aging·health·labor
CHRISTIAN DOPPLER LABORATORY

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

by

SYLVIA FRÜHWIRTH-SCHNATTER
STEFAN PITTNER
ANDREA WEBER
RUDOLF WINTER-EBMER

September 2016

Corresponding author: rudolf.winterebmer@jku.at

Christian Doppler Laboratory
Aging, Health and the Labor Market
cdecon.jku.at

Johannes Kepler University
Department of Economics
Altenberger Strasse 69
4040 Linz, Austria

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

Sylvia Frühwirth-Schnatter* Stefan Pittner† Andrea Weber‡
Rudolf Winter-Ebmer§

September 3, 2016

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 develop and 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 prior probability to belong to a certain cluster in dependence of 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 allow 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 consider 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. We can classify labor market states by quarter as being employed, sick, out of labor force, or retired.

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

Modelling transitions between discrete states over time is of interest not only in labor economics, but in many other 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 a longitudinal component.

When analyzing the effect of plant closure on career patterns in our specific application, we expect that unobserved heterogeneity in the response to job displacement from plant closure is present in the data. To account for unobserved heterogeneity and to identify subgroups 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, see Banfield and Raftery (1993); Fraley and Raftery (2002); McNicholas and Murphy (2010); Gollini and Murphy (2014) among many others. For model-based clustering of discrete-valued time series, typically first order Markov chain models are used to model transitions between states and separate clusters are distinguished by different transition matrices, see Frühwirth-Schnatter (2011) for a recent review.

Two important questions arise in this context. First, time-invariant or predetermined characteristics of a displaced worker may be correlated with group membership, i.e. persons with specific characteristics are more likely to belong to a certain cluster than to the other clusters. This issue can easily be addressed through the mixture-of-experts approach introduced by Peng et al. (1996), which allows to model the prior probability to belong to a specific cluster in dependence of covariates, see e.g. Gormley and Murphy (2008) for an application to model-based clustering of rank data, Frühwirth-Schnatter and Kaufmann (2008) and Juárez and Steel (2010) for an application to model-based clustering of time series of continuous outcomes and Frühwirth-Schnatter et al. (2012) for an application to model-based clustering of discrete-valued time series. To obtain a better understanding which workers in our data set are inclined toward which career pattern, such a mixture-of-experts approach based on a multinomial logit model is applied to model the prior probability to belong to a certain cluster in dependence of control variables, such as 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.

Second, previous approaches of Markov chain clustering of discrete-valued time series are typically based on time-homogeneous first order Markov chains, see e.g. Cadez et al. (2003); Ramoni et al. (2002); Frydman (2005); Pamminger and Frühwirth-Schnatter (2010). However, for our data the transition process is not necessarily stationary over time which poses an obvious challenge to time-invariant transition processes. 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 develop and apply a new method of Markov chain clustering and extend previous work on modelling transitions between labor market states through time-homogeneous Markov chain clustering. We extend the approach of Frühwirth-Schnatter et al. (2012) who introduced mixture-of-experts homogeneous Markov chain clustering for this type of time series by introducing inhomogeneous first order Markov transition processes with time-varying transition matrices as clustering kernels.

For our plant closure data, this new method of model-based 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. By addressing this unobserved heterogeneity explicitly, our paper contributes to the labor economics literature by revealing a variety of different shock-absorption patterns across multiple clusters, while previous research concentrated only on average effects of job displacements

The paper proceeds as follows. The next section introduces the empirical problem and the data from Austrian social security registers. Section 3 introduces the time-varying Markov chain clustering model and discusses 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 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 continue 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.

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 follow these 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 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 windows 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.

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 do 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 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. Control variables include 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. For computational reasons we transform all these variables into discrete categories; for summary statistics see Table 1.

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.

Worker's age (in years)	
Age 35–39	28 %
Age 40–44	28 %
Age 45–49	23 %
Age 50–55	21 %
Worker's professional experience (in days)	
Experience $\leq$ 1675 days	33 %
Experience from 1676 to 3938 days	31 %
Experience $\geq$ 3939 days	36 %
Worker's income at time of plant closure	
Income in lowest tertile	14 %
Income in middle tertile	32 %
Income in highest tertile	54 %

Firm's attributes	
Firm size $\leq$ 10	42 %
Firm size from 11 to 100	41 %
Firm size $>$ 100	17 %
Economic sector: service	31 %
Economic sector: industry	32 %
Economic sector: seasonal	2 %
Economic sector: unknown	35 %
White-collar workers	56 %
Blue-collar workers	44 %

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

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). 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 $\mathbf{y}_i = \{y_{i1}, \dots, y_{i,T_i}\}$, depending on an unknown class-specific parameter $\boldsymbol{\vartheta}_h$. A latent cluster indicator S_i taking a value in the set $\{1, \dots, H\}$ is introduced for each time series $\mathbf{y}_i$ to indicate which class the individual i belongs to, i.e. $p(\mathbf{y}_i|S_i, \boldsymbol{\vartheta}_1, \dots, \boldsymbol{\vartheta}_H) = p(\mathbf{y}_i|\boldsymbol{\vartheta}_{S_i})$.

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 distance to plant closure reveals that the employment rate does not converge to a steady state, but rather declines steadily with increasing distance to plant closure. Homogeneity would imply that the employment rate as well as all other state probabilities converge to a steady state within the observation period, both within each cluster as well marginalized over all clusters.

To obtain such a non-stationary pattern, we need to assume that the transition probabilities between the various states change with distance to 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 develop in the present paper Markov chain clustering based on a time-inhomogeneous first order Markov chain model with class-specific time-varying transition matrices $\boldsymbol{\vartheta}_h = (\boldsymbol{\pi}_h, \boldsymbol{\xi}_{h1}, \ldots, \boldsymbol{\xi}_{h10})$ as clustering kernel. More specifically, we assume that the transition behavior changes with distance to 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 the states y_{i1} 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:

$$p(\mathbf{y}_i|\boldsymbol{\vartheta}_h) = p(\mathbf{y}_{i,-1}|y_{i1}, \boldsymbol{\xi}_{h1}, \ldots, \boldsymbol{\xi}_{h10}) p(y_{i1}|S_i = h, \boldsymbol{\pi}_h), \tag{1}$$

where the truncated time series $\mathbf{y}_{i,-1} = \{y_{i2}, \ldots, y_{i,T_i}\}$ is described by time-varying transition matrices changing every year:

$$p(\mathbf{y}_{i,-1}|y_{i1}, \boldsymbol{\xi}_{h1}, \ldots, \boldsymbol{\xi}_{h10}) = \prod_{y=1}^{10} \prod_{j=1}^{3} \xi_{hy,jk}^{N_{iy,jk}}, \tag{2}$$

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). This issue is relevant for cases, where unobserved heterogeneity is either captured through an individual effect S_i following a continuous distribution, but, as discussed in Frühwirth-Schnatter et al. (2012), also for models where S_i follows a discrete distribution as for model-based clustering based on transition models. The key issue is to allow for dependence between the initial state y_{i1} and the latent variable S_i, which can be achieved by allowing the prior distribution of S_i to depend on y_{i1}, an approach that has been pursued for the analysis of labor market entry and earnings dynamics in Frühwirth-Schnatter et al. (2012).

In the present paper, we factorize the joint distribution of y_{i1} and S_i in a different way as $p(y_{i1}, S_i|\cdot) = p(y_{i1}|S_i, \boldsymbol{\pi}_{S_i}) p(S_i|\boldsymbol{\beta}_2, \ldots, \boldsymbol{\beta}_H, \mathbf{x}_i)$, where the entire initial state distribution changes across the clusters, i.e.:

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

where $\pi_{h,k} = \Pr(y_{i1} = k|S_i = h)$ 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 prior distribution of the latent indicator S_i is influenced by exogenous covariates and is modeled as following a multinomial logit (MNL) model:

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

where the row vector $\mathbf{x}_i = (1, x_{i1}, \ldots, x_{ir})$ includes the constant 1 for the intercept in addition to the exogenous covariates $(x_{i1}, \ldots, x_{ir})$. 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.

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.[1]

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 displaced workers with a control group who did not experience a job loss.

4.1 Model Selection

Statistical model selection criteria such as the AIC, the BIC or the AWE criterion as discussed e.g. in Frühwirth-Schnatter (2011) could be applied to the present data to select the number H of clusters, however, these statistical criteria typically are not unambiguous and do not give a clear answer. For this reason, we select the number of clusters based on the economic interpretability

[1] 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.

of the different results. We choose the model where the clusters are sufficiently distinct, both in statistical terms as well as in terms of allowing a meaningful economic interpretation. As we will discuss below, 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: a LOW-ATTACHED as well as a HIGHLY ATTACHED cluster are characterized by low and high levels of attachment to the labor market, respectively, with high persistence in the corresponding states; a MOBILE + LOW-ATTACHED and a MOBILE + HIGHLY ATTACHED cluster are characterized by a much higher level of mobility together with low and high levels of attachment to the labor market, respectively; and, finally, a cluster of RETIRING, where retirement is the predominant state.

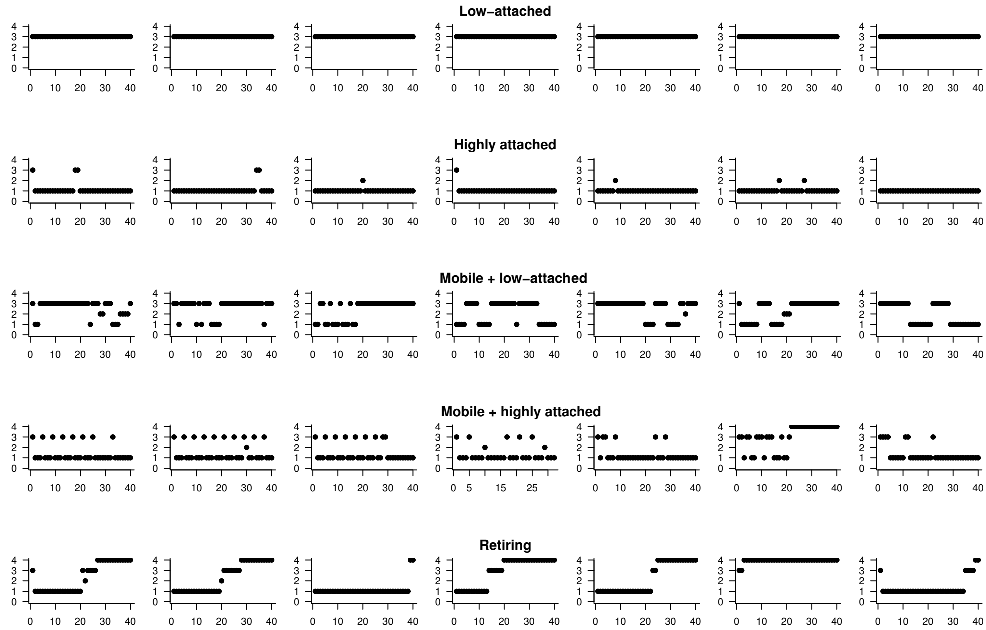

Figure 1: Employment profiles of typical cluster members within each cluster, showing the 10th, 25th, 50th, 70th, 100th, 200th and 350th highest classification probabilities. 1 = employed, 2 = sick, 3 = out of labor force, 4 = retirement.

In a six-cluster model, the distinctions between different clusters are less clear. Therefore, in the following, we concentrate on the five-cluster solution chiefly because this solution led to meaningful interpretations from an economic point of view.

4.2 Posterior Classification

Individuals are assigned to the five clusters of career-patterns using the posterior classification probabilities $t_{ih}(\boldsymbol{\theta}_5) = \Pr(S_i = h|\mathbf{y}_i, \boldsymbol{\theta}_5)$ given by eq. (8) in the Appendix. 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 in such a way that $\hat{t}_{i,\hat{S}_i} = \max_h \hat{t}_{ih}$. The closer $\hat{t}_{i,\hat{S}_i}$ is to 1, the higher is the segmentation power for individual i.

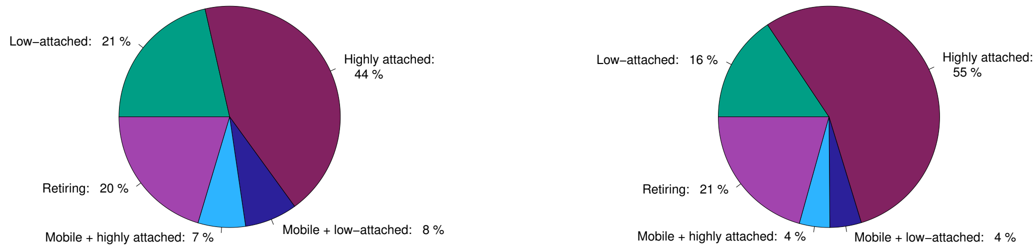

Figure 2: Group sizes for the five cluster solution. The cluster sizes are calculated based on the posterior classification probabilities. Left hand side: workers experiencing plant closure (displaced); right hand side: workers not experiencing plant closure (controls)

To obtain a first understanding of the transition patterns in the various clusters, typical group members are selected for each cluster and their individual time series are plotted in Figure 1. The career patterns are fairly similar within each cluster, but very different across clusters.

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 the LOW-ATTACHED, 44 % to the HIGHLY ATTACHED, 8 % to the MOBILE + LOW-ATTACHED, 7 % to the MOBILE + HIGHLY ATTACHED, and 20 % to the RETIRING cluster.

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 $\boldsymbol{\vartheta}_h = (\boldsymbol{\pi}_h, \boldsymbol{\xi}_{h1}, \ldots, \boldsymbol{\xi}_{h10})$ for $h = 1, \ldots, 5$. For all workers in our sample, the transition process starts with the shock of loosing employment 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 for selected transition probabilities of particular interest. In particular, the columns of Figure 4 display the probabilities of following events: persistence in the employment state (i.e. $j = 1 \rightarrow k = 1$), transition from employment to out of labor force (i.e. $j = 1 \rightarrow k = 3$), transition from out of labor force back to employment (i.e. $j = 3 \rightarrow k = 1$), and transition from employment to retirement (i.e. $j = 1 \rightarrow k = 4$). For each of the corresponding transition probabilities $\xi_{hy,jk}$, the marginal posterior distribution $p(\xi_{hy,jk}|\mathbf{y})$ is represented for each of the five clusters by a sequence of ten box plots of the corresponding MCMC draws, over the yearly distance to plant closure $y = 1, \ldots, 10$.

The 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 2.

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 2: Posterior expectations $\mathrm{E}(\pi_{h,k}|\mathbf{y})$ and, in parenthesis, posterior standard deviations $\mathrm{SD}(\pi_{h,k}|\mathbf{y})$ 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 $\mathrm{E}(\xi_{hy,jk}|\mathbf{y})$ and, in parenthesis, posterior standard deviations $\mathrm{SD}(\xi_{hy,jk}|\mathbf{y})$ of selected transition probabilities $\xi_{hy,jk}$ for selected years y in the various clusters. 1 = employed, 2 = sick, 3 = out of labor force, 4 = retirement.

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 $t = 3, \dots, 40$ recursively from the sequence of cluster-specific time-inhomogenous transition matrices by $\boldsymbol{\xi}_{h,1\to t} = \boldsymbol{\xi}_{h,1\to(t-1)}\boldsymbol{\xi}_{hy}$. 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 to plant closure.[2]

4.4 Understanding the Clusters

In this subsection we present a synthesis of posterior inference in Figure 1 to Figure 4 and Table 2 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.

Highly attached is the largest cluster 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 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 loose their job, they have quite a good chance to move back into employment within one quarter, however, with increasing distance to plant closure, the chance declines and is as small as 7.1% after 10 years.

Workers in the Low-attached cluster, the second largest cluster covering about 21% of the sample, are less successful than 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. The pattern in the first quarter after plant closure is similar for workers in the Mobile + low-attached and Mobile + highly attached clusters. Workers in these three clusters obviously suffer from the plant closure, at least in the short run.

What distinguishes workers in Low-attached from Mobile + low-attached and Mobile + highly attached is the subsequent transition behavior. Most strikingly, among Low-attached workers the chance of moving from out of labor force back into employment is extremely low in the years following plant closure and even decreases, being equal to only 1.3% five and 1% ten years after plant closure. In contrast, workers in Mobile + low-attached and Mobile + highly attached recover more easily from job displacement. Members of Low-attached hardly ever move back into employment after having lost their job due to plant closure. However, in the unlikely event, that workers in this cluster find a job, they have quite a good chance to keep their job, and this chance is larger than for Mobile + low-attached and Mobile + highly attached.

While Mobile + low-attached and Mobile + highly attached are similar to Low-attached immediately after plant closure, they show a subsequent transition pattern between out of labor force and employment that is quite different. Both clusters have about the same probability of remaining employed, which is nearly constant over time and, on average, equal

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

to 82%. They have a similar transition pattern from employment back into out of labor force, which again is nearly constant over time and is, on average, equal to about 15%. Members in both clusters have a good chance to move back into the labor market after plant closure, but they are at a high risk to loose their job again. These two clusters suffer from an intrinsically high risk of being out of labor force that appears to be unrelated to plant closure.

The main distinction between Mobile + low-attached and Mobile + highly attached is how the transition probability from out of labor force back into employment evolves with distance to plant closure. For workers in Mobile + highly attached the chance of moving back into the labour market is higher than in Mobile + low-attached 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 the typical members in Figure 1.

While Mobile + low-attached is similar to Mobile + highly attached in several respects, it is mainly the transition pattern from out of labor force back into employment that leads to career paths that are quite distinctive from Mobile + highly attached. Evidently, for Mobile + low-attached the transition probability from out of labor force back into employment is much smaller than in Mobile + highly attached and even declines over distance to plant closure. The corresponding transition probability is only 15% five years and as small as 7.8% ten years after plant closure. Workers in this cluster also switch between employment and being out of labor force; however, they have a much higher risk to remain out of labor force than workers in Mobile + highly attached. As a consequence, this leads to much longer spells of being out of labor force than for workers in Mobile + highly attached, where this duration is very short, see again Figure 1.

Workers in the Retiring cluster are more successful than Low-attached, Mobile + low-attached, and Mobile + highly attached to find a job in the first quarter after plant closure (42.2%), but less successful than Highly attached. This is the only cluster where immediate transition into retirement after plant closure happens with positive probability (2.7%), whereas this probability is practically 0 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 than workers in all other clusters. In addition, this cluster is characterized by an increasing transition probability from employment into retirement which is as large as 18.7% ten years after plant closure. For all other clusters, this probability practically remains zero. As a consequence, the probability to remain employed, which is relatively high in the first five or six years after plant closure, declines in later years and is the smallest among all clusters (72.2%) after 10 years.

The importance of using a time-inhomogeneous rather than a time-homogeneous Markov process for our application can be best seen in Figure 3 where the transition matrices change over time in all clusters. 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.

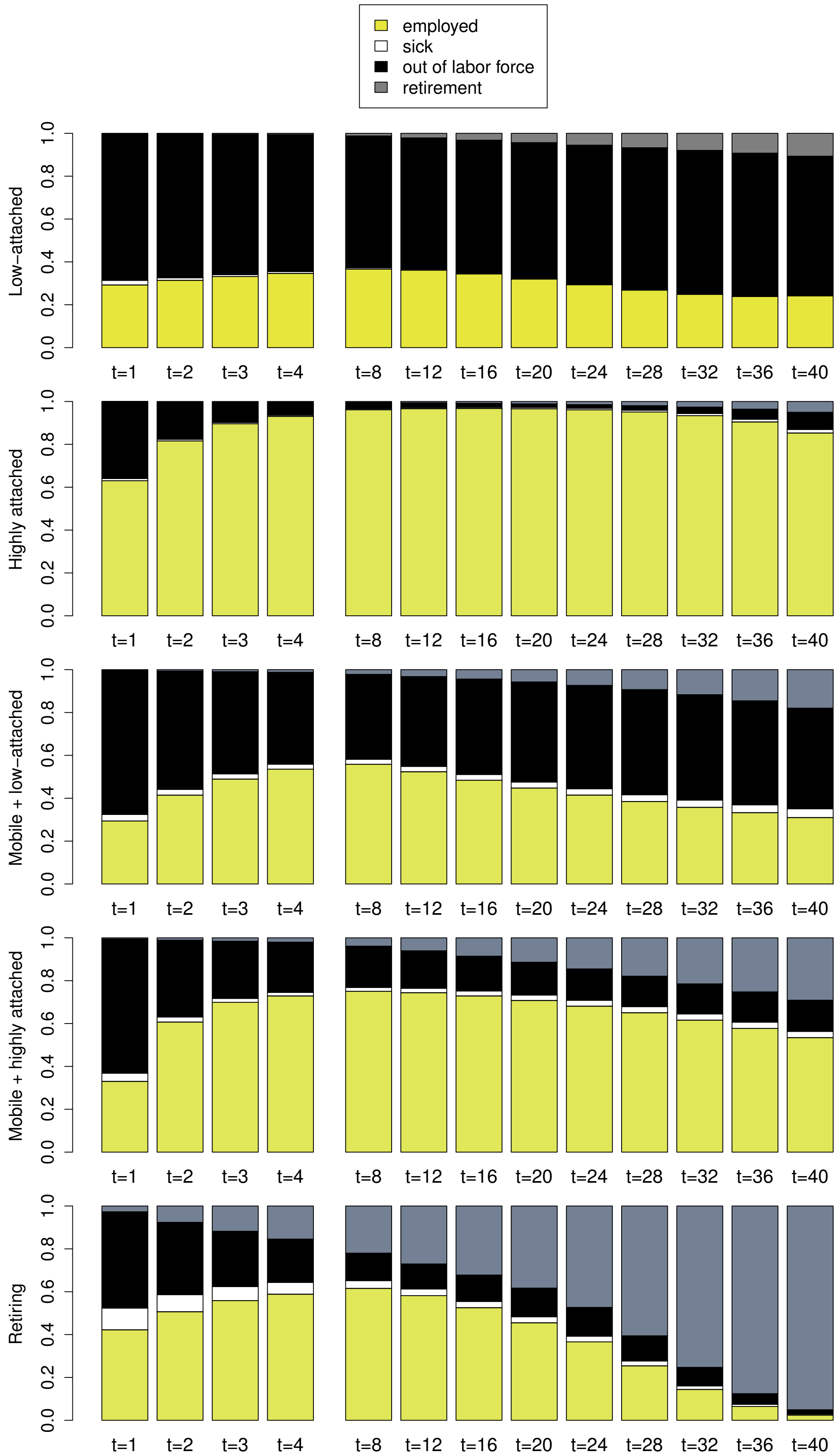

Figure 3: Posterior expectation of the distribution $\boldsymbol{\pi}_{h,t}$ over the 4 states (1 = employed, 2 = sick, 3 = out of labor force, 4 = retirement) after a period of t quarters in the various clusters (workers experiencing plant closure).

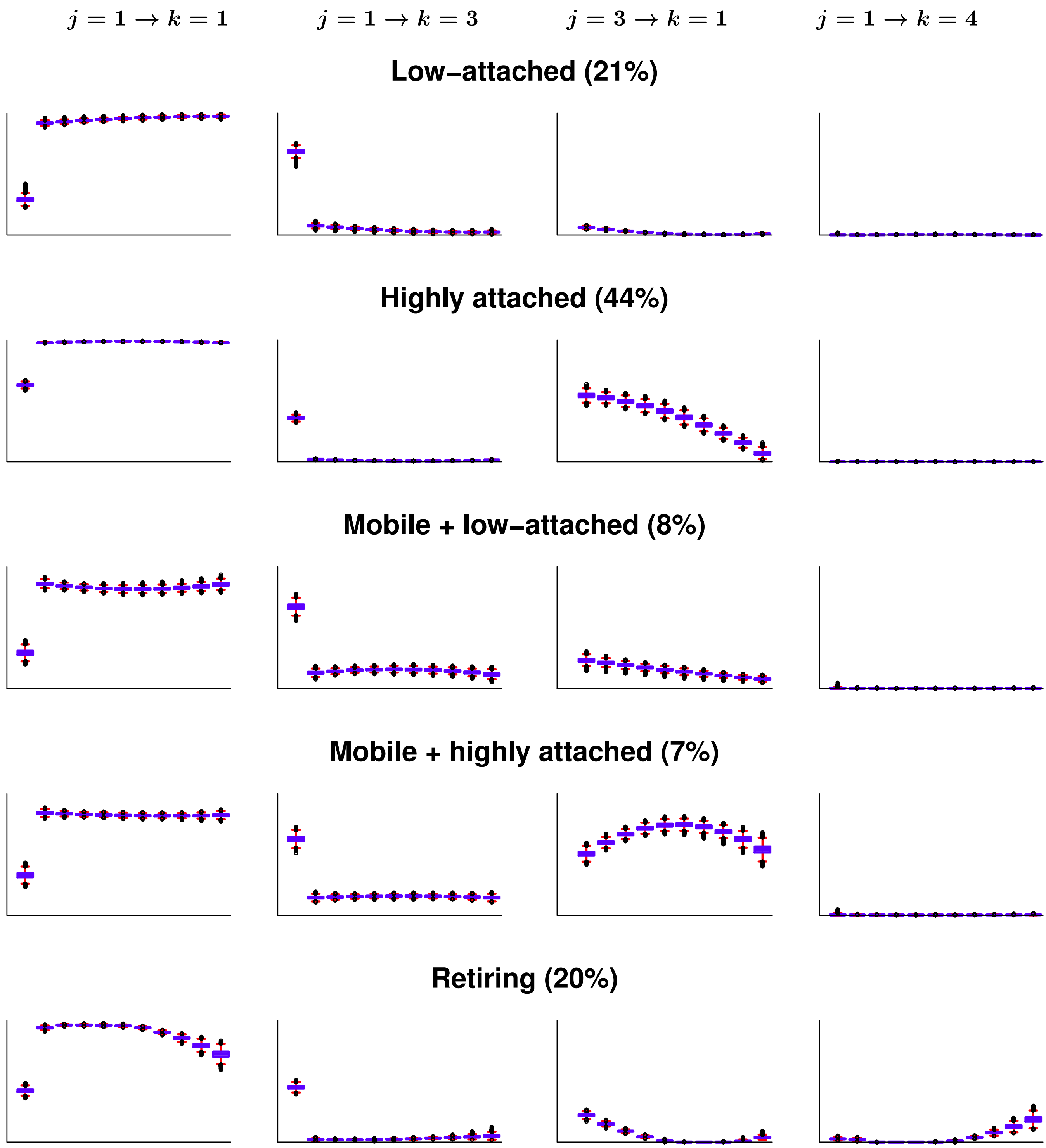

Figure 4: Visualization of the posterior distribution of 4 selected time-varying transition probabilities from state j to state k in the various clusters, obtained by time-inhomogeneous Markov chain clustering. The first box plot in columns 1, 2 and 4 displays the posterior distribution of the state probability $\pi_{h,k}$ at the end of the first quarter after plant closure for each cluster h. The remaining 10 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, 3 = out of labor force, 4 = retirement.

	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 3: 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.

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 are setting out to 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 Low-attached cluster than white collar workers?

The mixture-of-experts approach allows to answer these and similar questions, since we specify the prior probability of an individual to belong to a certain cluster by the multinomial logit (MNL) model given in equation (4). The regression framework 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.

Bayesian inference for the regression parameters $\boldsymbol{\beta}_h$ in the MNL model is summarized in Table 3, which reports the posterior expectation and the posterior standard deviation of all regression parameters relative to the baseline, which is equal to LOW-ATTACHED.

To visualize the main results, Figure 5 shows to which extent the prior probabilities of belonging to each of the five clusters are related to individual covariates; see also Table 4. For this evaluation, all other control variables are set to their mean values observed in the sample. The prior probability that a worker with certain predetermined characteristics belongs to any cluster is computed for all MCMC draws and the reported values are the average over all the MCMC draws. The diagrams can, therefore, be interpreted as giving the prior probability that a worker belongs to one of these five clusters based solely on his known characteristics before plant closure.

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

With respect to age at the time of plant closure, we see in Figure 5, panel (b), as well as in Table 4 that all workers younger than 45 years have similar prior probabilities to belong to the various clusters. Not surprisingly, young workers have a low probability to belong to the RETIRING cluster, but this probability strongly increases with age. While individuals with higher ages often belong to the RETIRING cluster, their probability of being in HIGHLY ATTACHED is reduced. For the oldest group, aged 50-55, the probability to be in the RETIRING cluster is particularly high (77%), while the probability to belong to HIGHLY ATTACHED is negligible. The probability to belong to MOBILE + HIGHLY ATTACHED is practically independent of age and the probability of belonging to LOW-ATTACHED is slightly decreasing with age.

Work experience is less strongly related to group membership than age; see Figure 5, panel (c), and Table 4. We see that the five clusters are most evenly distributed among individuals with low levels of work experience. There is not much variation in cluster membership for individuals with low levels of experience, while at high levels of experience HIGHLY ATTACHED and MOBILE + LOW-ATTACHED dominate. In particular, higher experience levels are correlated with higher probability to belong to HIGHLY ATTACHED and lower probability to belong to MOBILE + LOW-ATTACHED. Interestingly, the probability of belonging to RETIRING is practically independent of the amount of work experience. The pattern of distribution of cluster membership by tertiles of pre-displacement income resembles that of experience; see Figure 5, panel (d), and Table 4.

Figure 5, panel (e) and (f), as well as Table 4 show that group membership also varies with the size and industry affiliation of the firms from which workers are displaced. The groups with the largest LOW-ATTACHED portion are workers from small firms and from the service sector. The largest portion of the MOBILE + HIGHLY ATTACHED cluster is exhibited by the workers of medium size firms and workers from seasonal business outside of hotel and construction.

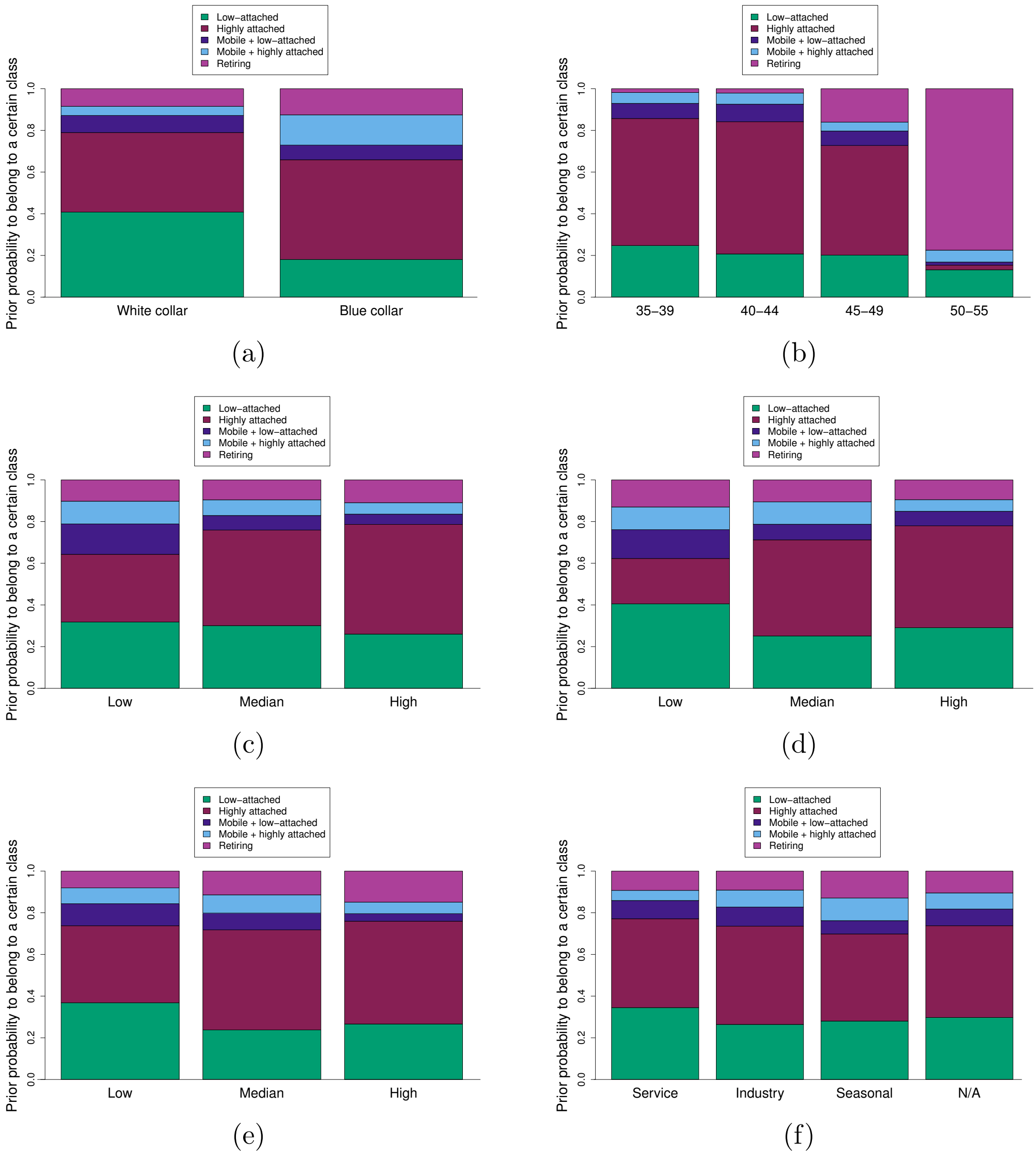

Figure 5: Impact of each covariate on the prior probability of a worker to belong to a certain cluster: (a) occupational state, (b) age, (c) experience, (d) income at time of plant closure, (e) firm size, (f) firm's economic sector (for each single covariate, all other covariates are set to their mean values observed in the sample).

	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 $\leq$ 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 $\geq$ 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 $\leq$ 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 4: Displaced persons: Prior cluster probabilities for a single covariate. All other control variables 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 that are described by the five separate clusters, we turn to a comparison of the careers of displaced workers with a 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. To evaluate the counterfactual career trajectories for each cluster, we perform a posterior classification of controls based on the clustering model that was 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.

The selection of the control group as a weighted sample with similar pre-displacement characteristics as the displaced group has been described in Section 2. The weighting procedure ensures that displaced and controls are similar with respect to the covariates, which determine prior group membership through the mixture-of-experts model specified in equation (4). The only feature that distinguishes 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 $\boldsymbol{\pi}_h$ of displaced workers at the end of the first quarter, with a very high rate of being out of labor force for practically all clusters. We thus have to take this event into account when simulating the career trajectories of control group members.

Our main modelling assumption is that the state distribution $\boldsymbol{\pi}_h$ at the end of the first quarter incorporates the entire effect of this shock. This means that the subsequent transition behaviour is independent from whether a person in this cluster experienced plant closure or not. The subsequent transition behaviour is characterized by the sequence of cluster-specific time-inhomogenous transition matrices $\boldsymbol{\xi}_h = (\boldsymbol{\xi}_{h1}, \ldots, \boldsymbol{\xi}_{h,10})$ and the person's given state at the end of the first quarter. 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 y_{i0} is employment, i.e. $y_{i0} = 1$, also for all controls, the first transition of the controls is described by a row vector $\boldsymbol{\pi}^c_h = (\pi^c_{h,1}, \ldots, \pi^c_{h,4})$ being different from $\boldsymbol{\pi}_h$ and containing the probability distribution over all states at the end of the first quarter after plant closure for the controls. Our assumption implies that beyond the first quarter, the transition matrices $\boldsymbol{\xi}_{h1}, \ldots, \boldsymbol{\xi}_{h,10}$, which were estimated from the displaced sample, can be used to classify the controls into the five clusters.

Based on this cluster model, the cluster assignment for control person i with observed individual time series denoted by $\mathbf{y}^c_i$ is performed by computing the posterior distribution $t^c_{ih}(\boldsymbol{\theta}_5) = \Pr(S^c_i = h|\mathbf{y}^c_i, \boldsymbol{\theta}_5)$ of the class indicator S^c_i over the 5 clusters by means of Bayes' rule:

$$t^c_{ih}(\boldsymbol{\theta}_5) \propto p(\mathbf{y}^c_{i,-1}|y^c_{i1}, \boldsymbol{\xi}_h) p(y^c_{i1}|S^c_i = h, \boldsymbol{\pi}^c_h) \Pr(S^c_i = h|\boldsymbol{\beta}_2, \ldots, \boldsymbol{\beta}_H, \mathbf{x}^c_i), \quad h = 1, \ldots, 5. \tag{6}$$

In (6), $p(\mathbf{y}^c_{i,-1}|y^c_{i1}, \boldsymbol{\xi}_h)$ is the clustering kernel based on a time-inhomogeneous first order Markov chain as introduced in (2), whereas the cluster-specific state distribution $\boldsymbol{\pi}^c_h = (\pi^c_{h,1}, \ldots, \pi^c_{h,4})$ for controls at the end of the first quarter after (potential) plant closure gives:

$$p(y^c_{i1}|S^c_i = h, \boldsymbol{\pi}^c_h) = \prod_{k=1}^{4} \left(\pi^c_{h,k}\right)^{C_{i,k}},$$

with $\pi^c_{h,k} = \Pr(y^c_{i1} = k|S^c_i = h)$ and $C_{i,k} = I\{y^c_{i1} = k\}$ being an indicator for a non-displaced worker's initial state. $\Pr(S^c_i = h|\boldsymbol{\beta}_2, \ldots, \boldsymbol{\beta}_H, \mathbf{x}^c_i)$ is the prior class assignment distribution introduced in (4), which is based on the individual characteristics $\mathbf{x}^c_i$ 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 panel, we use the MCMC draws obtained for the displaced persons to assign the individuals from the control panel to the five clusters of career patterns during an MCMC-type algorithm. Only the cluster-specific state distributions at the end of the first quarter are estimated by sampling $\boldsymbol{\pi}^c_h$ for each cluster from a Dirichlet distribution, $\boldsymbol{\pi}^c_h|\mathbf{S}^c, \mathbf{y} \sim \mathcal{D}\left(g_{0,1} + C^h_1, \ldots, g_{0,4} + C^h_4\right)$, where $C^h_k = \sum_{i:S_i=h} C_{i,k}$ is the total number of (non-displaced) workers in cluster h being in state k at the end of the first quarter after potential plant closure and $\boldsymbol{\pi}^c_h \sim \mathcal{D}\left(g_{0,1}, \ldots, g_{0,4}\right)$ follows a Dirichlet prior with hyperparameters analogous to those in Appendix A.

We assign individuals from the control panel using the posterior expectation $\hat{t}^c_{ih} = \mathrm{E}(t^c_{ih}(\boldsymbol{\theta}_5)|\mathbf{y}^c_i)$. $\hat{t}^c_{ih}$ is estimated by evaluating and averaging $t^c_{ih}(\boldsymbol{\theta}_5)$ 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 panel of displaced workers and the 15 000 MCMC draws of $\boldsymbol{\pi}^c_h$ obtained for the panel of controls.

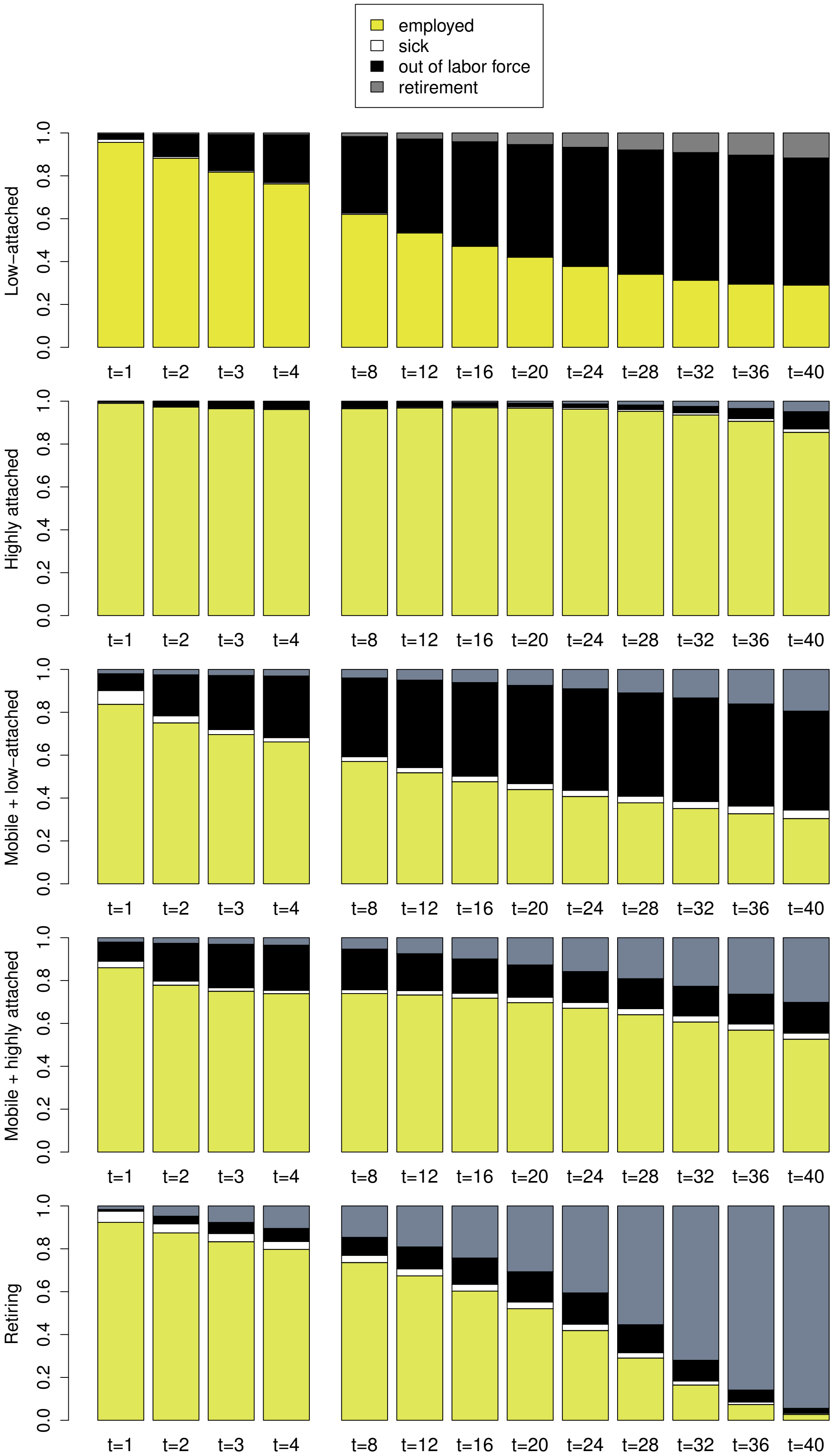

Figure 6: Posterior expectation of the distribution $\boldsymbol{\pi}^c_{h,t}$ over the 4 states ($1 =$ employed, $2 =$ sick, $3 =$ out of labor force, $4 =$ retirement) after a period of t quarters in the various clusters (control group).

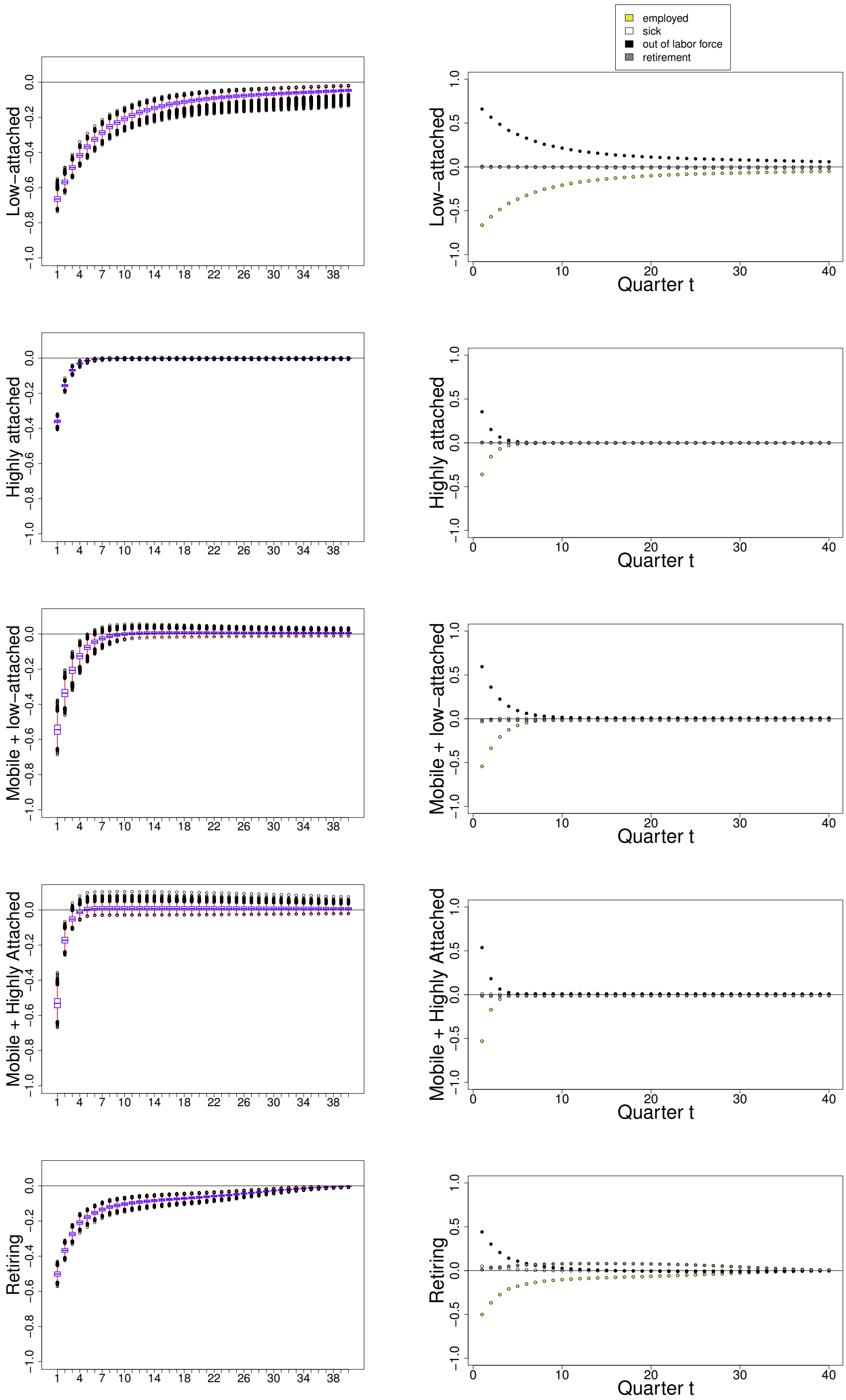

Figure 7: Analysing the difference $\pi_{h,k,t} - \pi^c_{h,k,t}$ in the probability to be in state k between persons experiencing plant closure ($\pi_{h,k,t} = \Pr(y_{it} = k|S_i = h)$) and controls ($\pi^c_{h,k,t} = \Pr(y^c_{it} = k|S^c_i = h)$) for the five clusters. Left hand side: posterior distribution of the difference $\pi_{h,1,t} - \pi^c_{h,1,t}$ in the probability to be in state "employed" between persons experiencing plant closure and controls; right hand side: posterior expectation of the difference $\pi_{h,k,t} - \pi^c_{h,k,t}$ in the probability to be in state k between persons experiencing plant closure and controls ($1 =$ employed, $2 =$ sick, $3 =$ out of labor force, $4 =$ retirement).

Each worker from the control panel 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 displaced and control group across clusters is shown in Figure 2 in the top and bottom graph, respectively. The figure shows that in absence of the plant closure event the cluster Highly attached would be considerably larger. The size of the Retiring cluster 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 $\boldsymbol{\pi}_{h,t}^c = \boldsymbol{\pi}_h^c \boldsymbol{\xi}_{h,1\to t}$ over distance t to plant closure for the control group, where the transition matrix $\boldsymbol{\xi}_{h,1\to t}$ 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 to 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 rates of convergence occuring in Highly attached and Mobile + highly attached and the lowest rate occuring in 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 labour 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 model-based clustering to explicitly address unobserved heterogeneity in reaction to loosing 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 transitions 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 the prior probability to belong to a certain cluster to depend on a set of covariates via a multinomial logit model.

For the plant closure data, this clustering procedure provides us with 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 can cope quite easily with a job loss and those who suffer large losses over extended periods of time. It turns out that around 50% of workers are highly attached after a plant closure, whereas 30% are low-attached – i.e. suffer large employment losses – and 20% belong to a group which takes early retirement as an option.

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.

Acknowledgements

The research was funded by the Austrian Science Fund (FWF): S10309-G16 (NRN "The Austrian Center for Labor Economics and the Analysis of the Welfare State") and the CD Laboratory "Ageing, Health and the Labor Market". Thanks to Oliver Ruf, Guido Schwerdt and Bernhard Schmidpeter for help with the data.

References

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

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.

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.

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

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.

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.

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.

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.

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.

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.

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 the unknown parameters $\boldsymbol{\theta}_H$ 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

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

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

Concerning the prior distribution $p(\boldsymbol{\theta}_H)$, we assume prior independence between the parameters $(\boldsymbol{\beta}_2, \ldots, \boldsymbol{\beta}_H)$, fully specifying the mixture-of-experts model, and the class-specific parameters $(\boldsymbol{\vartheta}_1, \ldots, \boldsymbol{\vartheta}_H)$ of the clustering kernel. All H parameter vectors $\boldsymbol{\beta}_h$ of regression coefficients 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 $N^0 = (N^0_{jk})$, where for each state $k = 1, \ldots, 4$ the element of the first row is equal to $N^0_{1k} := \#\{y_{i1} = k \text{ for some person } i\}$, and equal to 0 in the second and the third row (i.e. $N^0_{jk} = 0$ for $j = 2, 3$). Furthermore, we define for each year $y = 1, \ldots, 10$ the 3×4 empirical transition count matrix $N^y = (N^y_{jk})$ with elements

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

for $j = 1, 2, 3$ and $k = 1, 2, 3, 4$. For each $y = 0, 1, \ldots, 10$, we define the empirical transition matrices $\tilde{N}^y := (N^y_{jk}/r^y_j)$, where $r^y_j := \sum_{k=1}^{4} N^y_{jk}$ 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 the posterior distribution $p(\boldsymbol{\theta}_H|\mathbf{y})$, 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 $(\Pr(S_i = 1|\mathbf{y}_i, \boldsymbol{\theta}_H), \ldots, \Pr(S_i = H|\mathbf{y}_i, \boldsymbol{\theta}_H))$ given by:

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

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

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). 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 $\boldsymbol{\vartheta}_h = (\boldsymbol{\pi}_h, \boldsymbol{\xi}_{h1}, \ldots, \boldsymbol{\xi}_{h10})$ 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 $\boldsymbol{\xi}_{hy,j\cdot}$ are sampled row-by-row from a total of $30H$ Dirichlet distributions:

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

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 $\boldsymbol{\xi}_{hy,j\cdot}$ 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 $\boldsymbol{\xi}_{hy,j\cdot}$ is normalized by $\boldsymbol{\xi}_{hy,j\cdot} / \sum_{k=1}^4 \xi_{hy,jk}$ 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 $\mathbf{S}_0$. However, 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`.