\boldsymbol{\xi}_i^s$, $i = 1, \ldots, N$
- Latent **indicator variable** $\mathbf{S} = (S_1, \ldots, S_N)$ for group membership: $S_i = h$, if subject i belongs to group h
- Relative **group sizes** $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_H)$: $\mathrm{P}\{S_i = h|\boldsymbol{\eta}\} = \eta_h$, $h = 1, \ldots, H$

Modeling Heterogeneity

1. Simple model:

$$\boldsymbol{\xi}_i^s|(S_i = h) \;=\; \boldsymbol{\xi}_h \quad \text{(fixed)}$$

$\Rightarrow$ $\boldsymbol{\xi}_h|\mathbf{S}$ $\sim$ product of $(K+1$ indep.$)$ Dirichlet-distributions

2. Apply a **multinomial logit model with random effects** (Rossi et al., 2005). High-parametrical model including high-dimensional covariance matrices

3. **Dirichlet Multinomial Model**:

$$\boldsymbol{\xi}_{i,j\cdot}^s|(S_i = h) \sim \mathcal{D}(e_{h,j0}, \ldots, e_{h,jK})$$

with **group-specific** parameter $\mathbf{e}_h = \{\mathbf{e}_{h,j\cdot}\}$, $j = 0, \ldots, K$

Dirichlet Multinomial Model

Group-specific transition matrix $\boldsymbol{\xi}_h$ is given by

$$\xi_{h,jk} = E(\xi_{i,jk}^s|S_i = h, \mathbf{e}_h) = \frac{e_{h,jk}}{\sum_{k=0}^{K} e_{h,jk}}$$

So each row of $\mathbf{e}_h$ **determines** the corresponding row of $\boldsymbol{\xi}_h$

Finite mixture model representation:

$\mathbf{Y}_i \sim p_h(\mathbf{y}_i|\mathbf{e}_h)$ ... product of $K+1$ Dirichlet-distributions

Unconditional density:

$$p(\mathbf{Y}_i|\mathbf{e}_1, \ldots, \mathbf{e}_H) = \sum_{h=1}^{H} \eta_h \, p_h(\mathbf{y}_i|\mathbf{e}_h)$$

Group-specific parameter $\mathbf{e}_h$

The variance of $\xi^s_{i,jk}$ is given by

$$Var(\xi^s_{i,jk}|S_i = h, \mathbf{e}_h) = \xi^2_{h,jk} \cdot \frac{\sum_{l \neq k} e_{h,jl}}{\sum_{k=0}^{K} e_{h,jk} \cdot \left(1 + \sum_{k=0}^{K} e_{h,jk}\right)}$$

If $\sum_{k=0}^{K} e_{h,jk}$ is very **large** (for each row in each group) $\rightarrow$ amount of heterogeneity (in each group) is small $\Rightarrow$ leads to the simple model with fixed $\boldsymbol{\xi}_h$

If $\sum_{k=0}^{K} e_{h,jk}$ is **small** $\Rightarrow$ the individual transition matrices are allowed to **deviate** from the group mean within each group

Bayesian Analysis

Prior-assumptions:

- All $\mathbf{e}_{h,j\cdot}$ are *independent* and $\mathbf{e}_{h,j\cdot} - 1 \geq 0$ (to avoid problems with empty groups and non-informative priors)
- $\mathbf{e}_{h,j\cdot} - 1$ is a *discrete*-valued multivariate random variable
- $\mathbf{e}_{h,j\cdot} - 1 \sim$ negative multinomial distribution
- $\boldsymbol{\eta} \sim$ Dirichlet-distribution

All parameters $\mathbf{e}_1, \ldots, \mathbf{e}_H$, $\mathbf{S}$, $\boldsymbol{\eta}$ are jointly **estimated** by means of MCMC-Sampling

MCMC-Estimation (Gibbs Sampler)

Choose initial values for $\boldsymbol{\eta}$ and $\mathbf{e}_1, \ldots, \mathbf{e}_H$ (H fixed in advance) and repeat following steps ($m = 1, \ldots, M$):

1. **Bayes-classification** for each subject i:
   draw $S_i^{(m)}$ from $p(S_i|\mathbf{y}_i, \boldsymbol{\eta}^{(m-1)}, \mathbf{e}_1^{(m-1)}, \ldots, \mathbf{e}_H^{(m-1)})$.
2. sample **Group sizes** $\boldsymbol{\eta}$:
   draw $\boldsymbol{\eta}^{(m)}$ from $\mathcal{D}(\alpha_1^{(m)}, \ldots, \alpha_H^{(m)})$ with $\alpha_h^{(m)} = N_h^{(m)} + \alpha_0$ and $N_h^{(m)} = \#\{S_i^{(m)} = h\}$.
3. sample **group-specific parameters** $\mathbf{e}_1, \ldots, \mathbf{e}_H$:
   draw $\mathbf{e}_{h,j\cdot}^{(m)}$ row-by-row from $p(\mathbf{e}_{h,j\cdot}|\mathbf{y}, \mathbf{S}^{(m)})$ (not of closed form!) using a **Metropolis-Hastings step** (with discrete random walk proposal).

Estimation Results

Here we show the results for **3 groups** which allow very sensible interpretations according to our economist ($M =$ 10,000 with 2,000 burn-in)

- Transition probabilities
- Typical group members
- Classification probabilities
- Equilibrium distributions

Transition Probabilities

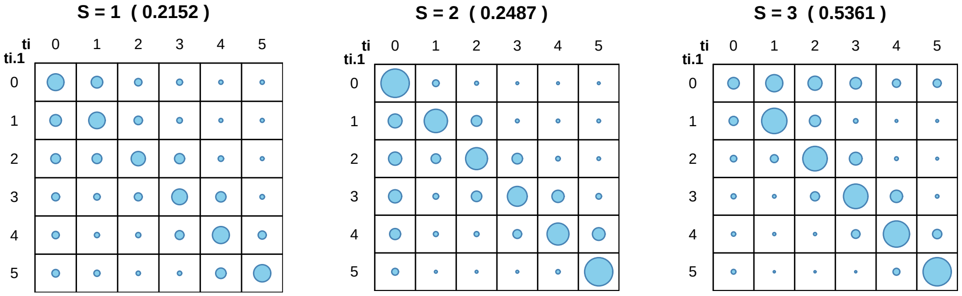

Figure 2: 3D-Visualizations of **transition probabilities** $\hat{\xi}_h$ (*volumes* of balls are proportional to probs) and estimated **group sizes** $\hat{\eta}$ indicated in brackets (posterior means).

Typical Group Members

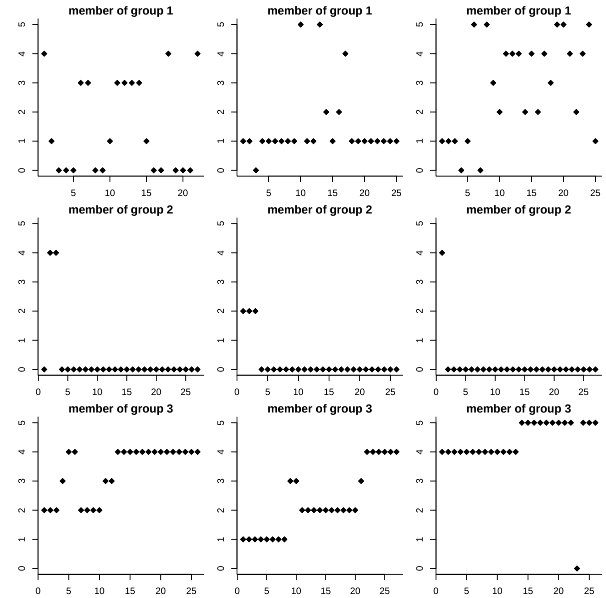

Figure 3: Selected typical group members (with high classification prob).

Classification Probabilities

$i\backslash h$	1	2	3
1	0.00016	0.35852	0.64132
2	0.01319	0.98676	0.00005
3	0.13440	0.25522	0.61039
4	0.34690	0.00462	0.64848
5	0.00035	0.99965	0.00000
6	0.13326	0.86632	0.00042
7	0.00011	0.99989	0.00000
8	0.81248	0.18748	0.00004
9	0.00008	0.99992	0.00000
10	0.05821	0.18316	0.75863
		⋮	
9809	0.51099	0.29038	0.19863

Table 1: Classification probabilities for each individual.

Equilibrium Distributions

$j\backslash h$	1	2	3
0	0.25028	0.60154	0.03993
1	0.22435	0.10482	0.10655
2	0.13299	0.06598	0.13688
3	0.14742	0.03524	0.16979
4	0.15030	0.03786	0.23205
5	0.09466	0.15456	0.31480

Table 2: Equilibrium distributions in each group.

Open Problem

Further research has to be done to find **formal criterions** to determine the number of groups.

Possible approaches:

- Model selection based on marginal likelihoods
- Classification likelihood information criterion (using entropy)
- Integrated classification likelihood

Summary

- Discrete-valued time series
- Categorical variable
- Markov chains
- Individual transition matrices
- Dirichlet multinomial model (allows for heterogeneity within groups):
  mixture model with (products of) Dirichlet-distributions with group-specific parameters
- Estimation via MCMC (number of groups fixed)
- → Group-specific transition matrices

References

Frühwirth-Schnatter, Sylvia (2006). Finite Mixture and Markov Switching Models. Springer Series in Statistics. New York: Springer (to appear).

Frühwirth-Schnatter, Sylvia and **Kaufmann**, Sylvia (2004). Model-Based Clustering of Multiple Time Series. IFAS Research Paper Series, 2004-02, http://www.ifas.jku.at/.

Rossi, Peter E., **Allenby**, Greg and **McCulloch**, Rob (2005). Bayesian Statistics and Marketing. John Wiley and Sons.

Weber, Andrea (2002). State Dependence and Wage Dynamics: A Heterogeneous Markov Chain Model for Wage Mobility in Austria, Economics Series 114, Institute for Advanced Studies.