Analysing Plant Closure Effects Using Time-Varying Mixture-of-Experts Markov Chain Clustering

Sylvia Frühwirth-Schnatter[*]   Stefan Pittner[†]   Andrea Weber[‡]

Rudolf Winter-Ebmer[§]

December 8, 2017

Abstract

In this paper, we study data on discrete labor market transitions from Austria. In particular, we follow the careers of workers who experience a job displacement due to plant closure and observe – over a period of forty quarters – whether these workers manage to return to a steady career path. To analyse these discrete-valued panel data, we apply a new method of Bayesian Markov chain clustering analysis based on inhomogeneous first order Markov transition processes with time-varying transition matrices. In addition, a mixture-of-experts approach allows us to model the probability of belonging to a certain cluster as depending on a set of covariates via a multinomial logit model. Our cluster analysis identifies five career patterns after plant closure and reveals that some workers cope quite easily with a job loss whereas others suffer large losses over extended periods of time.

Keywords: Transition data, Markov Chain Monte Carlo, Multinomial Logit, Panel data, Inhomogeneous Markov chains

1   Introduction

Long-term career outcomes after job loss due to a plant closure – where all workers are automatically displaced – are an often researched topic in labor economics, see e.g. Jacobson et al. (1993), Fallick (1996), Ruhm (1991) or more recently, for Austria, Ichino et al. (2016). Such a situation ideally allows us to observe how an economy absorbs exogenous shocks and how individuals react to perturbations to their stable career path. A plant closure has the advantage that displaced workers are neither predominantly ones who are dismissed nor those changing jobs voluntarily: a plant closure is close to an exogenous event where everybody gets displaced.

In the present paper, we study the evolution of career patterns after a job displacement due to plant closure in Austria. To observe the full recovery process after the employment shock, we follow workers over a period of 10 years. Using administrative data from social security registers, we represent career patterns by quarterly transitions between four different labor market states: being employed, sick, out of labor force, or retired. A particular focus in our analysis is on

---

[*]Institute for Statistics and Mathematics, Vienna University of Economics and Business
[†]Institute for Statistics and Mathematics, Vienna University of Economics and Business
[‡]CEU Budapest and WIFO, Vienna
[§]Department of Economics, Johannes Kepler University Linz and IHS, Vienna

heterogeneity in the career patterns due to observed or unobserved characteristics. The idea is that not all workers manage to return to stable employment paths after job displacement but that there are some types who recover faster or at slower rate and some whose career pattern changes completely.

To capture the impact of unobserved heterogeneity on transition patterns, we apply a model based clustering approach to identify cluster groups with similar career patterns after job displacement. The assumption is that the members of each group react in a common (group-specific) way within the group, but differently across groups. To identify cluster groups of workers that follow similar transition patterns in our data set, which is a collection of several thousands of discrete-valued time series, we apply model-based clustering in the spirit of Banfield and Raftery (1993); Fraley and Raftery (2002); McNicholas and Murphy (2010); Gollini and Murphy (2014), among many others. A popular method of model-based clustering of discrete-valued time series is based on separate Markov chain models for the latent subpopulations; see Frühwirth-Schnatter (2011) for a recent review. Under this approach, each subpopulation has group-specific initial and transition probabilities of the Markov chain, which distinguishes it from frailty models where a subject-specific effect for each individual is introduced (Diggle et al. (2002)).

Typically, a time-homogeneous first order Markov chain, characterized by group-specific transition matrices, is assumed as a clustering kernel, implying that the transition process within each cluster is stationary and reactions to a shock are only temporary. However, for our data the transition process is not necessarily stationary over time which poses a challenge to standard Markov chain clustering. An obvious reason for non-stationarity are the shocks to the stationary transition processes caused by an event out of the workers' control, such as job displacement. In this case, the patterns of transition during the recovery phase may differ significantly from stationary transitions and we expect that after a plant closure the intrinsically stable transition process of workers in and out of jobs might be disturbed for a period of time. Moreover, individual transitions will be shaped by changes over the life cycle – e.g. when it comes to transitions towards sick leave or retirement as workers age over time. To meet these challenges, we employ a more flexible method of Markov chain clustering, by introducing time-inhomogeneous first order Markov transition processes with time-varying transition matrices as clustering kernels.

To capture the role of observed heterogeneity and to obtain a better understanding of which workers in our data are inclined towards which career pattern, we assume that time-invariant or predetermined characteristics of a displaced worker may be correlated with group membership, i.e. persons with specific observable characteristics might be more likely to belong to a certain cluster than to the other clusters. To this aim, we follow the so-called mixture-of-experts approach introduced by Peng et al. (1996), and allow the probability of belonging to a certain subpopulation to depend on individual covariates.[1] From a statistical viewpoint, within a mixture-of-experts approach a multinomial logit model is applied to model the probability to belong to a certain cluster. In our application, these probabilities depend on pre-displacement characteristics such as the worker's age at job displacement, the years of labor market experience,

---

[1]Successful previous applications of this approach include, among many others, model-based clustering of rank data (Gormley and Murphy, 2008), model-based clustering of time series of continuous outcomes (Frühwirth-Schnatter and Kaufmann, 2008; Juárez and Steel, 2010) and model-based clustering of discrete-valued time series (Frühwirth-Schnatter et al., 2012).

the occupational type (i.e. blue versus white collar), and the income in the quarter preceding the job displacement.

To assess the effect of the job displacement shock on the change in career patterns relative to a counterfactual scenario without plant closure, we aim at comparing the post-displacement outcomes in each cluster group with a control group of workers who do not experience a plant closure. This involves the identification of a counterfactual group of non-displaced workers for each cluster group based on their unobserved characteristics. We propose a novel method to construct the counterfactual career patterns, based on the assumption that the full shock of job displacement is captured in the distribution of labor market states in the first quarter after displacement. This allows us to simulate group membership of non-displaced workers using the same clustering model that we estimated for displaced workers.

Our empirical analysis leads to the following main findings. First, applied to our sample of displaced workers, the time-varying Markov chain clustering approach identifies five distinct career patterns after plant closure. The group specific career patterns reveal a variety of different shock-absorption mechanisms, which are typically ignored in the literature. In particular, we find that almost 50% of workers cope relatively easily with job displacement, whereas others suffer considerable losses over extended periods of time. Second, modeling time-varying transition patterns is crucial in our application, as the adjustment processes show extensive variation by cluster group and over time. Third, using the time-varying mixture-of-experts Markov chain clustering approach, we find that observable characteristics are not evenly distributed across cluster groups, but individuals with different observable characteristics are found in different groups. For example, group membership strongly varies by age, occupation, or earnings prior to job displacement. Fourth, the comparison of career patterns of displaced workers with a control group of workers who do not experience a plant closure shows that – relative to the counterfactual scenario of non-displaced workers – displaced workers are less likely to be employed in the short run, but eventually employment rates of both groups converge to each other. Again, we find important heterogeneity by cluster groups.

Our paper contributes to several strands of the literature. In the field of labor economics, we contribute to the literature on the effects of job displacement and plant closures. Typical studies in this literature examine either short-term or long-term effects on employment or earnings and document that closing or downsizing plants leads to large and long-lasting effects on employment rates and earnings (Couch et al., 2010; Huttunen et al., 2011; Eliason and Storrie, 2004). Other studies investigate related outcomes such as health, mortality, fertility (Del Bono et al., 2012) or spillovers to family members (Sullivan and von Wachter, 2009). Winter-Ebmer (2016) gives a survey of recent papers. But there are no papers looking systematically at career patterns or labor market transitions and, in particular, on heterogeneous effects in such transitions. While existing studies focus on average effects of job displacement as well as effect heterogeneity by certain observable characteristics, our paper reveals an important role of unobserved heterogeneity in terms of the speed and type of labor market adjustment.

Our paper may also be instructive to the applied literature modeling transitions between discrete states over time in fields other than labor economics. Discrete transition patterns over time are of interest in many areas of applied research such as demography, finance, mathematical biology or genetics. Examples of topics to which these models are applied span a wide range: transitions between demographic states over the life cycles of individuals or households,

transitions between organisational characteristics, stock market participation or trading status of firms, changes in climate conditions across regions over time, or transitions of genetic determinants over generations of different species. These transition processes are typically captured by observations of unit-specific time series of discrete states over time.

We further contribute to the literature on finite mixtures of Markov chain modelling, where approaches similar to ours have been developed and applied in rather diverse contexts, such as clustering web-site users (Cadez et al., 2003; Dias and Vermunt, 2007), clustering sensor data from mobile robots (Ramoni et al., 2002), bond ratings migration (Frydman, 2005) or clustering employees according to their wage dynamics (Pamminger and Frühwirth-Schnatter, 2010; Pamminger and Tüchler, 2011); see also Goodman (1961) for an early discussion of the closely related mover-stayer model. Alternative, but related approaches to our clustering approach in the context of longitudinal data are based on finite mixtures of hidden Markov models (HMM), which were applied by Altman (2007) to panels of count data, by Maruotti and Rocci (2012) to panels of categorical data, by Bartolucci et al. (2014) to panels of ordinal time series and by Shirley et al. (2010) to a panel of alcohol consumption. While the hidden Markov model captures heterogeneity along the time axis, heterogeneity at the individual level is captured through an individual random effect with either a discrete (Maruotti and Rocci, 2012) or a continuous distribution (Altman, 2007). Bartolucci et al. (2014) consider an extension, where the individual random effect follows a first order Markov process. However, these approaches have some limitations in our context: While these models allow switching between different states, marginally they imply a stationary process within each cluster, whereas the time-varying Markov chain clustering approach is able to capture time-inhomogeneity in the marginal distribution also within each cluster.

The paper proceeds as follows. The next section introduces the empirical problem and the data from Austrian social security registers. Section 3 discusses the time-varying mixture-of-experts Markov chain clustering approach as well as Bayesian statistical inference. Estimation results and implications for labor market careers after job displacement are discussed in Section 4. We first comment on model selection and posterior assignment of individual cluster memberships. Then we interpret the different clusters of labor market transition processes and discuss the relationship between cluster membership and observable individual characteristics. Finally, we compare labor market trajectories of displaced workers with those of a control group of individuals who do not experience a plant closure.

2 Data Description

Our empirical analysis is based on administrative register data from the Austrian Social Security Database (ASSD), which combines detailed longitudinal information on employment and earnings of all private sector workers in Austria (Zweimüller et al., 2009). The data set includes the universe of private sector workers in Austria covered by the social security system. All employment spells record the identifier of the firm at which the worker is employed.

From the universe of employment records and employer identifiers, we can infer the characteristics of a firm's workforce at any point in time. Importantly for our application, we can observe firm entries and exits. Specifically, we define a firm's exit as the point in time when the

4

last employee leaves a firm. This is a fully data-driven definition, which in some cases identifies employer exits that do not correspond to a plant closure, for example due to a firm takeover or due to an administrative reassignment of the employer identifier. In these cases, we observe that a large group of employees continues their employment with a new identifier. To get a more precise definition of plant closure, we therefore drop an observation from the set of firm exits, if more than 50% of the employees continue under a single new employer identification number. As this method relying on worker flows does not work well for firms with high seasonal employment fluctuations, we exclude the construction and tourism sectors from our analysis. This leaves only a very small number of seasonal workers from other industries.

For the definition of our sample of displaced workers, we concentrate on all male workers employed during the years 1982 to 1988, who were experiencing a job displacement due to plant closure in this period. We do not consider female workers in the present study, because we do not have information on working time. We follow the displaced workers' detailed labor market careers for 4 years prior to job displacement and for 10 years afterwards. We further restrict the sample to workers displaced from firms that have more than 5 employees at least once during the period 1982 to 1988 and to workers who have at least one year of tenure prior to displacement. Moreover, we select workers who were between 35 and 55 years of age at the time of job displacement, leading to the analysis window being located before the official retirement age of 65 years in Austria. This procedure identifies 5,841 workers displaced by plant closures between 1982 and 1988. (Our panel is unbalanced, in the sense that we do not have 40 quarterly observations for each individual. This is due to problems with merging observations from several administrative sub-registers to create the longitudinal careers. Section 3 explains how our estimation procedure deals with unbalanced panel observations. A very small number of 320 individuals die during the observation period. Quarterly observations after death are coded as retired, which we regard as an absorbing state.)

To compare labor market careers after job loss with a counterfactual situation without job displacement, we extract a control group of workers who were employed during the years 1982 to 1988 in firms which did not close down. Our aim is to select controls who are very similar to the displaced group in terms of their pre-displacement labor market careers and observable individual characteristics. We therefore apply the following selection procedure. We start with the entire population of 1,087,705 male workers employed during the years 1982 to 1988 from which we draw a weighted sample of 5,841 workers, who are similar to the displaced group in terms of pre-displacement characteristics. Weights are constructed based on a logit regression estimating the probability of being displaced in the full set of displaced workers and potential controls (Imbens, 2004). The ASSD offers a rich set of covariates for this propensity score weighting procedure. In particular, we control for employment and earnings information in the 4 years prior to job displacement as well as age, occupational type, firm size, and industry affiliation. Sampling weights based on the logit model assure that the distribution of pre-displacement characteristics is similar among displaced and control observations.

To model employment careers we proceed by constructing a quarterly time series of labor market states for each individual. Specifically, we define the following categories: 1 denotes employed, 2 sick leave, 3 out of labor force (registered as unemployed or otherwise out of labor force), 4 retired (claiming government pension benefits). Retirement is coded as an absorbing state as virtually nobody in Austria returns to employment once he/she enters the public pension

5

system. These time series of labor market states are the basis of our empirical Markov chain clustering method.

To study characteristics that are correlated with different career patterns after job loss, we focus on variables which are pre-determined at the time of plant closure. The set of variables includes the worker's age at job displacement, the years of labor market experience, the occupational type (i.e. blue versus white collar), and the income in the quarter preceding the job displacement. Moreover, we control for firm size and industry. To capture possibly non-linear effects, we transform all these variables into discrete categories; for summary statistics see Table 1.

Worker's age (in years)		Firm's attributes	
Age 35–39	28 %	Firm size ≤ 10	42 %
Age 40–44	28 %	Firm size from 11 to 100	41 %
Age 45–49	23 %	Firm size > 100	17 %
Age 50–55	21 %	Economic sector: service	31 %
Worker's professional experience (in days)		Economic sector: industry	32 %
Experience ≤ 1675 days	33 %	Economic sector: seasonal	2 %
Experience from 1676 to 3938 days	31 %	Economic sector: unknown	35 %
Experience ≥ 3939 days	36 %		
Worker's income at time of plant closure			
Income in lowest tertile	14 %	White-collar workers	56 %
Income in middle tertile	32 %	Blue-collar workers	44 %
Income in highest tertile	54 %		

Table 1: Descriptive statistics for the control variables of all displaces persons in the mixture-of-experts model to explain group membership.

3 Time-varying Mixture-of-Experts Markov Chain Clustering

As for many data sets available for empirical labor market research, the structure of the individual level transition data introduced in Section 2 takes the form of a discrete-valued panel data. The categorical outcome variable y_{it} assumes one out of four states, labeled by $\{1, 2, 3, 4\}$, and is observed for N individuals $i = 1, \ldots, N$ over T_i quarters for a maximum of 10 years, i.e. $T_i \leq 40$ quarters. Moreover, we restrict ourselves to $T_i \geq 4$. For each individual i, we model the state of the outcome variable y_{it} in period t to depend on the past state $y_{i,t-1}$ through a time-inhomogeneous first order Markov transition model.

To capture the presence of unobserved heterogeneity in the dynamics in our discrete-valued panel data, we apply model-based clustering based on Markov transition models. The central assumption in model-based clustering is that the N time series in the panel arise from H hidden classes; see Frühwirth-Schnatter (2011) for a review. Within each class, say h, all time series can be characterized by the same data generating mechanism, called a clustering kernel, which is defined in terms of a probability distribution for the time series $\mathbf{y}_i = \{y_{i1}, \ldots, y_{i,T_i}\}$, depending on an unknown class-specific parameter $\boldsymbol{\vartheta}_h$. A latent cluster indicator S_i taking a value in the set $\{1, \ldots, H\}$ is introduced for each time series $\mathbf{y}_i$ to indicate which class the individual i

belongs to, i.e. $p(\mathbf{y}_i|S_i, \boldsymbol{\vartheta}_1, \ldots, \boldsymbol{\vartheta}_H) = p(\mathbf{y}_i|\boldsymbol{\vartheta}_{S_i})$. Note that we cluster the entire employment profile jointly; hence S_i is constant for each worker over the entire observation period.

To address serial dependence among the observations for each individual i, model-based clustering of time series data is typically based on dynamic clustering kernels derived from first order Markov processes, where the clustering kernel $p(\mathbf{y}_i|\boldsymbol{\vartheta}_h) = \prod_{t=1}^{T_i} p(y_{it}|y_{i,t-1}, \boldsymbol{\vartheta}_h)$ is formulated conditional on the initial state y_{i0}, which in our application is equal to 1 (employed) for all individuals. For discrete-valued time series, persistence is typically captured by assuming that $\mathbf{y}_i$ follows a time-homogeneous Markov chain of order one. Applications of time-homogeneous Markov Chain clustering to analyze individual wage careers in the Austrian labor market include Pamminger and Frühwirth-Schnatter (2010), Pamminger and Tüchler (2011), and Frühwirth-Schnatter et al. (2012).

However, the assumption that the long-run career paths of workers who experienced plant closure follow a time-homogeneous Markov chain is not realistic (see Ichino et al. (2016), Figure 2). A descriptive investigation of the evolution of the employment rate over time after plant closure reveals that the employment rate does not converge to a steady state, but rather declines steadily with distance from plant closure. Homogeneity would imply that all state probabilities, including the employment rate, converge to a steady state, both within each cluster as well as marginalized over all clusters. To obtain a non-stationary pattern, we need to assume that the transition probabilities between the various states change with distance from plant closure. Furthermore, it is to be expected that there is a lot of heterogeneity in this time-varying pattern across workers.

To capture this non-stationary feature of our data, we apply Markov chain clustering based on a time-inhomogeneous first order Markov chain model with class-specific time-varying transition matrices $\boldsymbol{\vartheta}_h = (\boldsymbol{\pi}_h, \boldsymbol{\xi}_{h1}, \ldots, \boldsymbol{\xi}_{h,10})$ as clustering kernel. More specifically, we assume that the transition behavior changes with distance from plant closure. Since the initial state is employment (i.e. $y_{i0} = 1$) for all workers, the first transition is described by the row vector $\boldsymbol{\pi}_h = (\pi_{h,1}, \ldots, \pi_{h,4})$, containing the cluster-specific probability distribution of y_{i1}, the state at the end of the first quarter after plant closure. The transition matrix $\boldsymbol{\xi}_{h1}$ describes the transition behavior between the various states in quarter two to four after plant closure, while the remaining transition matrices $\boldsymbol{\xi}_{hy}$, $y = 2, \ldots, 10$, describe the transition behavior for all four quarters in year y after plant closure. Since the fourth state, namely retirement, is an absorbing state, each of these time-varying transition matrices $\boldsymbol{\xi}_{hy}$ consists of three rows $\boldsymbol{\xi}_{hy,j\cdot} = (\xi_{hy,j1}, \ldots, \xi_{hy,j4})$, $j = 1, 2, 3$, representing a probability distribution over the states $\{1, 2, 3, 4\}$, i.e. $\sum_{k=1}^{4} \xi_{hy,jk} = 1$. Hence the clustering kernel reads:

$$p(\mathbf{y}_i|\boldsymbol{\vartheta}_h) = p(\mathbf{y}_{i,-1}|y_{i1}, \boldsymbol{\xi}_{h1}, \ldots, \boldsymbol{\xi}_{h,10}) p(y_{i1}|S_i = h, \boldsymbol{\pi}_h), \tag{1}$$

where $\mathbf{y}_{i,-1} = \{y_{i2}, \ldots, y_{i,T_i}\}$ denotes the truncated time series, excluding state y_{i1}. The distribution of $\mathbf{y}_{i,-1}$ is given by a sequence of time-varying transition matrices changing every year:

$$p(\mathbf{y}_{i,-1}|y_{i1}, \boldsymbol{\xi}_{h1}, \ldots, \boldsymbol{\xi}_{h,10}) = \prod_{y=1}^{10} \prod_{j=1}^{3} \prod_{k=1}^{4} \xi_{hy,jk}^{N_{iy,jk}}, \tag{2}$$

with transition probabilities $\xi_{h1,jk} = \Pr(y_{it} = k | y_{i,t-1} = j, S_i = h, t \in \{2,3,4\})$, and $\xi_{hy,jk} = \Pr(y_{it} = k | y_{i,t-1} = j, S_i = h, t \in \{4(y-1)+1, \ldots, 4y\})$ for $y = 2, \ldots, 10$. For each time series $\mathbf{y}_{i,-1}$, the cluster-specific sampling distribution (2) depends on the number of transitions from state j to state k observed in each year, i.e. $N_{i1,jk} = \#\{y_{i,t-1} = j, y_{it} = k | t \in \{2,3,4\}\}$ and $N_{iy,jk} = \#\{y_{i,t-1} = j, y_{it} = k | t \in \{4(y-1)+1, \ldots, 4y\}\}$ for $y = 2, \ldots, 10$. If $T_i < 40$, then all transition counts are zero for all unobserved quarters.

The choice of the distribution for the state y_{i1} at the end of the first quarter in (1) has to address the problem with initial conditions in non-linear dynamic models with unobserved heterogeneity, see e.g. Heckman (1981) and Wooldridge (2005). Skrondal and Rabe-Hesketh (2014) provides an excellent review of different approaches in the context of dynamic/transition models for binary data with unobserved heterogeneity where this problem was first discussed by Aitkin and Alfó (1998). This strand of literature focuses on the case where unobserved heterogeneity is captured through an individual random effect following a continuous distribution. However, as discussed in Frühwirth-Schnatter et al. (2012), this problem is also relevant for dynamic models where S_i follows a discrete distribution; hence in particular for model-based clustering based on transition models. As for the continuous case, the key issue is to allow for dependence between the initial state y_{i1} and the discrete-valued latent variable S_i. In Frühwirth-Schnatter et al. (2012), this dependence has been achieved by allowing the prior distribution of S_i to depend on y_{i1}. In the present paper, we suggest an alternative approach, based on factorizing the joint distribution of y_{i1} and S_i as $p(y_{i1}, S_i | \cdot) = p(y_{i1} | S_i, \cdot) p(S_i | \cdot)$, where the state distribution $p(y_{i1} | S_i, \cdot)$ in the first quarter after plant closure is allowed to be different across the clusters:

$$p(y_{i1} | S_i = h, \boldsymbol{\pi}_h) = \prod_{k=1}^{4} \pi_{h,k}^{I_{i,k}}, \tag{3}$$

where $\pi_{h,k} = \Pr(y_{i1} = k | S_i = h)$, $\boldsymbol{\pi}_h = (\pi_{h,1}, \ldots, \pi_{h,4})$, and $I_{i,k} = I\{y_{i1} = k\}$ is an indicator for a worker's state at the end of the first quarter after plant closure.

Following the mixture-of-experts approach introduced for Markov chain clustering methods by Frühwirth-Schnatter et al. (2012), the distribution $p(S_i | \cdot)$ is influenced by exogenous covariates $\mathbf{X} = (X_1, \ldots, X_r)$. For each individual i, exogenous covariates $(x_{i1}, \ldots, x_{ir})$ observed for $\mathbf{X}$ influence the probability to belong to a certain cluster through a multinomial logit (MNL) model:

$$\Pr(S_i = h | \boldsymbol{\beta}_2, \ldots, \boldsymbol{\beta}_H, \mathbf{X} = \mathbf{x}_i) = \frac{\exp(\mathbf{x}_i \boldsymbol{\beta}_h)}{1 + \sum_{l=2}^{H} \exp(\mathbf{x}_i \boldsymbol{\beta}_l)}, \quad h = 1, \ldots, H, \tag{4}$$

where the row vector $\mathbf{x}_i = (1, x_{i1}, \ldots, x_{ir})$ includes a constant intercept in addition to the exogenous covariates. For identifiability reasons $\boldsymbol{\beta}_1 = \mathbf{0}$, which means that $h = 1$ is the baseline class and $\boldsymbol{\beta}_h$ is the effect on the log-odds ratio relative to the baseline.

For estimation, we pursue a Bayesian approach. For a fixed number H of clusters, Markov chain Monte Carlo (MCMC) methods are used, to estimate the latent cluster indicators $\mathbf{S} = (S_1, \ldots, S_N)$ along with the unknown cluster-specific parameters $\boldsymbol{\theta}_H = (\boldsymbol{\vartheta}_1, \ldots, \boldsymbol{\vartheta}_H, \boldsymbol{\beta}_2, \ldots, \boldsymbol{\beta}_H)$ from the data $\mathbf{y} = (\mathbf{y}_1, \ldots, \mathbf{y}_N)$. To sample from the posterior distribution $p(\boldsymbol{\theta}_H, \mathbf{S} | \mathbf{y})$, we extend the sampler introduced in Frühwirth-Schnatter et al. (2012) to time-inhomogeneous mixture-of-experts Markov chain clustering; see Appendix A for computational details.

8

4 Analysing Plant Closure Effects

To identify clusters of individuals with similar career patterns after plant closure, we apply Markov chain clustering for 2 up to 6 clusters. All computations are based on the prior distributions introduced in Appendix A. For each number H of clusters we simulate 15 000 MCMC draws after a burn-in of 10 000 draws and use them for all posterior inference reported below.[2]

In the following, we start with a description of model selection and posterior classification. Second, we discuss the cluster-specific post-displacement career patterns that are implied by the estimated transition processes. Third, we describe the correlation between cluster membership and workers' characteristics. Finally, we compare the career paths of the displaced workers with workers in the control group who did not experience a job loss.

4.1 Model Selection

H	2	3	4	5	6
AIC	112160.9	110381.0	109113.4	**107567.6**	108057.0
BIC	113575.5	112549.6	112036.0	**111244.2**	112487.7
AWE	114402.1	114188.3	**114159.6**	114539.8	116356.4

Table 2: Various statistical criteria to select the number H of clusters.

Statistical model selection criteria such as AIC, BIC and the AWE criterion, as discussed e.g. in Frühwirth-Schnatter (2011), are applied to select the number H of clusters, see Table 2. As in previous studies (e.g. Pamminger and Frühwirth-Schnatter (2010); Frühwirth-Schnatter et al. (2012)), these statistical criteria are ambiguous and do not give a clear answer, with AIC and BIC supporting a five cluster solution, while AWE selects a model with four clusters. In addition, following another related study (Frühwirth-Schnatter et al., 2016), we found it useful to select H such that the resulting clusters are sufficiently distinct, both in statistical terms as well as in terms of allowing a meaningful economic interpretation.

As explained in detail in Subsection 4.4, we can conveniently interpret five distinct clusters of career patterns, which are characterized by a combination of mobility/persistence and attachment to the labor force – i.e. employment levels: cluster Low-attached and cluster Highly attached are characterized by low and high levels of attachment to the labor market, respectively, with high persistence in the corresponding states; cluster Mobile + low-attached and cluster Mobile + highly attached are characterized by a much higher level of mobility together with low and high levels of attachment to the labor market, respectively; and, finally, in cluster Retiring retirement is the predominant state. In a six-cluster model, the distinction between the different clusters is less clear. On the other hand, cluster solutions with less than five clusters lead to clusters that were still rather inhomogeneous in terms of the career patterns allocated to these clusters. Therefore, we concentrate in the following on the five cluster solution

---

[2]The computing time for all 25 000 draws is approx. 15 minutes for $H = 2$, 1 hour and 2 minutes for $H = 3$, 1 hour and 33 minutes for $H = 4$, 2 hours and 21 minutes for $H = 5$ and 4 hours and 45 minutes for $H = 6$ on a Lenovo Thinkpad T410s laptop equipped with 4 GB RAM and an Intel Core i5 processor with 2.67 GHz.

chosen by AIC and BIC, because this solution also leads to meaningful interpretations from an economic point of view.
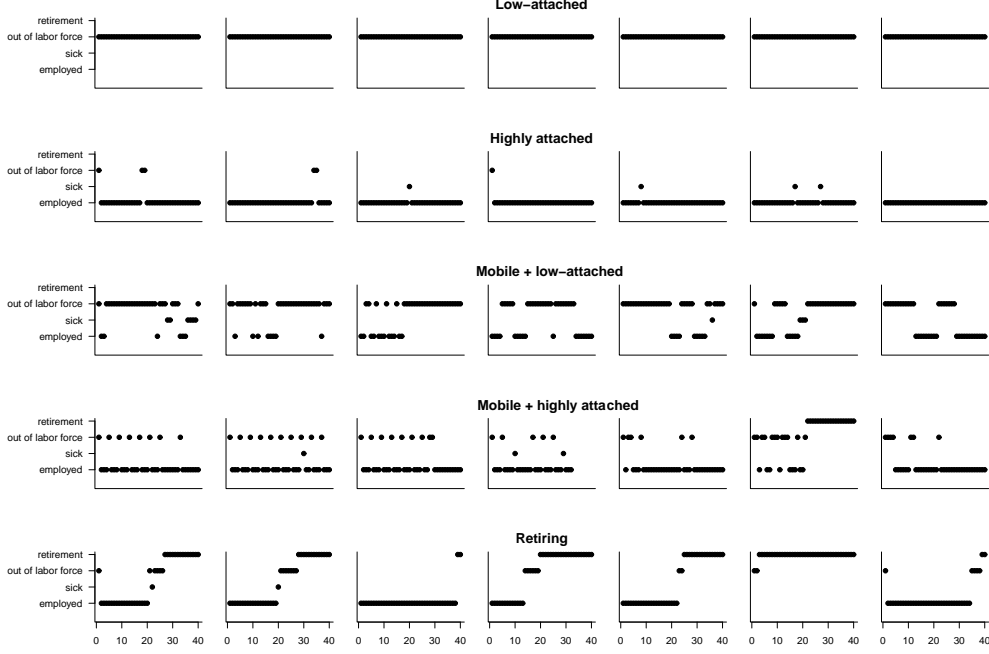

Figure 1: Employment profiles of typical cluster members within each cluster, showing the 10th, 25th, 50th, 70th, 100th, 200th and 350th highest classification probabilities.

4.2 Posterior Classification

As mentioned at the end of Section 3, we estimate the latent cluster indicators $\mathbf{S} = (S_1, \ldots, S_N)$ jointly with the unknown cluster-specific parameters $\boldsymbol{\theta}_H = (\boldsymbol{\vartheta}_1, \ldots, \boldsymbol{\vartheta}_H, \boldsymbol{\beta}_2, \ldots, \boldsymbol{\beta}_H)$, by sampling from the posterior distribution $p(\boldsymbol{\theta}_H, \mathbf{S}|\mathbf{y})$. Parameter estimation is then based on the marginal posterior distribution $p(\boldsymbol{\theta}_H|\mathbf{y})$ which is integrated over the unknown latent cluster indicators $\mathbf{S}$. Within full conditional Gibbs sampling, soft clustering is performed implicitly (see classification rule (8) in Appendix A) and each worker impacts the estimates of all cluster-specific transition matrices, weighted according to his probability to belong to a certain cluster.

To obtain a first understanding of the transition patterns in the various clusters, the posterior draws are post-processed and hard clustering is performed for all individuals. Individuals are assigned to the five clusters of career-patterns using the posterior classification probabilities $t_{ih}(\boldsymbol{\theta}_5) = \Pr(S_i = h|\mathbf{y}_i, \mathbf{X} = \mathbf{x}_i, \boldsymbol{\theta}_5)$ given by eq. (8) in Appendix A. The posterior expectation $\hat{t}_{ih} = \mathrm{E}(t_{ih}(\boldsymbol{\theta}_5)|\mathbf{y})$ of these probabilities is estimated by evaluating and averaging $t_{ih}(\boldsymbol{\theta}_5)$ over all MCMC draws of $\boldsymbol{\theta}_5$. Each worker is then allocated to that cluster $\hat{S}_i$, which exhibits the maximum posterior probability, i.e. $\hat{S}_i$ is defined such that $\hat{t}_{i,\hat{S}_i} = \max_h \hat{t}_{ih}$. This decision rules minimizes the misclassification risk for each worker, see e.g. Frühwirth-Schnatter (2006, Section 7.1). The closer $\hat{t}_{i,\hat{S}_i}$ is to 1, the higher is the segmentation power for individual i.

10

Typical group members are visualized in Figure 1 for each cluster through their individual time series. The career patterns in Figure 1 are fairly similar within each cluster, but very different across clusters.
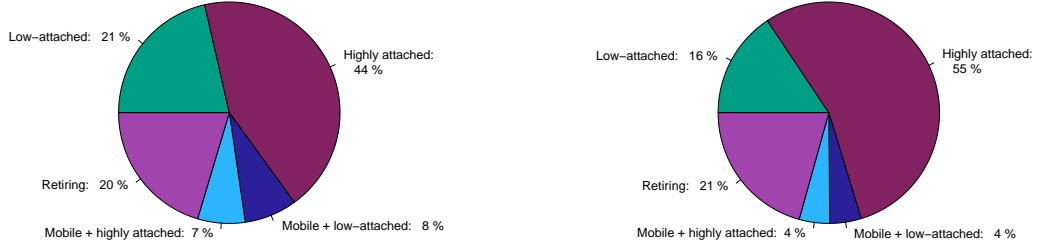

Figure 2: Group sizes for the five cluster solution. The cluster sizes are calculated based on the posterior classification probabilities. Left hand side: workers experiencing plant closure (displaced); right hand side: workers from the control group not experiencing plant closure

Based on the posterior classification probabilities of cluster membership for each of the N workers, we compute the average size of each cluster. The corresponding shares of individuals in each cluster are shown in the left hand graph of Figure 2. The displaced workers in our sample are relatively unevenly distributed across the five clusters: 21% of the persons belong to cluster LOW-ATTACHED, 44% to cluster HIGHLY ATTACHED, 8% to cluster MOBILE + LOW-ATTACHED, 7% to cluster MOBILE + HIGHLY ATTACHED, and 20% to cluster RETIRING.

4.3 Analyzing Career Mobility

To analyze career mobility patterns in the five different clusters we investigate for each cluster the posterior distribution of the time-varying cluster-specific transition matrices $\boldsymbol{\vartheta}_h = (\boldsymbol{\pi}_h, \boldsymbol{\xi}_{h1}, \ldots, \boldsymbol{\xi}_{h,10})$ for $h = 1, \ldots, 5$. For all workers in our sample, the transition process starts with the shock of job displacement due to plant closure. Thus the vector $\boldsymbol{\pi}_h$ defines, for each cluster, the worker's state distribution $\boldsymbol{\pi}_{h,1} = \boldsymbol{\pi}_h$ at the end of the first quarter after plant closure. The corresponding posterior expectation $\mathrm{E}(\boldsymbol{\pi}_{h,1}|\mathbf{y})$ is shown for each cluster in Figure 3 at $t = 1$.

The time-varying cluster-specific transition matrices are visualized in Figure 4, with each of the five rows corresponding to a specific cluster. The four columns of Figure 4 correspond to transition probabilities of particular interest, namely persistence in the employment state (i.e. $j = 1 \to k = 1$), transition from employment to out of labor force (i.e. $j = 1 \to k = 3$), transition from out of labor force back to employment (i.e. $j = 3 \to k = 1$), and transition from employment to retirement (i.e. $j = 1 \to k = 4$). Each single plot in Figure 4 shows how the posterior distribution of the transition probability $\xi_{hy,jk}$ from $j \to k$ changes in cluster h over time as the yearly distance from plant closure $y = 1, \ldots, 10$ increases. Note that each posterior distribution $p(\xi_{hy,jk}|\mathbf{y})$ is represented by box plots of the corresponding MCMC draws.

11

Furthermore, numerical estimates and standard deviations for the initial distribution $\boldsymbol{\pi}_h$ as well as the above selected transition probabilities $\xi_{hy,jk}$ are reported in Table 3.

To evaluate the long-term effect of the job loss experienced by all workers, the state distribution $\boldsymbol{\pi}_{h,t}$ was computed also for all subsequent quarters $t = 2, \ldots, 40$, individually for each cluster. Given the distribution of states at the end of the first quarter, described by $\boldsymbol{\pi}_h$, each state distribution $\boldsymbol{\pi}_{h,t}$ is computed by taking into account that the transition process evolves according to a time-inhomogenous Markov process:

$$\boldsymbol{\pi}_{h,t} = \boldsymbol{\pi}_h \boldsymbol{\xi}_{h,1 \to t}, \quad h = 1, \ldots, H. \tag{5}$$

Starting from $\boldsymbol{\xi}_{h,1 \to 2} := \boldsymbol{\xi}_{h1}$, the transition matrix $\boldsymbol{\xi}_{h,1 \to t}$ from the first to the qth quarter in year y, i.e. $t = 4(y-1)+q$, can be computed for $t = 3, \ldots, 40$ recursively from the sequence of cluster-specific time-inhomogenous transition matrices through $\boldsymbol{\xi}_{h,1 \to t} = \boldsymbol{\xi}_{h,1 \to (t-1)} \boldsymbol{\xi}_{hy}$. Figure 3 shows the evolution of the posterior expectations $\mathrm{E}(\boldsymbol{\pi}_{h,t}|\mathbf{y})$ of the cluster-specific state distribution over distance t from plant closure.[3]

4.4 Understanding the Clusters

In this subsection we present a synthesis of posterior inference in Figure 1 to Figure 4 and Table 3 and interpret the estimated transition processes after job displacement for the different clusters. The figures highlight remarkable differences across clusters in the state distribution at the end of the first quarter, as well as in the subsequent transition patterns. We will now discuss these career patterns cluster by cluster.

Cluster HIGHLY ATTACHED is the largest one with about 44% of the observations. Workers in this cluster have a relatively high probability to be employed again within one quarter after plant closure (63%), whereas this probability is considerably smaller for all other clusters. Only 35.9% of the cluster members are still out of labor force one quarter after plant closure. For workers in this cluster, the probability to remain employed is close to 1 over the whole 10 years (98.9% five and 97.8% ten years after plant closure). As a consequence, for workers in this cluster the risk of another job loss is very small (0.7% five and 1.5% ten years after plant closure). In the unlikely event that these workers lose their job, they have quite a good chance to move back into employment within one quarter, however, with increasing distance from plant closure, this chance declines and is as small as 7.1% after 10 years.

Workers in cluster LOW-ATTACHED, the second largest one containing about 21% of the sample, are less successful than workers in cluster HIGHLY ATTACHED in finding a new job in the first quarter after plant closure (only about 30%) and the majority (68.4%) are still out of labor force. Similar numbers can be observed for workers in cluster MOBILE + LOW-ATTACHED and cluster MOBILE + HIGHLY ATTACHED. Overall, workers in these three clusters suffer from the plant closure at least in the short run. However, what distinguishes workers in cluster LOW-ATTACHED from workers in the other two clusters is the subsequent transition behavior. Most strikingly, among workers in cluster LOW-ATTACHED the chance of moving from out of labor force back into employment is extremely low in the years following plant closure and even decreases,

---

[3]The posterior expectation is estimated by computing $\boldsymbol{\pi}_{h,t}$ for $t = 1, \ldots, 40$ for all 15 000 MCMC draws and averaging the resulting draws of $\boldsymbol{\pi}_{h,t}$ for each quarter t and each cluster h.

being equal to only 1.3% five and 1% ten years after plant closure. Members of this cluster hardly ever move back into employment after having lost their job due to plant closure and suffer from plant closure also in the long run.

While the clusters MOBILE + LOW-ATTACHED and MOBILE + HIGHLY ATTACHED are similar to cluster LOW-ATTACHED in the short-run after plant closure, they different from this cluster substantially in their subsequent transition pattern between out of labor force and employment. Workers in these two clusters recover more easily from job displacement and have about the same probability of remaining employed, which is nearly constant over time and, on average, equal to 82%. They have a similar transition pattern from employment back to out of the labor force, which again is nearly constant over time and is, on average, equal to about 15%. Obviously, members in these two clusters have a good chance to move back into the labor market after plant closure, but they are at a high risk to lose their job again. Workers in these two clusters which are characterized by frequent switches between employment and being out of labor force suffer from an intrinsically high risk of being out of labor force that appears to be unrelated to plant closure.

The main distinction between cluster MOBILE + HIGHLY ATTACHED and cluster MOBILE + LOW-ATTACHED is the transition pattern from out of labor force back into employment and how it evolves with distance from plant closure. This difference leads to career paths that are quite distinctive. For workers in cluster MOBILE + HIGHLY ATTACHED, the chance of moving back into the labor market is higher than in the other cluster and even increases in the first five years after plant closure. The corresponding transition probability is as large as 74% five years and still equal to 54% ten years after plant closure. This leads to career patterns that are characterized by frequent transitions between employment and out of labor force, see also some typical members of this cluster in Figure 1. For cluster MOBILE + LOW-ATTACHED, the transition probability from out of labor force back into employment is much smaller and declines, being only 15% five years and as small as 7.8% ten years after plant closure. Workers in both clusters switch between employment and being out of labor force; however, workers in cluster MOBILE + LOW-ATTACHED have a much higher risk to remain out of labor force. As a consequence, this leads to much longer spells of being out of labor force than for workers in cluster MOBILE + HIGHLY ATTACHED, where this duration is very short, see again Figure 1.

Finally, workers in cluster RETIRING are less successful than workers in cluster HIGHLY ATTACHED to find a job in the first quarter after plant closure (42.2%), but more successful than workers in the other clusters. In cluster RETIRING immediate transition into retirement after plant closure happens with positive probability (2.7%), whereas this probability is practically zero for all other clusters. Workers in this cluster also have a much higher risk (10.1%) to be on sick leave immediately after plant closure. In addition, we find an increasing transition probability from employment into retirement which is as large as 18.7% ten years after plant closure, whereas this probability practically remains zero for all other clusters. As a consequence, the probability to remain employed, which is relatively high in the first years after plant closure, declines in later years and is the smallest among all clusters (72.2%) after 10 years.
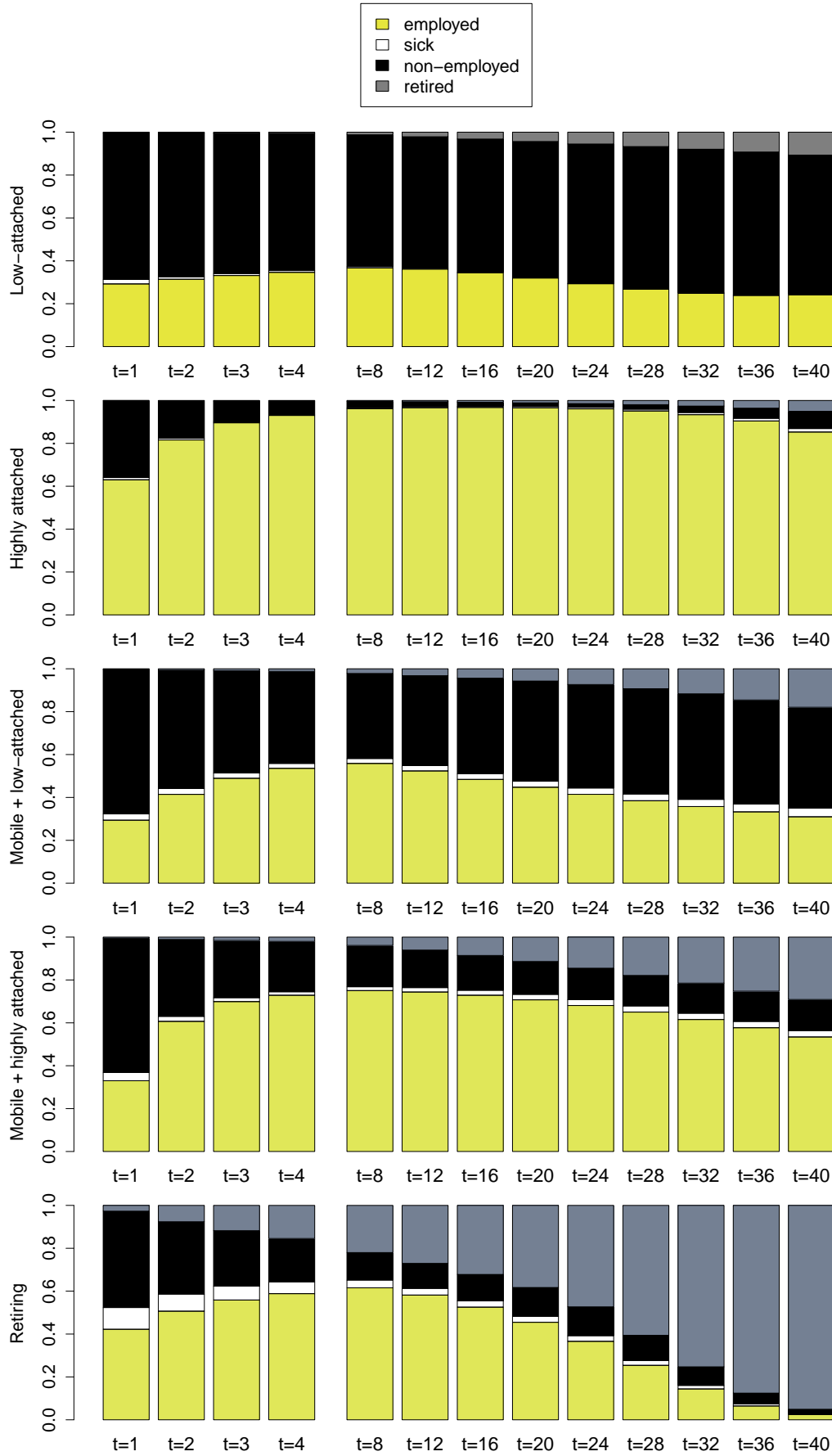
Figure 3: Posterior expectation of the distribution $\boldsymbol{\pi}_{h,t}$ over the 4 states ($1 = $ employed, $2 = $ sick leave, $3 = $ out of labor force, $4 = $ retired) after a period of t quarters in the various clusters (workers experiencing plant closure).
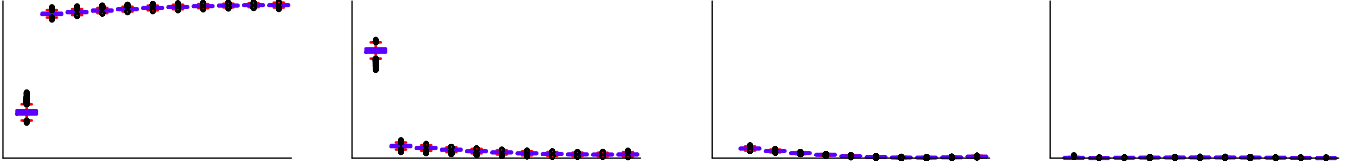
Figure 4: Visualization of the posterior distribution of four selected time-varying transition probabilities from state j to state k in the various clusters, with each row corresponding to a specific cluster. The first box plot in columns 1, 2 and 4 displays the posterior distribution of the state probability $\pi_{h,k}$ at the end of the first quarter after plant closure for each cluster h. The remaining ten box plots display the posterior distribution of the transition probabilities $\xi_{hy,jk}$ over the years $y = 1, 2, \ldots, 10$ for each cluster h. $1 =$ employed, $2 =$ sick leave, $3 =$ out of labor force, $4 =$ retired.

The importance of using a time-inhomogeneous rather than a time-homogeneous Markov chain clustering method for our application can be best seen in Figure 3, which shows for each cluster how the state distribution evolves over time. The largest changes can be seen in the clusters RETIRING and MOBILE + LOW-ATTACHED, which is due to the varying importance of the states employment and retirement. The inhomogeneous modeling approach deals with such non-linear patterns in a very flexible way. Our time series data, where a stable equilibrium process is shocked by a plant closure, require flexibility in particular at the beginning. The importance of allowing for a separate transition process in the first quarter can clearly be seen in the large turbulence in the first year in Figure 3.

h	$\pi_{h,1}$	$\pi_{h,2}$	$\pi_{h,3}$	$\pi_{h,4}$
LOW-ATTACHED	0.292 (0.021)	0.021 (0.005)	0.684 (0.022)	0.002 (0.002)
HIGHLY ATTACHED	0.630 (0.011)	0.010 (0.002)	0.359 (0.011)	0.001 (0.001)
MOBILE + LOW-ATTACHED	0.294 (0.026)	0.030 (0.010)	0.672 (0.028)	0.003 (0.004)
MOBILE + HIGHLY ATTACHED	0.330 (0.026)	0.038 (0.012)	0.627 (0.027)	0.005 (0.006)
RETIRING	0.422 (0.016)	0.101 (0.009)	0.449 (0.016)	0.027 (0.005)

year y	$j=1 \to k=1$	$j=1 \to k=3$	$j=3 \to k=1$	$j=1 \to k=4$
	LOW-ATTACHED			
$y = 1$	0.918 (0.009)	0.077 (0.009)	0.062 (0.005)	0.001 (0.001)
$y = 5$	0.956 (0.006)	0.037 (0.005)	0.013 (0.002)	0.005 (0.001)
$y = 10$	0.974 (0.006)	0.024 (0.005)	0.010 (0.002)	0.000 (0.000)
	HIGHLY ATTACHED			
$y = 1$	0.978 (0.002)	0.019 (0.001)	0.545 (0.022)	0.000 (0.000)
$y = 5$	0.989 (0.001)	0.007 (0.001)	0.416 (0.022)	0.000 (0.000)
$y = 10$	0.978 (0.001)	0.015 (0.001)	0.071 (0.019)	0.000 (0.000)
	MOBILE + LOW-ATTACHED			
$y = 1$	0.860 (0.014)	0.130 (0.013)	0.232 (0.020)	0.001 (0.001)
$y = 5$	0.817 (0.012)	0.158 (0.010)	0.154 (0.013)	0.001 (0.001)
$y = 10$	0.856 (0.018)	0.117 (0.016)	0.078 (0.009)	0.001 (0.001)
	MOBILE + HIGHLY ATTACHED			
$y = 1$	0.841 (0.012)	0.146 (0.012)	0.506 (0.024)	0.003 (0.001)
$y = 5$	0.821 (0.008)	0.158 (0.007)	0.740 (0.019)	0.003 (0.001)
$y = 10$	0.822 (0.013)	0.146 (0.011)	0.540 (0.037)	0.005 (0.002)
	RETIRING			
$y = 1$	0.938 (0.007)	0.021 (0.004)	0.221 (0.012)	0.021 (0.005)
$y = 5$	0.955 (0.004)	0.024 (0.003)	0.011 (0.003)	0.000 (0.000)
$y = 10$	0.722 (0.031)	0.052 (0.012)	0.040 (0.009)	0.187 (0.027)

Table 3: Posterior expectations $E(\pi_{h,k}|\mathbf{y})$ and, in parenthesis, posterior standard deviations $SD(\pi_{h,k}|\mathbf{y})$ of the state probability $\pi_{h,k}$ at the end of the first quarter after plant closure for all states $k = 1, \ldots, 4$ as well as posterior expectations $E(\xi_{hy,jk}|\mathbf{y})$ and, in parenthesis, posterior standard deviations $SD(\xi_{hy,jk}|\mathbf{y})$ of selected transition probabilities $\xi_{hy,jk}$ for selected years y in the various clusters. 1 = employed, 2 = sick leave, 3 = out of labor force, 4 = retired.

4.5 The Impact of Observables on Group Membership

After having established differences in labor market careers following plant closure across five different clusters of workers, we now investigate how individual characteristics relate to cluster membership. From a social policy point of view, it is interesting to understand if the characteristics of a particular worker make him more prone to belong to a specific cluster. In particular, we would like to answer questions such as: Is the career adjustment after plant closure easier for younger workers than for older workers? Who might be forced into early retirement? Do blue collar workers have a higher risk to belong to the most disadvantaged cluster LOW-ATTACHED than white collar workers?

The mixture-of-experts approach allows to answer these and similar questions, since we specify the probability of an individual to belong to a certain cluster by the multinomial logit (MNL) model given in equation (4). The regression framework flexibly controls for the impact of six covariates in the MNL model, namely age at the time of plant closure, experience, broad occupational status (i.e. blue versus white collar), income, firm size, and the economic sector, each with dummy coding. More specifically, we introduce five age groups (35-39, 40-44, 45-49, 50-55), three levels of experience (low, medium, high), a dummy for white-collar workers, three levels of income before plant closure (low, medium, high) based on the tertiles of the general income distribution at time of plant closure, three categories of firm size (1-10, 11-100, and more than 100 employees), and four broad economic sectors (service, industry, remaining seasonal business (outside of hotel and construction), unknown); see also Table 1. Alternatively, it would be possible to include all continuous covariates in the mixture-of-experts approach without discretization, as exemplified by a related paper on mothers' long-run career patterns after first birth (Frühwirth-Schnatter et al., 2016) which uses a time-homogeneous mixture-of-experts Markov chain clustering approach.

Bayesian inference for the regression parameters $\boldsymbol{\beta}_h$ in the MNL model (4) is summarized in Table 4, which reports the posterior expectation and the posterior standard deviation of all regression parameters relative to the baseline, which is equal to cluster LOW-ATTACHED.

To visualize the main results, Figure 5 shows to which extent the probability of belonging to each of the five clusters is related to each individual covariate X_j; see also Table 5. For this evaluation, all covariates in $\mathbf{X}$ apart from X_j are set to their mean values observed in the sample. The probability $\Pr(S_i = h | \boldsymbol{\beta}_2, \ldots, \boldsymbol{\beta}_H, \mathbf{X})$ that a worker with certain predetermined characteristics $\mathbf{X}$ belongs to cluster h is computed for all MCMC draws and the reported values are averages over all MCMC draws. Since the probabilities $\Pr(S_i = h | \boldsymbol{\beta}_2, \ldots, \boldsymbol{\beta}_H, \mathbf{X})$ act as a "prior" probabilities in the Bayes' classification rule (8), as outlined in Appendix A, the various diagrams in Figure 5 can be interpreted as providing the prior probability that a worker belongs to any of the five clusters based solely on characteristics $\mathbf{X}$ known before plant closure.

A worker's broad occupational status is highly related to cluster membership; see Figure 5, panel (a), as well as Table 5. Most strikingly, blue collar workers have about half the risk of white collar workers (18% versus 41%) to belong to cluster LOW-ATTACHED, which suffers most from plant closure. This is a specific feature of plant closure events, see also Schwerdt et al. (2010). Not surprisingly, white collar workers have a small prior probability to belong to cluster MOBILE + HIGHLY ATTACHED (4%).

With respect to age at the time of plant closure, we see in Figure 5, panel (b), as well as in

	Highly attached	Mobile + low-attached	Mobile + highly attached	Retiring
Intercept	-1.522 (0.177)	-0.762 (0.249)	-3.002 (0.261)	-4.114 (0.294)
Age 35–39 (basis)				
Age 40–44	0.220 (0.106)	0.334 (0.163)	0.201 (0.175)	0.307 (0.323)
Age 45–49	0.061 (0.118)	0.160 (0.186)	0.001 (0.196)	2.398 (0.246)
Age 50–55	-2.740 (0.388)	-0.988 (0.436)	0.725 (0.236)	4.410 (0.249)
Experience $\leq$ 1675 days (basis)				
Experience from				
1676 to 3938 days	0.404 (0.107)	-0.687 (0.163)	-0.318 (0.164)	-0.010 (0.172)
Experience $\geq$ 3939 days	0.687 (0.108)	-0.891 (0.190)	-0.490 (0.176)	0.272 (0.163)
Blue collar	1.045 (0.111)	0.665 (0.183)	2.020 (0.179)	1.212 (0.166)
Income in lowest tertile (basis)				
Income in middle tertile	1.235 (0.156)	-0.134 (0.197)	0.469 (0.191)	0.274 (0.202)
Income in highest tertile	1.146 (0.153)	-0.352 (0.186)	-0.334 (0.213)	0.022 (0.201)
Firm size $\leq$ 10 (basis)				
Firm size from 11 to 100	0.701 (0.100)	0.163 (0.159)	0.578 (0.155)	0.787 (0.157)
Firm size > 100	0.617 (0.142)	-0.761 (0.286)	-0.002 (0.233)	0.941 (0.190)
Economic sector: service (basis)				
Economic sector: industry	0.368 (0.114)	0.314 (0.173)	0.785 (0.193)	0.253 (0.173)
Economic sector: seasonal	-0.224 (0.318)	-0.065 (0.490)	0.588 (0.534)	0.282 (0.465)
Economic sector: unknown	0.188 (0.103)	-0.110 (0.164)	1.017 (0.179)	0.542 (0.165)

Table 4: Multinomial logit model to explain cluster membership in a particular cluster (baseline: Low-attached); the numbers are the posterior expectation and, in parenthesis, the posterior standard deviation of the various regression coefficients.

Table 5 that workers younger than 45 years have similar probabilities to belong to the various clusters. In particular, their probability to belong to cluster Retiring is low, but this probability strongly increases with age. Individuals with higher age more often belong to cluster Retiring, and this probability is particularly high (77%) for the oldest group, aged 50-55. At the same time, the probability of being in cluster Highly attached reduces with age and is negligible for the oldest age group. The probability to belong to cluster Mobile + highly attached is practically independent of age and the probability of belonging to cluster Low-attached is slightly decreasing with age.

Work experience is less strongly related to cluster membership than age; see Figure 5, panel (c), and Table 5. We see that the five clusters are quite evenly distributed among individuals with low level of work experience. On the other hand, higher experience levels are correlated with a higher probability to belong to cluster Highly attached and a lower probability to belong to cluster Mobile + low-attached. Interestingly, the probability of belonging to cluster Retiring is practically independent of the amount of work experience.

The influence of pre-displacement income, measured in tertiles of the income distribution, can be studied in Figure 5, panel (d); see also Table 5. Low income workers have a particularly high probability to belong to cluster Low-attached and, at the same time, a comparably low probability to belong to cluster Highly attached. For the other income groups, cluster

membership resembles that of medium and high experience.

Figure 5, panel (e) and (f), as well as Table 5 show that cluster membership also varies with the size and industry affiliation of the firms from which workers are displaced. The groups with the largest portion in cluster LOW-ATTACHED are workers from small firms and from the service sector. The largest portion in cluster MOBILE + HIGHLY ATTACHED is exhibited by the workers of medium size firms and workers from seasonal business outside of hotel and construction.

	LOW-ATTACHED	HIGHLY ATTACHED	Mobile LA	Mobile HA	RETIRING
White collar	0.408	0.382	0.082	0.044	0.085
Blue collar	0.181	0.478	0.071	0.145	0.125
Age 35–39	0.248	0.609	0.072	0.052	0.018
Age 40–44	0.207	0.634	0.084	0.053	0.021
Age 45–49	0.201	0.526	0.069	0.043	0.160
Age 50–55	0.131	0.022	0.015	0.057	0.775
Experience ≤ 1675 days	0.318	0.325	0.146	0.109	0.102
Experience from					
1676 to 3938 days	0.301	0.459	0.070	0.075	0.096
Experience ≥ 3939 days	0.260	0.526	0.049	0.055	0.109
Income in lowest tertile	0.405	0.218	0.138	0.109	0.130
Income in middle tertile	0.251	0.461	0.075	0.107	0.105
Income in highest tertile	0.291	0.489	0.070	0.056	0.095
Firm size ≤ 10	0.368	0.370	0.106	0.076	0.081
Firm size from 11 to 100	0.238	0.480	0.080	0.087	0.114
Firm size > 100	0.266	0.493	0.036	0.055	0.149
Economic sector: service	0.345	0.426	0.088	0.049	0.093
Economic sector: industry	0.264	0.471	0.092	0.081	0.091
Economic sector: seasonal	0.280	0.418	0.064	0.109	0.129
Economic sector: unknown	0.297	0.440	0.080	0.077	0.105

Table 5: Displaced workers: cluster membership probabilities for a single covariate. All other covariates are set to their mean values observed in the sample.

4.6 Comparison to the control group

After analyzing the career paths of displaced workers in the five different clusters, we now turn to a comparison of the careers of displaced workers with the control group of workers not affected by a plant closure. This gives us some insights in the counterfactual situation that would have arisen, if the plant closure had not taken place. The literature on job displacements typically compares mean post-displacement outcomes among displaced workers with those in a control group of non-displaced worker (Jacobson et al., 2005). Our objective is more complex, as we want to create a separate counterfactual scenario for each cluster group and compare the mean outcome in each group with the counterfactual. To achieve this goal, we propose a novel method that relies on posterior classification of control individuals based on the clustering model that we estimated for the displaced workers. In the following, we describe the corresponding classification of the controls and the simulation of the counterfactual career patterns in each cluster.
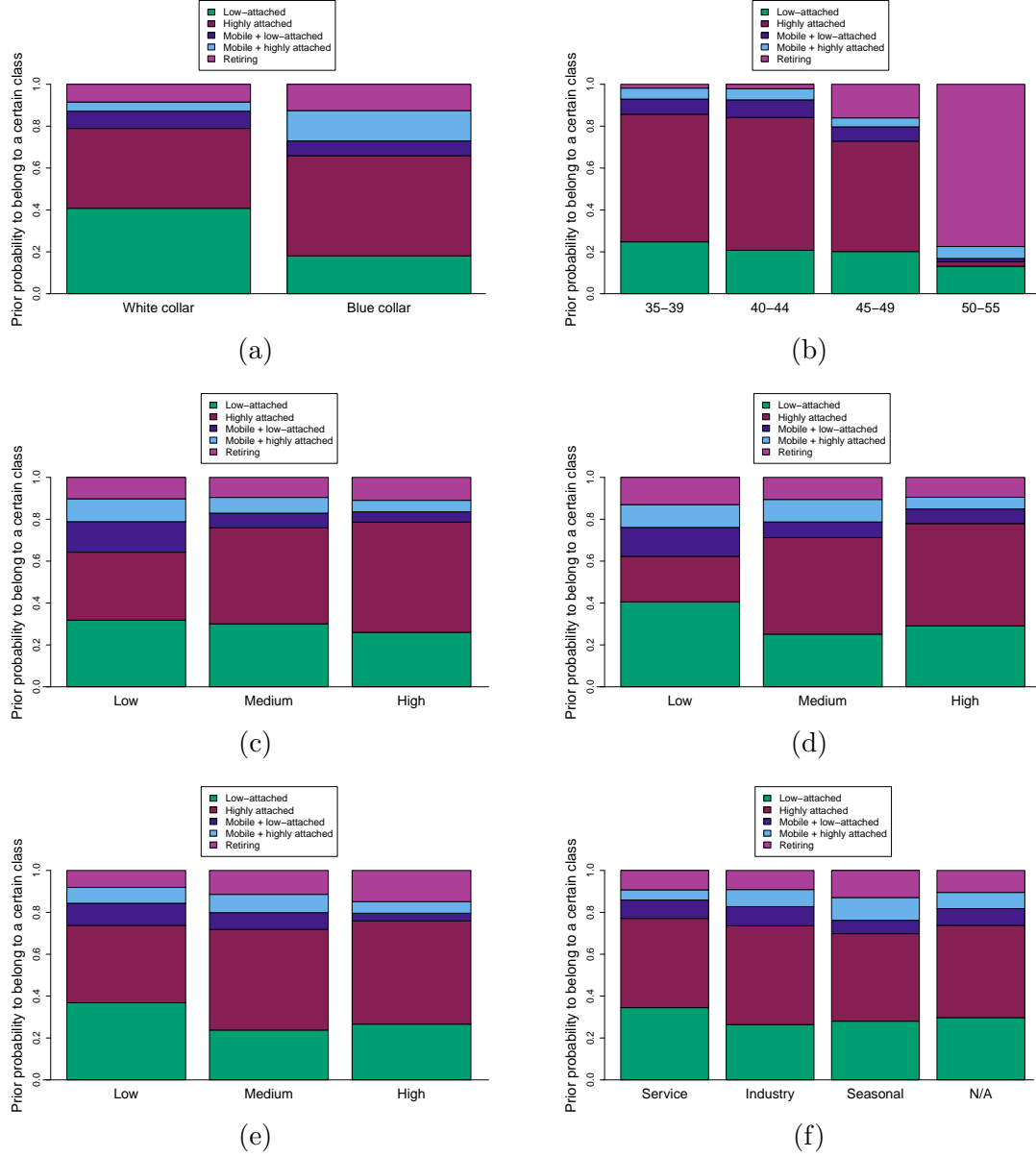
Figure 5: Impact of each covariate on the probability of a worker to belong to a certain cluster: (a) occupational state, (b) age, (c) experience, (d) income at time of plant closure, (e) firm size, (f) firm's economic sector (for each single covariate, all other covariates are set to their mean values observed in the sample). For each covariate, the probabilities of belonging to, respectively, cluster LOW-ATTACHED, HIGHLY ATTACHED, MOBILE + LOW-ATTACHED, MOBILE + HIGHLY ATTACHED and RETIRING are stacked from bottom to the top.

In Section 2 we applied a weighted sampling procedure to construct a control group of non-displaced workers, ensuring that displaced and control individuals are similar with respect to the set of individual characteristics which determine cluster membership in the mixture-of-experts model specified in equation (4). Under the assumption that job displacement due to plant closure is random given these covariates, the only feature that distinguishes the labor market careers of the two groups is the experience of a plant closure. It is evident from Figure 3 that this shock has a dramatic effect on the state distribution $\boldsymbol{\pi}_h = \boldsymbol{\pi}_{h,1}$ of displaced workers in the first quarter after displacement, with a very high rate of being out of labor force in all cluster groups. We thus have to take this event into account when constructing counterfactual career trajectories based on the control group.

The classification of control observations into cluster groups relies on the following thought experiment. What would have happened to the transition patterns in each cluster of displaced workers, if we eliminated the displacement shock in the initial quarter? To implement this thought experiment we assume that the state distribution $\boldsymbol{\pi}_h$ in the first quarter after job displacement incorporates the full extent of the displacement shock and that the subsequent transition patterns are independent from the experience of job displacement. In other words, we assume that the post-displacement behaviour from the second quarter onward is determined by the person's state in the first quarter after (potential) job displacement, the sequence of cluster-specific time-inhomogenous transition matrices $\boldsymbol{\xi}_h = (\boldsymbol{\xi}_{h1}, \ldots, \boldsymbol{\xi}_{h,10})$ estimated in Subsection 4.3 and the vector of individual characteristics $\mathbf{X}$ via the parameters $\boldsymbol{\beta}_2, \ldots, \boldsymbol{\beta}_H$ of the MNL model estimated in Subsection 4.5.

While the typical career transitions are assumed to be the same for all persons within each cluster, regardless of whether the person experienced plant closure or not, it is to be expected that the state distribution at the end of the first quarter after (potential) plant closure is different for the displaced and the controls. Since the initial state in the quarter before displacement is employment also in the control group, i.e. $y_{i0}^c = 1$, their first transition is described by a row vector $\boldsymbol{\pi}_h^c = (\pi_{h,1}^c, \ldots, \pi_{h,4}^c)$. $\boldsymbol{\pi}_h^c$ contains the probability distribution over all states in the first quarter for the controls and is assumed to be different from the state distribution $\boldsymbol{\pi}_h$ of displaced workers due to the absence of the displacement shock. Our assumption implies that beyond the first quarter, the transition matrices $\boldsymbol{\xi}_{h1}, \ldots, \boldsymbol{\xi}_{h,10}$, which were estimated in the displaced sample, can be used to classify the controls into the five clusters.

Based on this cluster model, the cluster assignment of control person i with the observed individual time series denoted by $\mathbf{y}_i^c$ and individual characteristics $\mathbf{x}_i^c$ is performed by computing the posterior distribution $t_{ih}^c(\boldsymbol{\theta}_5) = \Pr(S_i^c = h | \mathbf{y}_i^c, \mathbf{X} = \mathbf{x}_i^c, \boldsymbol{\theta}_5)$ of the class indicator S_i^c over the five clusters for $h = 1, \ldots, 5$ by means of Bayes' rule:

$$t_{ih}^c(\boldsymbol{\theta}_5) \propto p(\mathbf{y}_{i,-1}^c | y_{i1}^c, \boldsymbol{\xi}_h) p(y_{i1}^c | S_i^c = h, \boldsymbol{\pi}_h^c) \Pr(S_i^c = h | \boldsymbol{\beta}_2, \ldots, \boldsymbol{\beta}_H, \mathbf{X} = \mathbf{x}_i^c). \tag{6}$$

In (6), $p(\mathbf{y}_{i,-1}^c | y_{i1}^c, \boldsymbol{\xi}_h)$ is the clustering kernel based on a time-inhomogeneous first order Markov chain as introduced in (2), whereas the cluster-specific state distribution $\boldsymbol{\pi}_h^c = (\pi_{h,1}^c, \ldots, \pi_{h,4}^c)$ for the control group in the first quarter after (potential) plant closure defines:
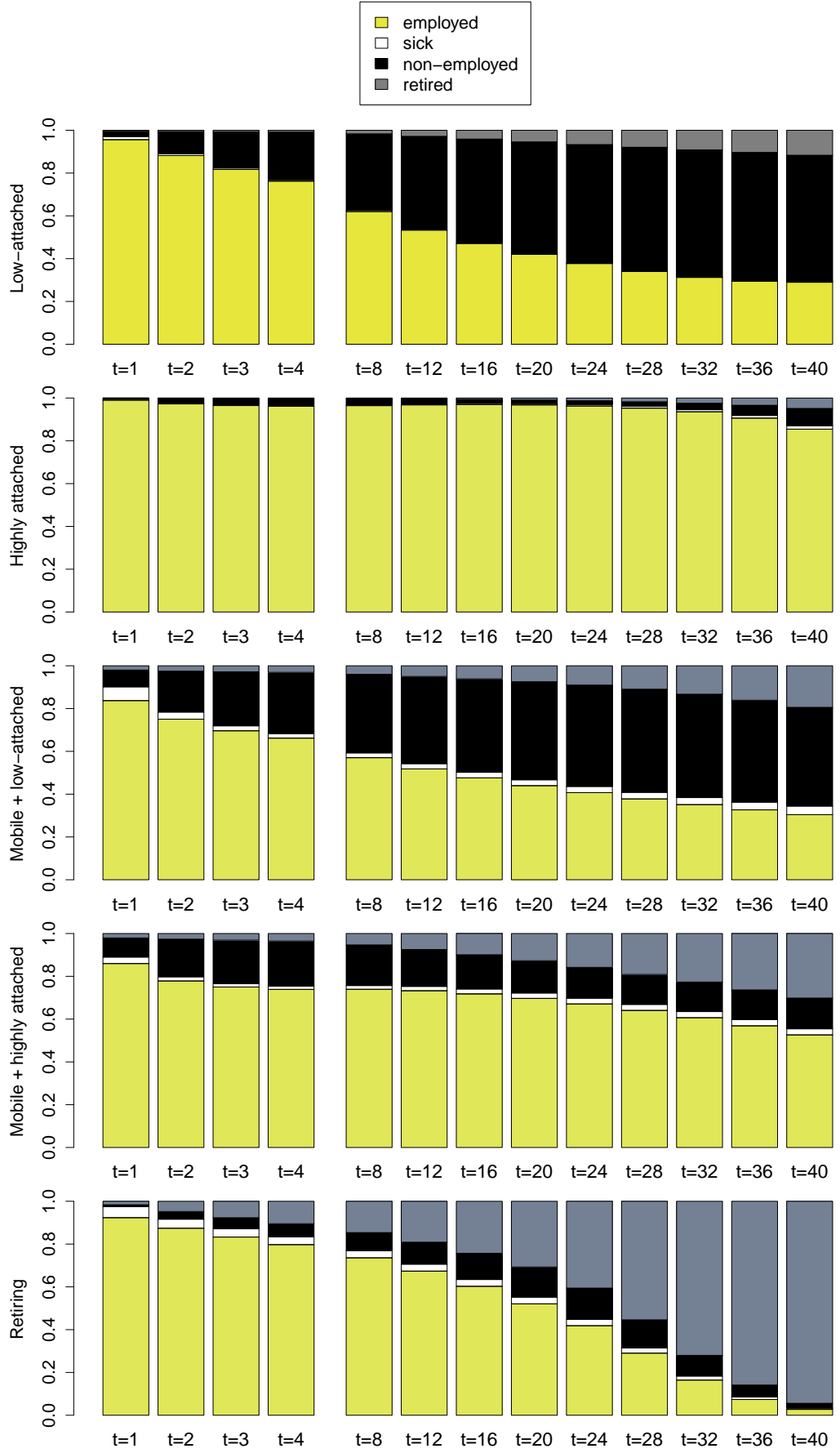
Figure 6: Posterior expectation of the distribution $\boldsymbol{\pi}_{h,t}^c$ over the 4 states ($1=$ employed, $2=$ sick leave, $3=$ out of labor force, $4=$ retired) after a period of t quarters in the various clusters (control group).
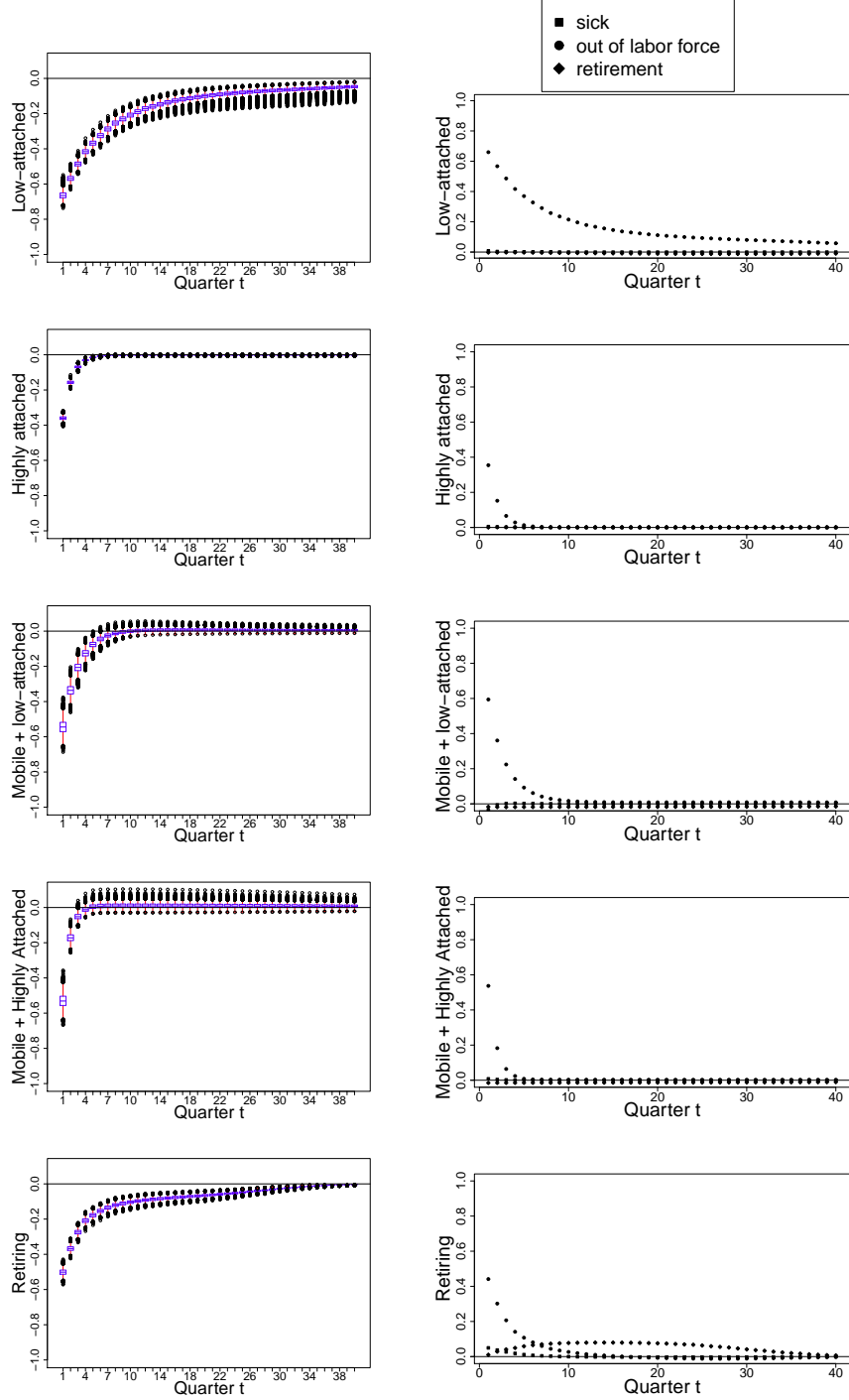
Figure 7: Analysing the difference $\pi_{h,k,t} - \pi_{h,k,t}^c$ in the probability to be in state k ($1 =$ employed, $2 =$ sick leave, $3 =$ out of labor force, $4 =$ retired) between persons experiencing plant closure ($\pi_{h,k,t} = \Pr(y_{it} = k | S_i = h)$) and controls ($\pi_{h,k,t}^c = \Pr(y_{it}^c = k | S_i^c = h)$) for the five clusters (arranged from top to bottom). Left hand side: posterior distribution of the difference $\pi_{h,1,t} - \pi_{h,1,t}^c$ in the probability to be in state "employed" between persons experiencing plant closure and controls; right hand side: posterior expectation of the difference $\pi_{h,k,t} - \pi_{h,k,t}^c$, $k = 2, 3, 4$, in the probability to be in one of the remaining states between persons experiencing plant closure and controls.

$$p(y_{i1}^c|S_i^c = h, \boldsymbol{\pi}_h^c) = \prod_{k=1}^{4} \left(\pi_{h,k}^c\right)^{C_{i,k}},$$

with $\pi_{h,k}^c = \Pr(y_{i1}^c = k|S_i^c = h)$ and $C_{i,k} = I\{y_{i1}^c = k\}$ being an indicator for a non-displaced worker's state in the first quarter. $\Pr(S_i^c = h|\boldsymbol{\beta}_2, \ldots, \boldsymbol{\beta}_H, \mathbf{X} = \mathbf{x}_i^c)$ is the cluster membership distribution introduced in (4), which is based on the individual characteristics $\mathbf{x}_i^c$ of the control person under consideration.

Rather than estimating $(\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_5, \boldsymbol{\beta}_2, \ldots, \boldsymbol{\beta}_5)$ again for the control group, we use the MCMC draws obtained for the displaced persons to assign the individuals from the control group to the five clusters of career patterns during an MCMC-type algorithm. Only the cluster-specific state distributions in the first quarter are estimated by sampling $\boldsymbol{\pi}_h^c$ for each cluster from a Dirichlet distribution, $\boldsymbol{\pi}_h^c|\mathbf{S}^c, \mathbf{y} \sim \mathcal{D}\left(g_{0,1} + C_1^h, \ldots, g_{0,4} + C_4^h\right)$, where $C_k^h = \sum_{i:S_i=h} C_{i,k}$ is the total number of control group workers in cluster h being in state k in the first quarter and $\boldsymbol{\pi}_h^c \sim \mathcal{D}\left(g_{0,1}, \ldots, g_{0,4}\right)$ follows a Dirichlet prior with hyperparameters analogous to those in Appendix A.

We assign individuals in the control group using the posterior expectation $\hat{t}_{ih}^c = \mathrm{E}(t_{ih}^c(\boldsymbol{\theta}_5)|\mathbf{y}_i^c)$. $\hat{t}_{ih}^c$ is estimated by evaluating and averaging $t_{ih}^c(\boldsymbol{\theta}_5)$ as given by (6) using the $15\,000$ MCMC draws of $(\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_5, \boldsymbol{\beta}_2, \ldots, \boldsymbol{\beta}_5)$ obtained for the group of displaced workers and $15\,000$ MCMC draws of $\boldsymbol{\pi}_h^c$ obtained for the group of controls as described above. Each worker from the control group is then allocated to that cluster $\hat{S}_i^c$ which exhibits the maximum posterior probability, i.e. $\hat{S}_i^c$ is defined in such a way that $\hat{t}_{i,\hat{S}_i^c}^c = \max_h \hat{t}_{ih}^c$.

Based on the posterior classification $\hat{S}_i^c$ of all controls, we compute the size of each cluster for the controls. The distribution of individuals in the control group across clusters is shown in Figure 2 and compared to displaced workers. This comparison shows that in absence of the plant closure event the cluster HIGHLY ATTACHED would be considerably larger. The size of cluster RETIRING does not differ much when comparing displaced and control persons, whereas the three remaining clusters are significantly smaller in the absence of a plant closure.

Figure 6 shows the evolution of the posterior expectations $\mathrm{E}(\boldsymbol{\pi}_{h,t}^c|\mathbf{y})$ of the cluster-specific state distribution $\boldsymbol{\pi}_{h,t}^c = \boldsymbol{\pi}_h^c \boldsymbol{\xi}_{h,1 \rightarrow t}$ over distance t from plant closure for the control group, where the transition matrix $\boldsymbol{\xi}_{h,1 \rightarrow t}$ has been defined in (5). Turning to the impact of job displacement from plant closure on career trajectories in the different clusters, the left hand side of Figure 7 shows the posterior distribution of the difference $\Pr(y_{it} = 1|S_i = h) - \Pr(y_{it}^c = 1|S_i^c = h) = \pi_{h,1,t} - \pi_{h,1,t}^c$ for the employment states between displaced and control individuals over distance t from plant closure. Career paths of displaced individuals are characterized by significantly lower employment rates in the initial periods after plant closure throughout all clusters, but eventually employment rates of both groups converge to each other. The speed of convergence varies by cluster, with the fastest convergence rate occurring in clusters HIGHLY ATTACHED and MOBILE + HIGHLY ATTACHED and the lowest convergence rate occurring in cluster LOW-ATTACHED.

Another way to interpret career trajectories in the displaced and counterfactual cases is a direct comparison of Figure 3 and Figure 6, which show the posterior expectation of $\pi_{h,k,t} = \Pr(y_{it} = k|S_i = h)$ and $\pi_{h,k,t}^c = \Pr(y_{it}^c = k|S_i^c = h)$ for all labor market states $k = 1, \ldots, 4$ by cluster. During the first 8 to 12 quarters counterfactual trajectories in all clusters are dominated

by the employment state. At larger distances from the job displacement shock, profiles of displaced and control individuals become very similar, as is also evident from the right hand side of Figure 7.

5   Concluding Remarks

In this paper, we have analysed labor market data from Austria on discrete labor market transitions after a plant closure, where we follow workers over ten years. Economists have shown that the loss of a job due to a plant closure can have major disruptive effects on future careers of workers (Jacobson et al. (1993), Fallick (1996) or Ichino et al. (2016)). They studied only plant closure effects for average persons, whereas our analysis applies elaborated statistical techniques such as model-based clustering using finite mixtures of time-inhomogeneous Markov chain models to explicitly address unobserved heterogeneity in reaction to losing a job due to an exogenous event such as a plant closure.

Modelling workers' transition patterns in such a setting, however, has to address several issues: i) transition patterns immediately after the job loss are very specific, and ii) moreover, as workers age transitions into sick leave and retirement spells become more prevalent. Such – predictable – changes of transition patterns over the life cycle cannot be handled, if time-invariant transition matrices in each cluster are assumed as in Pamminger and Tüchler (2011) or Frühwirth-Schnatter et al. (2012). To address these issues, we developed and applied a more general method of Markov chain clustering analysis, based on inhomogeneous first order Markov transition processes with time-varying transition matrices. As in previous work, a mixture-of-experts model is applied that allows cluster membership to depend on a set of covariates via a multinomial logit model.

Applied to labor market careers after job displacement, the clustering procedure identifies five distinctive clusters which are characterized by a combination of mobility/persistence and attachment to the labor force. Our analysis allows to distinguish between workers who are hardly affected by job displacement and quickly return to stable employment careers job and others who suffer large losses over extended periods of time. It turns out that roughly 50% of workers remain "highly attached" to the labor market even after a plant closure, whereas 30% are "low-attached", they have difficulties with returning to stable jobs and thus suffer large employment and earnings losses, and 20% belong to a group which takes early retirement as an option to exit the labor market.

The empirical findings have important policy implications. In order to implement efficient policies that provide support to workers hit by an economic downturn or a local shock due to the closure of a large plant, it is essential to target groups who are potentially most adversely affected. Our results provide some guidance with respect to observed characteristics of the least attached groups and the dynamics of the effects of displacement. This allows not only targeting individuals by their characteristics but also to identify the relevant timing for interventions.

The model-based clustering approach developed in this paper for the analysis of the plant closure data might be useful in other areas of applied research, whenever transition processes have to be modelled that are not necessarily stationary over time. This situation typically occurs when transition processes are analyzed over the entire life cycle of an entity, and transition rates

differ between the beginning and the end of the life cycle. Other reasons for nonstationarity are shocks to the stationary transition processes caused by events out of the entities' control, such as stock market crashes or natural disasters. In these cases the patterns of transition during the recovery phase may differ significantly from stationary transitions.

Furthermore, we see several interesting avenues for future statistical research for these type of models. It is possible to include covariates not only in modelling the latent group indicator, as we did in the present study, but also in the conditional distribution of the observed outcome, given the group indicator, see e.g. Frühwirth-Schnatter (2011) for a review. As common for mixtures of discrete data, another important issue is verifying generic identifiability. Examples are easily constructed, where generic identifiability fails, because discrete-valued covariates with too little heterogeneity such as, for instance, age groups, appear both in the mixture-of experts part and in the group-specific transition model. To our knowledge, generic identifiability of mixtures of Markov chain models has not been discussed in full mathematical depth and might be an interesting venue for future research.

Finally, in the present paper, we were interested in clustering the entire employment profile of a worker after plant closure. Statistically, this implied to hold the cluster indicator S_i for each worker constant over each time series. Dynamic mixture models where the cluster indicator $S_{i,t}$ changes over time and allows worker to switch clusters is another interesting line of future research.

Acknowledgements

References

Aitkin, M. and M. Alfó (1998). Regression models for binary longitudinal responses. *Statistics and Computing 8*, 289–307.

Altman, R. (2007). Mixed hidden Markov models: An extension of the hidden Markov model to the longitudinal data setting. *Journal of The American Statistical Association 102*, 201–210.

Banfield, J. D. and A. E. Raftery (1993). Model-based Gaussian and non-Gaussian clustering. *Biometrics 49*, 803–821.

Bartolucci, F., S. Bacci, and F. Pennoni (2014). Longitudinal analysis of self-reported health status by mixture latent auto-regressive models. *Journal of the Royal Statistical Society, Series C 63*, 267–288.

Cadez, I., D. Heckerman, C. Meek, P. Smyth, and S. White (2003). Model-based clustering and visualization of navigation patterns on a web site. *Data Mining and Knowledge Discovery 7*(4), 399–424.

Couch, K., N. Jolley, and D. Placzek (2010). Earnings impact of job displacement revisited. *American Economic Review 100*, 572–589.

Del Bono, E., A. Weber, and R. Winter-Ebmer (2012). Clash of career and family: fertility decisions after job displacement. *Journal of the European Economic Association 10*(4), 659–683.

Dias, J. G. and J. K. Vermunt (2007). Latent class modeling of website users' search patterns: Implications for online market segmentation. *Journal of Retailing and Consumer Services 14*, 359–368.

Diebolt, J. and C. P. Robert (1994). Estimation of finite mixture distributions through Bayesian sampling. *Journal of the Royal Statistical Society, Ser. B 56*, 363–375.

Diggle, P. J., P. Heagerty, K.-Y. Liang, and S. L. Zeger (2002). *Analysis of Longitudinal Data* (2 ed.). Oxford University Press.

Draper, N. R. and H. Smith (1998). *Applied Regression Analysis* (3 ed.). New York: John Wiley & Sons.

Eliason, M. and D. Storrie (2004). The echo of job displacement. *ISER Working Paper 20*.

Fallick, B. (1996). A review of the recent empirical literature on displaced workers. *Industrial and Labor Relations Review 50*, 5–16.

Fraley, C. and A. E. Raftery (2002). Model-based clustering, discriminant analysis, and density estimation. *Journal of the American Statistical Association 97*, 611–631.

Frühwirth-Schnatter, S. (2006). *Finite Mixture and Markov Switching Models*. New York: Springer-Verlag.

Frühwirth-Schnatter, S. (2011). Panel data analysis - a survey on model-based clustering of time series. *Advances in Data Analysis and Classification 5*, 251–280.

Frühwirth-Schnatter, S. and R. Frühwirth (2010). Data augmentation and MCMC for binary and multinomial logit models. In T. Kneib and 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/ifas/content/e114480`, IFAS Research Paper Series 2010-48.

Frühwirth-Schnatter, S. and S. Kaufmann (2008). Model-based clustering of multiple time series. *Journal of Business & Economic Statistics 26*, 78–89.

Frühwirth-Schnatter, S., C. Pamminger, A. Weber, and R. Winter-Ebmer (2012). Labor market entry and earnings dynamics: Bayesian inference using mixtures-of-experts Markov chain clustering. *Journal of Applied Econometrics 27*, 1116–1137.

Frühwirth-Schnatter, S., C. Pamminger, A. Weber, and R. Winter-Ebmer (2016). Mothers' long-run career patterns after first birth. *Journal of the Royal Statistical Society, Series A 179*, 707–725.

Frydman, H. (2005). Estimation in the mixture of Markov chains moving with different speeds. *Journal of the American Statistical Association 100*, 1046–1053.

Gamerman, D. and H. F. Lopes (2006). *Markov Chain Monte Carlo. Stochastic Simulation for Bayesian Inference* (2 ed.). Boca Raton, FL: Chapman & Hall/CRC.

Gollini, I. and T. B. Murphy (2014). Mixture of latent trait analyzers for model-based clustering of categorical data. *Statistics and Computing 24*, 569–588.

Goodman, L. A. (1961). Statistical methods for the mover-stayer model. *Journal of the American Statistical Association 56*, 841–868.

Gormley, I. C. and T. B. Murphy (2008). A mixture of experts model for rank data with applications in election studies. *Annals of Applied Statistics 2*, 1452–1477.

Heckman, J. (1981). The incidental parameters problem and the problem of initial conditions in estimating a discrete time-discrete data stochastic process. In C. F. Manski and D. McFadden (Eds.), *Structural Analysis of Discrete Data with Econometric Applications*, pp. 179–195. Cambridge MA: MIT Press.

Huttunen, K., J. Moen, and K. G. Salvanes (2011). How destructive is creative destruction? Effects of job loss on job mobility, withdrawal and income. *Journal of the European Economic Association 9(5)*, 840–870.

Ichino, A., G. Schwerdt, R. Winter-Ebmer, and J. Zweimüller (2016). Too old to work, too young to retire? *Journal of the Economics of Ageing*.

Imbens, G. W. (2004). Nonparametric estimation of average treatment effects under exogeneity: A review. *Review of Economics and statistics 86*(1), 4–29.

Jacobson, L. S., R. J. LaLonde, and D. G. Sullivan (1993). Earnings losses of displaced workers. *American Economic Review 83*(4), 685–709.

Jacobson, L. S., R. J. LaLonde, and D. G. Sullivan (2005). Estimating the returns to community college schooling for displaced workers. *Journal of Econometrics 125*, 271–304.

Juárez, M. A. and M. F. J. Steel (2010). Model-based clustering of non-Gaussian panel data based on skew-t distributions. *Journal of Business & Economic Statistics 28*, 52–66.

Maruotti, A. and R. Rocci (2012). A mixed non-homogeneous hidden Markov model for categorical data, with application to alcohol consumption. *Statistics in Medicine 31*, 871–886.

McNicholas, P. D. and T. B. Murphy (2010). Model-based clustering of longitudinal data. *The Canadian Journal of Statistics 38*, 153–168.

Pamminger, C. and S. Frühwirth-Schnatter (2010). Model-based clustering of categorical time series. *Bayesian Analysis 5*, 345–368.

Pamminger, C. and R. Tüchler (2011). A Bayesian analysis of female wage dynamics using Markov chain clustering. *Austrian Journal of Statistics 40*, 281–296.

Peng, F., R. A. Jacobs, and M. A. Tanner (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.

Ramoni, M., P. Sebastiani, and P. Cohen (2002). Bayesian clustering by dynamics. *Machine Learning 47*, 91–121.

Ruhm, C. J. (1991). Are workers permanently scarred by job displacements? *American Economic Review 81*(1), 319–324.

Schwerdt, G., A. Ichino, O. Ruf, R. Winter-Ebmer, and J. Zweimüller (2010). Does the color of the collar matter? Employment and earnings after plant closure. *Economics Letters 108*(2), 137 – 140.

Shirley, K. E., D. S. Small, K. G. Lynch, S. A. Maisto, and D. W. Oslin (2010). Hidden Markov models for alcoholism treatment trial data. *Annals of Applied Statistics 4*, 366–395.

Skrondal, A. and S. Rabe-Hesketh (2014). Handling initial conditions and endogenous covariates in dynamic/transition models for binary data with unobserved heterogeneity. *Journal of the Royal Statistical Society, Series C 63*, 211–237.

Sullivan, D. and T. von Wachter (2009). Job displacement and mortality: an analysis using administrative data. *The Quarterly Journal of Economics 124*(3), 1265–1306.

Winter-Ebmer, R. (2016). Long-term effects of unemployment: What can we learn from plant-closure studies? In S. Bentolila and M. Jansen (Eds.), *Long-term unemployment after the Great Recession*, pp. 33–42. CEPR Press.

Wooldridge, J. M. (2005). Simple solutions to the initial conditions problem in dynamic, nonlinear panel data models with unobserved heterogeneity. *Journal of Applied Econometrics 20*, 39–54.

Zweimüller, J., R. Winter-Ebmer, R. Lalive, A. Kuhn, J.-P. Wuellrich, O. Ruf, and S. Büchi (2009). The Austrian Social Security Database (ASSD). Technical report.

A Computational Details

In this section, we summarize the Bayesian approach toward estimating the unknown parameters $\boldsymbol{\theta}_H = (\boldsymbol{\vartheta}_1, \ldots, \boldsymbol{\vartheta}_H, \boldsymbol{\beta}_2, \ldots, \boldsymbol{\beta}_H)$ and the latent cluster indicators $\mathbf{S} = (S_1, \ldots, S_N)$ from categorical panel data $\mathbf{y} = \{\mathbf{y}_1, \ldots, \mathbf{y}_N\}$ for a fixed number H of clusters.

In a Bayesian framework, estimation of $\boldsymbol{\theta}_H$ is based on the posterior distribution $p(\boldsymbol{\theta}_H|\mathbf{y})$ of $\boldsymbol{\theta}_H$ given $\mathbf{y}$. Using Bayes' theorem, the posterior distribution $p(\boldsymbol{\theta}_H|\mathbf{y})$, given by $p(\boldsymbol{\theta}_H|\mathbf{y}) \propto p(\mathbf{y}|\boldsymbol{\theta}_H)p(\boldsymbol{\theta}_H)$, is derived as the product of the prior distribution $p(\boldsymbol{\theta}_H)$ and the observed-data (mixture) likelihood function $p(\mathbf{y}|\boldsymbol{\theta}_H)$ given by

$$p(\mathbf{y}|\boldsymbol{\theta}_H) = \prod_{i=1}^{N} \left(\sum_{h=1}^{H} p(\mathbf{y}_i|\boldsymbol{\vartheta}_h) \Pr(S_i = h|\boldsymbol{\beta}_2, \ldots, \boldsymbol{\beta}_H, \mathbf{X} = \mathbf{x}_i) \right), \tag{7}$$

where $p(\mathbf{y}_i|\boldsymbol{\vartheta}_h)$ is the clustering kernel defined in (1) and $\Pr(S_i = h|\boldsymbol{\beta}_2, \ldots, \boldsymbol{\beta}_H, \mathbf{X} = \mathbf{x}_i)$ is given by the mixture-of-experts model (4).

Concerning the prior distribution $p(\boldsymbol{\theta}_H)$, we assume prior independence between the parameters $(\boldsymbol{\beta}_2, \ldots, \boldsymbol{\beta}_H)$ of the mixture-of-experts model and the class-specific parameters $(\boldsymbol{\vartheta}_1, \ldots, \boldsymbol{\vartheta}_H)$ of the clustering kernel. All parameter vectors $\boldsymbol{\beta}_h, h = 2, \ldots, H$, are assumed to be independent a priori, each following a standard normal distribution of dimension $r + 1$. This means that also the individual regression coefficients inside a single vector $\boldsymbol{\beta}_h = (\beta_{h0}, \ldots, \beta_{hr})$ are independent a priori, each having a $\mathcal{N}(0, 1)$ distribution.

The prior distribution for each class-specific time-varying transition matrix $\boldsymbol{\vartheta}_h$ is composed of priors being conditionally conjugate to the time-varying Markov chain clustering kernel $p(\mathbf{y}_i|\boldsymbol{\vartheta}_h)$ defined in (1). This choice implies that each state distribution $\boldsymbol{\pi}_h$ follows a priori a Dirichlet distribution $\mathcal{D}(g_{0,1}, \ldots, g_{0,4})$ with hyperparameters $g_{0,1}, \ldots, g_{0,4}$. Furthermore, the three rows $\boldsymbol{\xi}_{hy,1\cdot}, \ldots, \boldsymbol{\xi}_{hy,3\cdot}$ of all transition matrices $\boldsymbol{\xi}_{hy}$, $y = 1, \ldots, 10$, $h = 1, \ldots, H$, are independent a priori, each following a Dirichlet distribution $\mathcal{D}(e_{0,yj1}, \ldots, e_{0,yj4})$ with hyperparameters $e_{0,yj1}, \ldots, e_{0,yj4}$, for $j = 1, 2, 3$.

We use empirical transition counts to define weakly informative hyperparameters for these prior distributions. More specifically, we define the 3×4 empirical initial count matrix $N^0 = (N_{jk}^0)$, where for each state $k = 1, \ldots, 4$ the element of the first row is equal to $N_{1k}^0 := \#\{y_{i1} = k$ for some person $i\}$, and equal to 0 in the second and the third row (i.e. $N_{jk}^0 = 0$ for $j = 2, 3$). Furthermore, we define for each year $y = 1, \ldots, 10$ the 3×4 empirical transition count matrix $N^y = (N_{jk}^y)$ with elements

$$N_{jk}^y = \#\{y_{i,t-1} = j, \ y_{it} = k \quad \text{for some person } i \text{ and some quarter } t \text{ in year } y\},$$

for $j = 1, 2, 3$ and $k = 1, 2, 3, 4$. For each $y = 0, 1, \ldots, 10$, we define the empirical transition matrices $\tilde{N}^y := (N_{jk}^y/r_j^y)$, where $r_j^y := \sum_{k=1}^{4} N_{jk}^y$ are the row sums for each $j = 1, 2, 3$. In our special application, we had all of these row sums greater than zero except for those two rows in N^0 whose sum is trivially equal to zero. The matrix $\bar{N}$ is then defined as the average over these

11 matrices:

$$\bar{N} = (\bar{N}_{jk}) := \sum_{y=0}^{10} \tilde{N}^y / 11.$$

The initial distribution $\boldsymbol{\pi}_h$ follows a $\mathcal{D}(g_{0,1}, \ldots, g_{0,4})$ prior with $g_{0,k} := \max\{17\bar{N}_{1k}, \ 0.5\}$, whereas the rows $\boldsymbol{\xi}_{hy,1\cdot}, \ldots, \boldsymbol{\xi}_{hy,3\cdot}$ of each transition matrix $\boldsymbol{\xi}_{hy}$ follow a $\mathcal{D}(e_{0,yj1}, \ldots, e_{0,yj4})$ prior with $e_{0,yjk} := \max\{17\bar{N}_{jk}, \ 0.5\}$.

Since the posterior distribution $p(\boldsymbol{\theta}_H|\mathbf{y})$ does not have a closed form, Bayesian inference is carried out by sampling M draws from $p(\boldsymbol{\theta}_H|\mathbf{y})$, using Markov chain Monte Carlo (MCMC) methods based on data augmentation – a method that has been introduced for finite mixture models by Diebolt and Robert (1994). See Gamerman and Lopes (2006) for a review of MCMC-based statistical inference and Frühwirth-Schnatter (2006) for a review of MCMC estimation of mixture models. The data augmentation technique underlying MCMC estimation also provides estimates of the latent class indicators $\mathbf{S} = (S_1, \ldots, S_N)$.

After starting MCMC with some initial classification (partition) of the N subjects into H disjoint classes, by assigning an initial value $\mathbf{S}_0$ to the latent cluster indicators $\mathbf{S} = (S_1, \ldots, S_N)$, the following steps are repeated during a burn-in period to achieve convergence and additional M iteration steps are performed to produce the desired number of draws:

(a) Sample the unknown parameters $\boldsymbol{\beta}_2, \ldots, \boldsymbol{\beta}_H$ in the mixture-of-experts model (4) from the conditional posterior distribution $p(\boldsymbol{\beta}_2, \ldots, \boldsymbol{\beta}_H|\mathbf{S}) \propto \prod_{i=1}^N p(S_i|\boldsymbol{\beta}_2, \ldots, \boldsymbol{\beta}_H) p(\boldsymbol{\beta}_2, \ldots, \boldsymbol{\beta}_H)$.

(b) Sample the class-specific parameters $\boldsymbol{\vartheta}_1, \ldots, \boldsymbol{\vartheta}_H$: draw $\boldsymbol{\vartheta}_h$ independently from the conditional posterior distribution $p(\boldsymbol{\vartheta}_h|\mathbf{S}, \mathbf{y}) \propto \prod_{i=1}^N p(\mathbf{y}_i|\boldsymbol{\vartheta}_h) p(\boldsymbol{\vartheta}_h)$ for each $h = 1, \ldots, H$.

(c) Bayes' classification for each subject i: determine a random clustering $\mathbf{S} = (S_1, \ldots, S_N)$ of the N subjects into H classes by sampling, independently for all $i = 1, \ldots, N$, S_i from the discrete posterior distribution $\Pr(S_i = 1|\mathbf{y}_i, \cdot), \ldots, \Pr(S_i = H|\mathbf{y}_i, \cdot)$ given by:

$$\Pr(S_i = h|\mathbf{y}_i, \boldsymbol{\theta}_H, \mathbf{X}) \propto p(\mathbf{y}_i|\boldsymbol{\vartheta}_h)\Pr(S_i = h|\boldsymbol{\beta}_2, \ldots, \boldsymbol{\beta}_H, \mathbf{X}), \quad h = 1, \ldots, H, \qquad (8)$$

where $p(\mathbf{y}_i|\boldsymbol{\vartheta}_h)$ is the clustering kernel defined in (1) and $\mathbf{X} = \mathbf{x}_i$.

For the mixture-of-experts model (4), the regression coefficients $(\boldsymbol{\beta}_2, \ldots, \boldsymbol{\beta}_H)$ are sampled in step (a) from the posterior distribution $p(\boldsymbol{\beta}_2, \ldots, \boldsymbol{\beta}_H|\mathbf{S})$, where the likelihood $p(S_i|\boldsymbol{\beta}_2, \ldots, \boldsymbol{\beta}_H)$ is obtained from the MNL model (4) with $\mathbf{X} = \mathbf{x}_i$. To sample $\boldsymbol{\beta}_2, \ldots, \boldsymbol{\beta}_H$, we follow Frühwirth-Schnatter et al. (2012) and apply auxiliary mixture sampling in the differenced random utility model representation of the MNL model (Frühwirth-Schnatter and Frühwirth, 2010), because this method seems to be superior to other MCMC methods in terms of the effective sampling rate.

Closed form Gibbs sampling of $\boldsymbol{\vartheta}_h = (\boldsymbol{\pi}_h, \boldsymbol{\xi}_{h1}, \ldots, \boldsymbol{\xi}_{h,10})$ in Step (b) is possible, since the prior $p(\boldsymbol{\vartheta}_h)$ is conditionally conjugate to the clustering kernel $p(\mathbf{y}_i|\boldsymbol{\vartheta}_h)$. For each cluster, the initial distribution $\boldsymbol{\pi}_h$ and the various rows $\boldsymbol{\xi}_{hy,j\cdot}$ of the time-varying transition matrix $\boldsymbol{\xi}_{hy}$ are conditionally independent, given $\mathbf{S}$ and $\mathbf{y}$. In each cluster, the initial distribution $\boldsymbol{\pi}_h$ is sampled

from the Dirichlet distribution,

$$\boldsymbol{\pi}_h | \mathbf{S}, \mathbf{y} \sim \mathcal{D}\left(g_{0,1} + I_1^h, \ldots, g_{0,4} + I_4^h\right), \tag{9}$$

where $I_k^h := \sum_{i:S_i=h} I_{i,k}$ is the total number of workers in cluster h being in state k at the end of the first quarter after plant closure. I_k^h is the sum of the individual indicators $I_{i,k}$, defined after (3), over all cluster members.

The various rows $\boldsymbol{\xi}_{hy,j.}$ are sampled row-by-row from a total of $30H$ Dirichlet distributions:

$$\boldsymbol{\xi}_{hy,j.} | \mathbf{S}, \mathbf{y} \sim \mathcal{D}\left(e_{0,yj1} + N_{y,j1}^h, \ldots, e_{0,yj4} + N_{y,j4}^h\right), \quad y = 1, \ldots, 10, j = 1, \ldots, 3, \ h = 1, \ldots, H,$$

where $N_{y,jk}^h := \sum_{i:S_i=h} N_{iy,jk}$ is the total number of transitions from state j into state k observed in cluster h in period y. $N_{y,jk}^h$ is the sum of the individual counts $N_{iy,jk}$, defined after (2), over all cluster members.

At the end of Step (b) the following smoothing procedure is applied to the transition probabilities. For each cluster h, for each row j and for each column k, we apply a standard polynomial regression technique with a quadratic polynomial (Draper and Smith, 1998) to smooth the ten time-varying transition probabilities $\xi_{h1,jk}, \xi_{h2,jk}, \ldots, \xi_{h10,jk}$ over time. After this smoothing step, we consider each row $\boldsymbol{\xi}_{hy,j.}$ of the smoothed transition matrices $\boldsymbol{\xi}_{hy}$. Whenever one element of such a row is below zero, i.e. $\xi_{hy,jk} < 0$, it is set to zero: $\xi_{hy,jk} = 0$ and each row $\boldsymbol{\xi}_{hy,j.}$ is normalized by $\boldsymbol{\xi}_{hy,j.}/\sum_{k=1}^{4} \xi_{hy,jk}$ to ensure that all row sums are equal to one as required for transition matrices.

We start MCMC estimation by choosing the initial values $\mathbf{S}_0$ for the cluster indicators $\mathbf{S}$ through the following procedure. For each person i, we define a vector $\boldsymbol{q}_i$ containing the four indicators $N_{i0,k}$, where for each $k = 1, \ldots, 4$,

$$N_{i0,k} := \begin{cases} 1 & \text{for } y_{i,1} = k \\ 0 & \text{else.} \end{cases}$$

as well as all 120 empirical transition counts $N_{i1,jk}$ and $N_{iy,jk}$ defined after formula (2). Adding 0.5 to each element of $\boldsymbol{q}_i$ gives the vector $\boldsymbol{v}_i$. Clustering all N resulting vectors $\log(\boldsymbol{v}_i)$ into H clusters using the k-means algorithm gives the desired initial classification $\mathbf{S}_0$.

To perform step (a) of our MCMC scheme, we also need starting values for the parameters $\boldsymbol{\beta}_2, \ldots, \boldsymbol{\beta}_H$ in the mixture-of-experts model in addition to $\mathbf{S}_0$. Given both the covariate vectors $\mathbf{x}_i$ for all N persons under consideration as well as the initial classification vector $\mathbf{S}_0$, we are dealing with a multinomial logit regression (MNL) model. We use the estimated coefficients of this MNL model as starting values for $\boldsymbol{\beta}_2, \ldots, \boldsymbol{\beta}_H$ in our MCMC procedure. To this aim, we applied the function `multinom` from the R package `nnet`.

Like any mixture model, mixtures of Markov chain models are invariant to relabelling the clusters and as a consequence, MCMC draws might suffer from label switching. However, in our empirical application, we are dealing with a large data set, which typically leads to a clear separation of the various (equivalent) modes of the posterior distribution (see, e.g. Frühwirth-Schnatter (2006, Chapter 3)). As a consequence, post-processesing of the MCMC draws did not reveal any signs of label switching.