



**A Bayesian Analysis of Female Wage Dynamics Using
Markov Chain Clustering**

by

Christoph PAMMINGER
Regina TÜCHLER

Working Paper No. 1104

July 2011

Supported by the
Austrian Science Funds

FWF

**The Austrian Center for Labor
Economics and the Analysis of
the Welfare State**

JKU Linz
Department of Economics
Altenberger Strasse 69
4040 Linz, Austria
www.laborrn.at

christoph.pamminger@jku.at
phone +43 (0)70 2468 - 1294

A Bayesian Analysis of Female Wage Dynamics Using Markov Chain Clustering

Christoph Pamminer¹ and Regina Tüchler²

¹Department of Applied Statistics, Johannes Kepler University Linz,
Austria

²Wirtschaftskammer Österreich, Vienna, Austria

Abstract: In this work, we analyze wage careers of women in Austria. We identify groups of female employees with similar patterns in their earnings development. Covariates such as e.g. the age of entry, the number of children or maternity leave help to detect these groups. We find three different types of female employees: (1) “high-wage mums”, women with high income and one or two children, (2) “low-wage mums”, women with low income and ‘many’ children and (3) “childless careers”, women who climb up the career ladder and do not have children.

We use a Markov chain clustering approach to find groups in the discrete-valued time series of income states. Additional covariates are included when modeling group membership via a multinomial logit model.

Zusammenfassung: In dieser Arbeit analysieren wir Einkommensverläufe von Frauen in Österreich. Unser Ziel ist es, mit Hilfe von zusätzlichen erklärenden Variablen wie z.B. dem Alter bei Berufseintritt, der Kinderanzahl oder einer Karenzvariable, Gruppen von erwerbstätigen Frauen mit ähnlichen Einkommensmustern zu finden. Es ergeben sich drei verschiedene Gruppen: (1) Frauen mit hohem Einkommen und ein bis zwei Kindern, (2) Frauen mit geringem Einkommen und ‘vielen’ Kindern und (3) Frauen, die im Laufe des Erwerbslebens die “Karriereleiter” hinaufklettern und keine Kinder haben.

Wir verwenden eine *Markov Chain Clustering* Methode, um auf Basis der diskreten Einkommensdaten Gruppen von Frauen zu identifizieren. Die Gruppenzugehörigkeit wird mit Hilfe eines multinomialen logistischen Modells, in dem zusätzliche erklärende Variable berücksichtigt werden, geschätzt.

Keywords: Income Career, Transition Data, Multinomial Logit, Auxiliary Mixture Sampler, Markov Chain Monte Carlo.

1 Introduction

In this paper, we analyze individual wage careers of Austrian women who started to work between 1980 and 1985. Our administrative data set covers a long time period of about twenty years and ends in 2001.

Many studies analyze the dependence of the amount of earnings on different attributes at one specific point of time. On the one hand, individual attributes are assumed to be correlated with individual productivity and therefore to be influential on the amount of individual earnings. On the other hand, it was found out that general economic conditions and

firm specific attributes are also important. Individual attributes, which are often included in such studies, are for example age, gender, education, professional status of occupation, whereas economic activity is an influential firm-specific attribute and the general situation of the labor market is for example reflected in the unemployment rate. For Austrian employees the interrelationship between earnings and different attributes was investigated on the basis of structure of earnings data by Geisberger (2007) and on the basis of data on income and living conditions by Grünberger and Zulehner (2009).

In the present paper, we take a different approach. We are interested in the development of earnings over time and how this development depends on individual attributes. We aim at finding typical female career patterns and at classifying women into groups with similar patterns. We will see that career patterns of women highly depend on individual attributes and that variables like maternity leave or motherhood are especially important. Therefore, we concentrate on individual attributes in this work and do not take into account the economic situation. A similar analysis including individual as well as economic variables was carried out for male Austrian employees by Frühwirth-Schnatter et al. (2011).

For building and estimating our model we follow the ideas of Frühwirth-Schnatter et al. (2011). Clustering of Markov chain models is used to find homogeneous groups of the discrete-valued income data. It is assumed that the income states of all individuals of one specific group are modeled by the same group-specific transition matrix. This model was presented by Pamminer and Frühwirth-Schnatter (2010). As in Frühwirth-Schnatter et al. (2011) we additionally include covariates via a multinomial logit model in the prior of the cluster weights.

The paper is structured as follows. In Section 2, the administrative data set is introduced. Section 3 gives an overview over the model used to analyze female wage dynamics and the results are described in Section 4. The paper is summarized in Section 5.

2 Data Description

For our investigation, we consider data from the Austrian Social Security Data Base (ASSD), which contains detailed longitudinal information on employment and earnings of all private sector employees in Austria since 1972 (Zweimüller et al., 2009).

Our cohort study is based on a data set consisting of $N = 183\,805$ female Austrian employees, who entered the labor market for the first time in the years 1980 to 1985 and were 14 to at most 25 years old at entry. Two out of three women were between 17 and 19 years old at entry.

Due to a change of the qualifying conditions for maternity leave in the beginning of 2002 we had to cut the observation period after the year 2001 to omit inconsistencies concerning maternity leave as since then much more ‘newly mums’ became eligible for maternity leave and hence childcare benefit (Kinderbetreuungsgeld).

As in Frühwirth-Schnatter et al. (2011) we take yearly earnings observations measured by gross monthly wages representing May. The wages are observed for a period between 2 (to have at least one transition observed) to 22 successive years for each individual. The median time length in our panel is equal to 14 years.

Following Weber (2001), we divide the gross monthly wage into six categories labeled with 0 up to 5. Category zero corresponds to zero-income, i.e. unemployment or out of labor force. The categories one to five correspond to the quintiles of the income distribution which are calculated for each year from all non-zero wages observed in that year for the total population of female employees in Austria. Note that the data set excludes marginal employees and, as hours of work are not available in our data set, we are not able to distinguish between part- and full-time jobs.

There are two important advantages by using wage categories: we do not need any inflation adjustment; and we do not have to be worried about the fact that recorded wages are right-censored because of a social security payroll tax cap.

As in Frühwirth-Schnatter et al. (2011) we cut the time series of individuals after observing more than five consecutive years with zero income, because these individuals have most likely stopped being employed. For individuals first observed in the data as apprentices, we consider their first wage after finishing education as the point of job entry.

As we are interested in characterizing the wage path of women since their first job with respect to covariates, we include the following predetermined variables:

We use a dummy for motherhood, which indicates if the (female) job entrant has ever been at least once in maternity leave (that are 72.6 % of the women in our data). This means that there was either a notice of the (live) birth of a child followed by maternity leave within nine months or an adoption leave.

The number of children indicates how often a person has had a live birth announcement. The average number of children is equal to 1.417 and the median equals 1.

We incorporate the age at entry into the model and introduce a linear and a quadratic age effect.

We include the color of the collar at job entry. 59.2 % of the women in our data started as white collar worker and 40.8 % as blue collar worker.

To include interaction effects between maternity leave and the color of the collar we build four categories out of these two variables: we define the combination blue collar & no maternity leave as baseline category (that are 9.49 % of the women in our data), further blue collar & maternity leave (31.35 %), white collar & no maternity leave (17.90 %) and white collar & maternity leave (41.26 %).

As in Frühwirth-Schnatter et al. (2011), we also introduce dummies for the initial state (wage category) to take the initial conditions problem into account (see Appendix A).

3 The Model

3.1 Markov Chain Clustering

Let $\mathbf{y}_i = \{y_{i1}, \dots, y_{iT_i}\}$ denote for each individual $i = 1, \dots, N$ the time series of income states, excluding the initial state y_{i0} . The categorical income variables $y_{it}, i = 1, \dots, N, t = 1, \dots, T_i$ take one of K states labeled $\{1, \dots, K\}$. Following Pamminger and Frühwirth-Schnatter (2010) we classify the individuals into H groups. We assume that the clustering kernel $p(\mathbf{y}_i | \boldsymbol{\xi}_h)$ depends on the cluster-specific unknown parameter vector $\boldsymbol{\xi}_h$ for each of the $h = 1, \dots, H$ hidden clusters and that the clustering kernel $p(\mathbf{y}_i | \boldsymbol{\xi}_h)$ sufficiently describes all time series within group h , see Frühwirth-Schnatter

and Kaufmann (2008). As we introduce a group indicator $S_i \in \{1, \dots, H\}$ for each individual i we obtain $p(\mathbf{y}_i | S_i, \boldsymbol{\xi}_h) = p(\mathbf{y}_i | \boldsymbol{\xi}_{S_i})$.

For Markov chain clustering we introduce separate transition processes for each (hidden) group through a first-order time-homogeneous Markov chain model with a cluster-specific transition matrix $\boldsymbol{\xi}_h$. The elements of $\boldsymbol{\xi}_h$ are equal to

$$\xi_{h,jk} = \Pr(y_{it} = k | y_{i,t-1} = j, S_i = h), \quad j, k = 1, \dots, K. \quad (1)$$

Hence, each row of $\boldsymbol{\xi}_h$ represents a probability distribution over the discrete set $\{1, \dots, K\}$, i.e. $\sum_{k=1}^K \xi_{h,jk} = 1$. The clustering kernel $p(\mathbf{y}_i | \boldsymbol{\xi}_h)$ reads:

$$p(\mathbf{y}_i | \boldsymbol{\xi}_h) = \prod_{t=1}^{T_i} p(y_{it} | y_{i,t-1}, \boldsymbol{\xi}_h) = \prod_{j=1}^K \prod_{k=1}^K \xi_{h,jk}^{N_{i,jk}}, \quad (2)$$

where $N_{i,jk} = \#\{y_{it} = k, y_{i,t-1} = j\}$ is the number of transitions from state j to state k observed in time series i . Note that we condition in (2) on the first observation y_{i0} and the actual number of observations is equal to T_i for each time series.

3.2 Mixture-of-Experts Model for Modeling Prior Group Membership

We assume that individual attributes help to identify groups of female employees with similar career patterns. Therefore, we use a mixture-of-experts model (see e.g. Frühwirth-Schnatter, 2006, Chapter 8.6.3; Peng et al., 1996) to specify the prior for the individual group indicators. Prior group membership $\Pr(S_i = h)$ is modeled as a multinomial logit model (MNL):

$$\Pr(S_i = h | \boldsymbol{\beta}_2, \dots, \boldsymbol{\beta}_H) = \frac{\exp(\mathbf{x}_i \boldsymbol{\beta}_h)}{1 + \sum_{l=2}^H \exp(\mathbf{x}_i \boldsymbol{\beta}_l)}. \quad (3)$$

\mathbf{x}_i is the row vector of predetermined individual covariates, including 1 for the intercept. To determine group $h = 1$ as baseline group in the MNL we set $\boldsymbol{\beta}_1 = \mathbf{0}$ and the unknown group-specific regression coefficients $\boldsymbol{\beta}_2, \dots, \boldsymbol{\beta}_H$ are the effects on the log-odds ratio relative to the baseline group.

3.3 Model Estimation

For model estimation we pursue the Bayesian approach of Frühwirth-Schnatter et al. (2011) for fixed H . \mathbf{S} is estimated along with the group-specific transition matrices $\boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_H$ and regression coefficients $\boldsymbol{\beta}_2, \dots, \boldsymbol{\beta}_H$ from the data. Details on the initial conditions problem and on the choice of the prior may be found in Appendix A.

After choosing initial values for the group indicators \mathbf{S} we repeat the following steps:

- (i) **Sample the cluster-specific transition matrices $\boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_H$ given \mathbf{S} and \mathbf{y} :**

The conditionally independent rows are sampled from a total of $K \cdot H$ Dirichlet

distributions:

$$\xi_{h,j} \cdot | \mathbf{S}, \mathbf{y} \sim \mathcal{D} (e_{0,j1} + N_{j1}^h(\mathbf{S}), \dots, e_{0,jK} + N_{jK}^h(\mathbf{S})),$$

$$j = 1, \dots, K, \quad h = 1, \dots, H,$$

where $N_{jk}^h(\mathbf{S}) = \sum_{i:S_i=h} N_{i,jk}$ is the total number of transitions from j to k observed in group h .

(ii) **Sample the regression coefficients β_2, \dots, β_H conditional on \mathbf{S} :**

The likelihood $p(\mathbf{S} | \beta_2, \dots, \beta_H)$ is obtained from the multinomial logit model (3). To sample β_2, \dots, β_H we apply auxiliary mixture sampling in the differenced random utility model representation as introduced by Frühwirth-Schnatter and Frühwirth (2010). Details are given in Appendix B and C.

(iii) **Sample \mathbf{S} conditional on $\beta_2, \dots, \beta_H, \xi_1, \dots, \xi_H$ and \mathbf{y} :**

For each individual $i = 1, \dots, N$ we draw S_i from:

$$\Pr(S_i = h | \mathbf{y}_i, \mathbf{x}_i, \beta_2, \dots, \beta_H, \xi_1, \dots, \xi_H) \propto p(\mathbf{y}_i | \xi_h) \cdot \frac{\exp(\mathbf{x}_i \beta_h)}{1 + \sum_{l=2}^H \exp(\mathbf{x}_i \beta_l)},$$

$$h = 1, \dots, H.$$

The MCMC algorithm is carried out for a fixed number H of clusters. This fixed number H is selected by using model selection criteria summarized and described in Frühwirth-Schnatter et al. (2011): we use the integrated classification likelihood (ICL) and the approximate weight of evidence (AWE) to determine the number of clusters H .

4 Results

We analyze patterns in the earnings development of young female labor market entrants in Austria over their life cycle. Further, we investigate the effects of individual's observable characteristics such as number of children or maternity leave on the probability of belonging to each transition type.

Therefore, we applied Markov chain clustering for two up to four groups and simulated 10 000 MCMC draws including a burn-in of 5000 draws. Finally, we applied a thinning parameter equal to 5 and used the remaining 1000 draws for posterior inference. Results were confirmed by starting the algorithm from different starting values. We used the priors specified in Appendix A. Additionally, we ran the algorithm with an uninformative prior for the transition matrices and the results proved to be extremely robust.

4.1 Choosing the Number of Groups

To choose the number of clusters we apply two criteria, namely the integrated classification likelihood criterion (ICL) and the approximate weight of evidence criterion (AWE). Both criteria agree in choosing three groups as the optimal number of groups, see Figure 1. As this choice additionally proves to yield meaningful results from the economic point of view we carry out the rest of the analysis for this three-cluster solution.

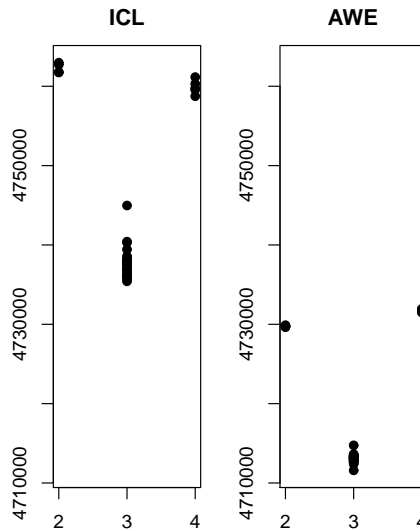


Figure 1: Model selection criteria ICL (integrated classification likelihood) and AWE (approximate weight of evidence) for various numbers H of clusters and several independent MCMC runs.

4.2 Results of Markov Chain Clustering

Figure 2 visualizes the posterior transition probabilities ξ_1 , ξ_2 and ξ_3 . In these balloon plots, the size of each circle corresponds to the respective entry in the transition matrix. Numerical results for the posterior transition probabilities as well as their standard deviations are given in Table 3 in Appendix D. On the other hand, in Figure 3 the balloon plots represent the contingency tables. From this picture it can be seen where the transitions mainly occur.

The largest group is the “low-wage” group with 54.62 %. Individuals of that group have a higher risk to move into (next) lower or no income categories. Transitions occur mainly in the lower wage categories. On the contrary, the “high-wage” and the “career ladder” clusters are of group size 28.02 % and 17.36 %, respectively. Women in these two clusters have a higher chance to move upwards into (the next) higher wage categories. The “career ladder” group more likely stays within the (same) wage category. We observe transitions mainly in the higher wage categories, whereas, for the “high-wage” cluster we find transitions in either end of the income distribution. This fact separates the “high-wage” from the “career ladder” group. Both end up in the higher wage categories, but contrary to the “career-ladder” group “high-wage” employees also move into the no income category.

Figure 4 shows the most typical group members and confirms the interpretation of the transition probabilities. “High-wage” women earn high wages but also spend time in the no income category whereas the “career ladder” employees climb up the ladder, what can be perfectly seen from these pictures. The “low-wage” group remains mainly in the lower quintiles of the income distribution with ‘visits’ to the no income category.

Figure 5 shows the development of the distribution of the wage categories over the years starting in the initial distribution (determined from the initial states in the data) up

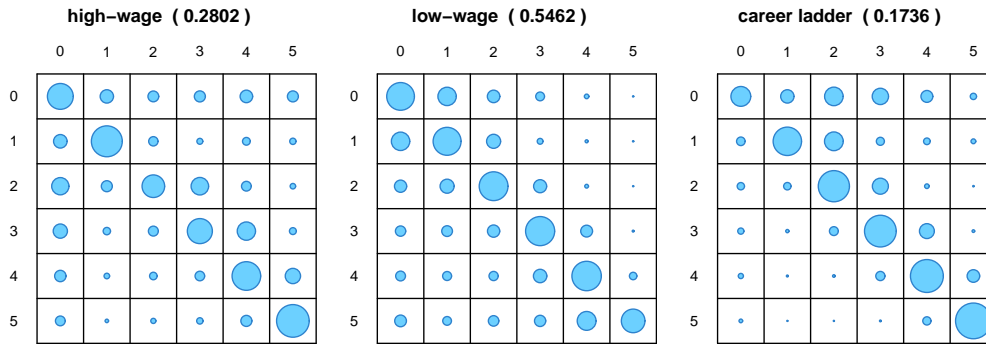


Figure 2: Balloon plots of posterior expectation of the transition matrices ξ_1 , ξ_2 and ξ_3 obtained by Markov chain clustering. The circular areas are proportional to the size of the corresponding entry in the transition matrix. The corresponding group sizes are calculated based on the posterior classification probabilities and are indicated in parentheses.

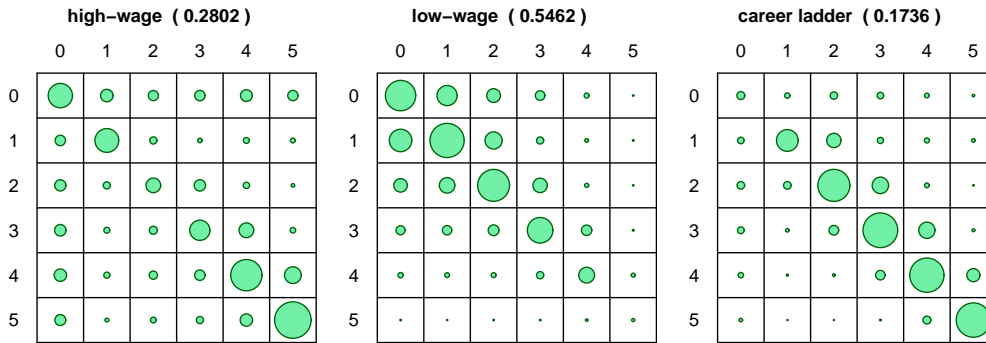


Figure 3: Balloon plots of cluster specific contingency tables. For each cluster we find in cell (j, k) the probability $\Pr(y_{i,t-1} = j, y_{it} = k | S_i = h)$ of observing wage categories (j, k) in consecutive years for an individual in this cluster. The entries of this table sum to one.

to the steady state. The majority of the “career ladder” group ends up in the highest wage category whereas the “low-wage” women stay within the lower income categories as well as in the no income state. The “high-wage” employees show some kind of mixed tendencies: approximately 50% end up in the two highest wage categories and approximately 30% in the lowest or no income category. The “high-wage” and “low-wage” employees reach the steady state within a decade, whereas the “career ladder” individuals maybe will never reach the steady state within their working life.

An important criterion indicating the quality of the clustering procedure is the segmentation power (which equals one minus the misclassification risk) shown in Table 1: three out of four individuals are assigned with at least 73.8% probability to their respective cluster. The segmentation power for the “low-wage” cluster is much higher than for the other groups, whereas the segmentation power for the “high-wage” group is slightly lower than for the “career ladder” cluster. Three in four “low-wage” women are assigned with at least 79.6% to their group. Three-quarters of the “high-wage” individuals have an

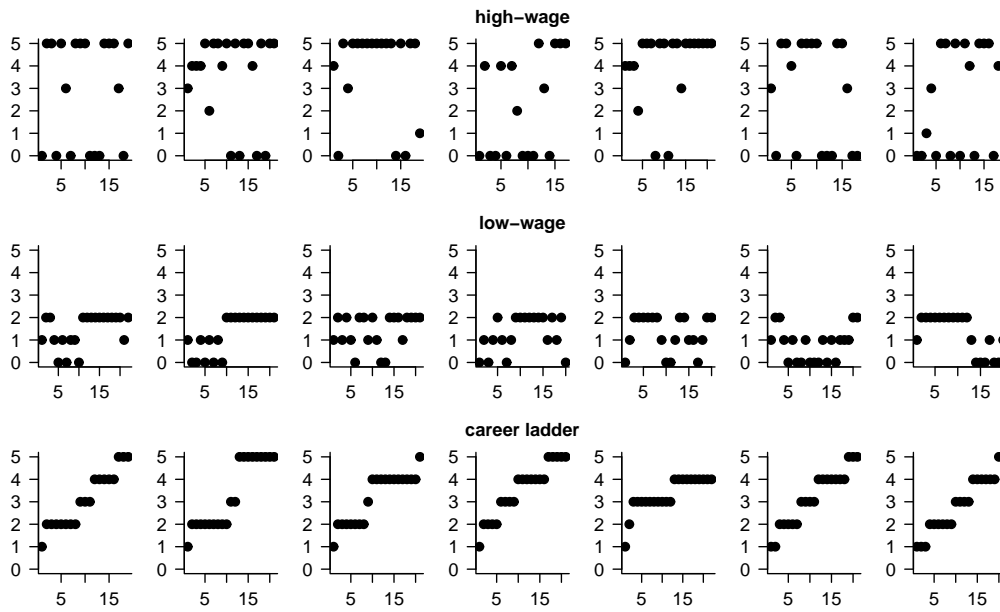


Figure 4: Typical group members showing the highest classification probabilities.

assignment probability of at least 65.7 % and three-quarters of the “career ladder” women have an assignment probability of at least 67.9 %.

Table 1: Segmentation power of Markov chain clustering; reported are the lower quartile, the median and the upper quartile of the individual posterior classification probabilities for all individuals within a certain cluster as well as for all individuals.

	Markov chain clustering		
	1st Qu.	Median	3rd Qu.
“high-wage”	0.6574	0.8613	0.9770
“low-wage”	0.7956	0.9323	0.9843
“career ladder”	0.6790	0.8703	0.9674
overall	0.7378	0.9110	0.9807

4.3 Results of the Multinomial Logit Model

Table 2 shows the posterior regression coefficients for the MNL (3). We choose the “low-wage” group as the baseline group. The credibility intervals do not cover zero for any coefficient.

The interaction model shows that blue collar workers who were at least once in maternity leave (compared to blue collar workers never been in maternity leave) more likely belong to the “low-wage” group than to the other groups. White collar workers never been in maternity leave (compared to blue collar workers never been in maternity leave) belong

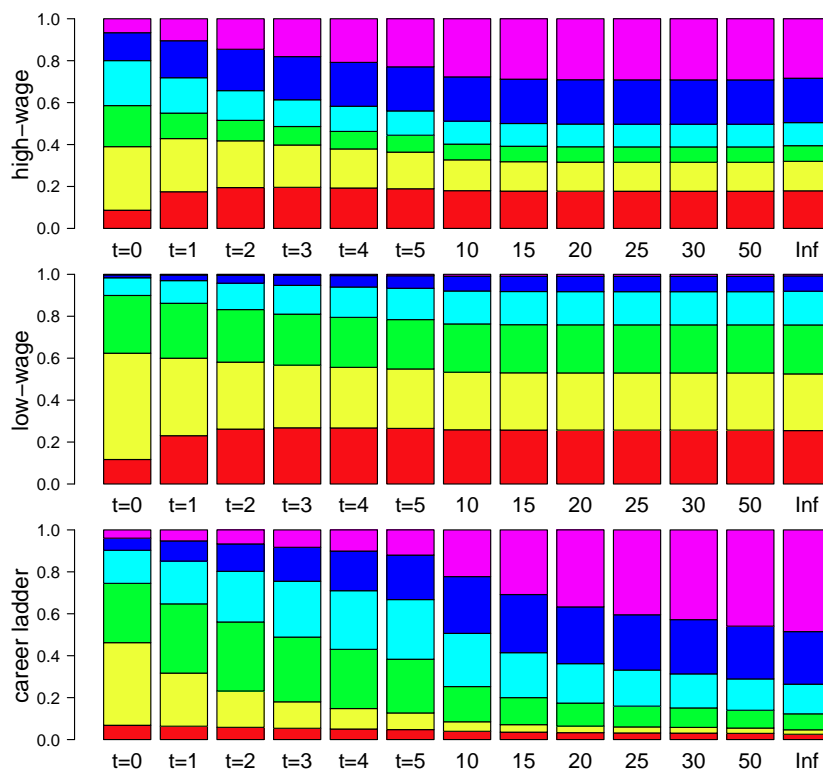


Figure 5: Posterior expectation of the wage distribution for wage category 0 (bottom) to category 5 (top) after a period of t years in the various clusters.

more likely to the “high-wage” or the “career ladder” cluster than to the “low-wage” cluster, whereas white collar workers having been in maternity leave have higher probability to belong to the “high-wage” group than to the “low-wage” group and much less likely to the “career ladder” than to the “low-wage” group. To sum up, having been in maternity leave reduces the chance to belong either to the “high-wage” or to the “career ladder” group (instead of belonging to the “low-wage” group) except for white collar workers.

Not surprisingly, with growing number of children it becomes less likely to belong to either “high-wage” or “career ladder” group. The ex post analyses show that the median number of children within group “career ladder” is equal to 0, in the “high-wage” group 1 and in the “low-wage” group 2, which confirms these results.

We include the age and the quadratic age effect in our model and find that with growing age at entry the allocation to either “high-wage” or “career ladder” group is more likely.

Finally, the higher a woman starts in the income distribution the more likely she belongs to the “high-wage” or to the “career ladder” group instead of belonging to the “low-wage” group.

To summarize the interpretation of the estimation results we finally found the following ‘labeling’ for the three types of earnings careers:

“high-wage mums:” women with high income and one or two children

“low-wage mums:” women with low income and ‘many’ children

Table 2: Multinomial logit model to explain group membership in a particular cluster (baseline: “low-wage” cluster); the numbers are the posterior expectation and, in parenthesis, the posterior standard deviation of the various regression coefficients.

	“high-wage”	“career ladder”
Intercept	7.98821 (0.60443)	5.05117 (0.90168)
blue collar \times maternity leave	-1.11074 (0.04926)	-6.60941 (0.18408)
white collar \times no maternity leave	2.38276 (0.07639)	2.66115 (0.06684)
white collar \times maternity leave	0.42969 (0.04155)	-1.79956 (0.09450)
Number of children	-0.22701 (0.01394)	-0.63768 (0.03005)
Age at start	-1.10084 (0.06391)	-0.53595 (0.09568)
Age at start (squared)	0.03443 (0.00170)	0.01344 (0.00255)
Start in wage category 1	-0.08364 (0.03538)	0.24558 (0.05567)
Start in wage category 2	-0.20876 (0.03512)	0.59434 (0.05880)
Start in wage category 3	0.84706 (0.04141)	1.31027 (0.06498)
Start in wage category 4	1.80361 (0.05405)	1.95150 (0.08756)
Start in wage category 5	2.05103 (0.10402)	2.58711 (0.13330)

“**childless careers:**” women who climb up the career ladder and do not have children

5 Conclusions

In this paper, we investigated wage careers of women in Austria. More precisely, we analyzed patterns in the earnings development of female employees in order to identify groups of women with similar transition patterns between the discrete earnings states. We were interested whether covariates such as age at entry, maternity leave, collar of work, etc., which are commonly known to be influential on the amount of the wage, help to find those groups of employees with similar wage careers.

The results, which are based on a large administrative data set, suggest that motherhood in general, the number of children, the color of the collar and the entry age have a strong impact on the mobility pattern of women throughout their working life. We are able to identify three distinctly different types of female employees: (1) “high-wage mums”, women with high income and one or two children; (2) “low-wage mums”, women with low income and ‘many’ children and (3) “childless careers”, women climbing up the career ladder and do not have children.

To estimate our statistical model we use Markov chain clustering, proposed by Paminger and Frühwirth-Schnatter (2010), which is a model-based clustering approach for clustering discrete-valued time series obtained by observing a categorical variable with several states. This method is based on finite mixtures of first-order time-homogeneous Markov chain models. In order to analyze group membership in dependence on additional covariates, we follow Frühwirth-Schnatter et al. (2011) and include a probabilistic model for the latent group indicators within the Bayesian classification rule using a multinomial logit model.

The choice of the number of groups was based on the AWE (approximate weight

of evidence) and ICL (integrated classification likelihood) to take also into account the quality of the partitioning. Remarkably, the segmentation power is quite high: three out of four individuals are assigned with at least 73.8 % probability to their respective cluster.

Acknowledgments We thank Sylvia Frühwirth-Schnatter for her helpful comments and discussions and also our colleagues at the Department of Applied Statistics at the Johannes Kepler University, Linz, Austria for their remarks.

This research is supported by the Austrian Science Foundation (FWF) under the grant S 10309-G14 (NRN “The Austrian Center for Labor Economics and the Analysis of the Welfare State”, Subproject “Bayesian Econometrics”).

References

- Frühwirth-Schnatter, S. (2006). *Finite Mixture and Markov Switching Models*. New York: Springer.
- Frühwirth-Schnatter, S., & Frühwirth, R. (2010). Data augmentation and MCMC for binary and multinomial logit models. In T. Kneib & G. Tutz (Eds.), *Statistical Modelling and Regression Structures – Festschrift in Honour of Ludwig Fahrmeir* (pp. 111–132). Heidelberg: Physica-Verlag. (Also available at http://www.ifas.jku.at/e2550/e2756/index_ger.html, IFAS Research Paper Series 2010-48)
- Frühwirth-Schnatter, S., & Kaufmann, S. (2008). Model-based clustering of multiple time series. *Journal of Business & Economic Statistics*, 26, 78–89.
- Frühwirth-Schnatter, S., Pamminer, C., Weber, A., & Winter-Ebmer, R. (2011). Labor market entry and earnings dynamics: Bayesian inference using mixtures-of-experts Markov chain clustering. *Journal of Applied Econometrics*, 26, n/a–n/a. Available from <http://dx.doi.org/10.1002/jae.1249>
- Geisberger, T. (2007). Geschlechtsspezifische Lohn- und Gehaltsunterschiede. *Statistische Nachrichten*, 7, 633–642.
- Grünberger, K., & Zulehner, C. (2009). Geschlechtsspezifische Lohnunterschiede in Österreich. *Wifo Monatsberichte*, 2, 139–150.
- McFadden, D. (1974). Conditional logit analysis of qualitative choice behaviour. In P. Zarembka (Ed.), *Frontiers of econometrics* (pp. 105–142). New York: Academic.
- Pamminer, C., & Frühwirth-Schnatter, S. (2010). Model-based Clustering of Categorical Time Series. *Bayesian Analysis*, 5(2), 345–368. Available from <http://ba.stat.cmu.edu/journal/2010/vol05/issue02/pamminer.pdf>
- Peng, F., Jacobs, R. A., & Tanner, M. A. (1996). Bayesian inference in mixtures-of-experts and hierarchical mixtures-of-experts models with an application to speech recognition. *Journal of the American Statistical Association*, 91, 953–960.
- Weber, A. (2001). *State dependence and wage dynamics: A heterogeneous Markov chain model for wage mobility in Austria* (Research report). Vienna: Institute for Advanced Studies.
- Zweimüller, J., Winter-Ebmer, R., Lalive, R., Kuhn, A., Wuellrich, J.-P., Ruf, O., et al. (2009). *The Austrian Social Security Database (ASSD)* (Working Paper No. 0903).

Appendix

A The Initial Conditions Problem and the Prior

Frühwirth-Schnatter et al. (2011) found a simple solution to the initial conditions problem for the discrete case present in this paper. They allow for dependence between the initial states y_{i0} and the group indicators S_i which capture unobserved heterogeneity. The initial states are included in the design matrix \mathbf{x}_i of equation (3). Since the two distributions $p(S_i|y_{i0}, \cdot)$ and $p(y_{i0}|\cdot)$ appearing in the factorization of the joint distribution $p(y_{i0}, S_i|\cdot) = p(S_i|y_{i0}, \cdot) \cdot p(y_{i0}|\cdot)$ do not have parameters in common the marginal distribution $p(y_{i0}|\cdot)$ cancels out and we do not need to specify it.

We assume prior independence between ξ_1, \dots, ξ_H and β_2, \dots, β_H . The a priori independent regression coefficients β_{hj} follow a standard normal prior distribution and the a priori independent K rows $\xi_{h,1}, \dots, \xi_{h,K}$ of ξ_h following a Dirichlet distribution, i.e. $\xi_{h,j} \sim \mathcal{D}(e_{0,j1}, \dots, e_{0,jK})$, with prior parameters $\mathbf{e}_{0,j} = (e_{0,j1}, \dots, e_{0,jK}) = N_0 \cdot \xi_j^*$ where $N_0 = 10$ and

$$\xi^* = \begin{pmatrix} 0.7 & 0.2 & 0.025 & 0.025 & 0.025 & 0.025 \\ 0.15 & 0.6 & 0.15 & 0.03 & 0.03 & 0.03 \\ 0.03 & 0.15 & 0.6 & 0.15 & 0.03 & 0.03 \\ 0.03 & 0.03 & 0.15 & 0.6 & 0.15 & 0.03 \\ 0.03 & 0.03 & 0.03 & 0.15 & 0.6 & 0.15 \\ 0.025 & 0.025 & 0.025 & 0.025 & 0.2 & 0.7 \end{pmatrix}$$

as in Frühwirth-Schnatter et al. (2011).

B MNL as Random Utility Model (RUM)

We rewrite model (3) as random utility model (RUM) as it was introduced by McFadden (1974). Let y_{hi}^u be the utility of choosing group h . Then

$$y_{hi}^u = \mathbf{x}_i \beta_h + \delta_{hi}, \quad h = 1, \dots, H \quad (4)$$

$$S_i = h \Leftrightarrow y_{hi}^u = \max_{l \in \{1, \dots, H\}} y_{li}^u \quad (5)$$

$$i = 1, \dots, N. \quad (6)$$

If $\delta_{1i}, \dots, \delta_{Hi}$ are i.i.d. following a type I extreme value distribution, the MNL (3) results as the marginal distribution of S_i .

An alternative way to write the MNL as an augmented model involving random utilities is the differenced RUM (dRUM), which is obtained by choosing a baseline category (here $h_0 = 1$) and considering the model involving the differences of the utilities:

$z_{hi} = \mathbf{x}_i \boldsymbol{\beta}_h + \varepsilon_{hi}$, where $z_{hi} = y_{hi}^u - y_{1i}^u$. Marginally, the errors $\varepsilon_{hi} = \delta_{hi} - \delta_{1i}$ follow a logistic distribution but are no longer independent across categories.

It can be shown (see Frühwirth-Schnatter & Frühwirth, 2010) that for each h , the MNL has the following representation as partial (binary) dRUM:

$$z_{hi} = \mathbf{x}_i \boldsymbol{\beta}_h - \log\left(\sum_{l \neq h} \lambda_{li}\right) + \varepsilon_{hi}, \quad (7)$$

where $\varepsilon_{hi}, h \neq 1$ are now i.i.d. following a logistic distribution.

C Auxiliary Mixture Sampling

The logistic distribution is approximated for each ε_{hi} by a finite scale mixture of six normal distributions with zero means, variances s_r^2 and weights $w_r, r = 1, \dots, 6$ (for the values of the parameters see Frühwirth-Schnatter & Frühwirth, 2010). Conditional on the latent utilities $\mathbf{z} = \{z_{2i}, \dots, z_{Hi}, i = 1, \dots, N\}$ and the indicators $\mathbf{R} = \{r_{2i}, \dots, r_{Hi}, i = 1, \dots, N\}$ the dRUM (7) reduces to a Gaussian regression model:

$$z_{hi} = \mathbf{x}_i \boldsymbol{\beta}_h - \log\left(\sum_{l \neq h} \lambda_{li}\right) + \varepsilon_i, \quad \varepsilon_i | r_{hi} \sim \mathcal{N}(0, s_{r_{hi}}^2). \quad (8)$$

Thus, we select starting values for \mathbf{z} and \mathbf{R} and repeat the following steps:

- (ii-a) Sample the regression coefficients $\boldsymbol{\beta}_2, \dots, \boldsymbol{\beta}_H$ conditional on \mathbf{z} and \mathbf{R} based on the normal regression model (8) from a multivariate normal density.
- (ii-b) Sample the latent variables z_{hi} and r_{hi} conditional on $\boldsymbol{\beta}_2, \dots, \boldsymbol{\beta}_H$ and \mathbf{S} for $i = 1, \dots, N$ and $h = 2, \dots, H$ with $\lambda_{hi} = \exp(\mathbf{x}_i \boldsymbol{\beta}_h)$:
 - (ii-b-1) Sample all utilities z_{2i}, \dots, z_{Hi} simultaneously for each i from:

$$z_{hi} = \log(\lambda_{hi}^* U_{hi} + I\{S_i = h\}) - \log(1 - U_{hi} + \lambda_{hi}^* I\{S_i \neq h\}),$$

where $U_{ih} \sim \mathcal{U}[0, 1]$ and $\lambda_{hi}^* = \lambda_{hi} / (\sum_{l \neq h} \lambda_{li})$.

- (ii-b-2) Sample the component indicators r_{hi} conditional on z_{hi} from:

$$\Pr(r_{hi} = j | z_{hi}, \boldsymbol{\beta}_h) \propto \frac{w_j}{s_j} \exp \left\{ -\frac{1}{2} \left(\frac{z_{hi} - \mathbf{x}_i \boldsymbol{\beta}_h + \log(\sum_{l \neq h} \lambda_{li})}{s_j} \right)^2 \right\}.$$

D Numerical Results of Transition Probabilities

Table 3: Posterior expectation $E(\xi_h|\mathbf{y})$ and, in parenthesis, posterior standard deviations $SD(\xi_h|\mathbf{y})$ (multiplied by 100) of the transition matrices ξ_h in the various clusters.

"high-wage"						
	0	1	2	3	4	5
0	0.4895(0.691)	0.1310(0.473)	0.0856(0.218)	0.0899(0.198)	0.1154(0.242)	0.0886(0.200)
1	0.1347(0.667)	0.7018(1.213)	0.0642(0.435)	0.0259(0.151)	0.0441(0.227)	0.0293(0.147)
2	0.2188(0.679)	0.0889(0.904)	0.3747(1.235)	0.2256(0.726)	0.0697(0.227)	0.0224(0.091)
3	0.1484(0.283)	0.0398(0.230)	0.0722(0.206)	0.4597(0.715)	0.2449(0.377)	0.0350(0.091)
4	0.0949(0.123)	0.0228(0.063)	0.0419(0.088)	0.0678(0.118)	0.6007(0.294)	0.1719(0.180)
5	0.0702(0.102)	0.0089(0.030)	0.0194(0.049)	0.0298(0.063)	0.0900(0.132)	0.7817(0.263)
"low-wage"						
	0	1	2	3	4	5
0	0.5636(0.272)	0.2477(0.214)	0.1151(0.116)	0.0569(0.074)	0.0155(0.046)	0.0012(0.012)
1	0.2479(0.335)	0.5737(0.546)	0.1451(0.229)	0.0250(0.052)	0.0069(0.030)	0.0013(0.011)
2	0.1099(0.111)	0.1450(0.197)	0.6081(0.214)	0.1253(0.130)	0.0106(0.028)	0.0011(0.008)
3	0.0782(0.102)	0.0842(0.130)	0.1107(0.123)	0.6214(0.265)	0.1027(0.123)	0.0028(0.019)
4	0.0707(0.145)	0.0589(0.126)	0.0620(0.136)	0.1327(0.202)	0.6347(0.335)	0.0410(0.155)
5	0.1009(0.569)	0.0558(0.386)	0.0812(0.532)	0.0925(0.536)	0.2608(1.069)	0.4089(1.563)
"career ladder"						
	0	1	2	3	4	5
0	0.2879(1.391)	0.1308(1.023)	0.2525(0.892)	0.1956(0.642)	0.1040(0.482)	0.0293(0.335)
1	0.0540(0.464)	0.5911(2.031)	0.2576(1.322)	0.0477(0.319)	0.0325(0.282)	0.0170(0.203)
2	0.0397(0.138)	0.0409(0.256)	0.7136(0.351)	0.1886(0.316)	0.0157(0.064)	0.0015(0.021)
3	0.0284(0.083)	0.0080(0.052)	0.0579(0.121)	0.7378(0.261)	0.1626(0.198)	0.0054(0.035)
4	0.0198(0.081)	0.0028(0.028)	0.0051(0.040)	0.0620(0.113)	0.7941(0.221)	0.1161(0.176)
5	0.0100(0.080)	0.0009(0.018)	0.0010(0.017)	0.0024(0.028)	0.0498(0.148)	0.9359(0.182)

Authors' addresses:

Christoph Pamminger
Department of Applied Statistics
Johannes Kepler University Linz
Altenberger Straße 69
A-4040 Linz

Regina Tüchler
Wirtschaftskammer Österreich
Stabsabteilung Statistik
Wiedner Hauptstraße 63
A-1045 Vienna