When Is The Best Time To Give Birth?

by

Sylvia FRÜHWIRTH-SCHNATTER
Christoph PAMMINGER
Andrea WEBER
Rudolf WINTER-EBMER*)

Working Paper No. 1409

June 2014

Supported by the Austrian Science Funds

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

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

Corresponding author: rudolf.winterebmer@jku.at
Phone: +43-732-2468-8236

When Is The Best Time To Give Birth?

Sylvia Frühwirth-Schnatter* Christoph Pamminger† Andrea Weber‡
Rudolf Winter-Ebmer§

June 2014

Abstract

Using Bayesian Markov chain clustering analysis we investigate career paths of Austrian women after their first birth. This data-driven method allows characterizing long-term career paths of mothers over up to 19 years by transitions between parental leave, non-employment and different forms of employment. We, thus, classify women into five cluster-groups with very different long-run career costs of childbearing. We model group membership with a multinomial specification within the finite mixture model. This approach gives insights into the determinants of the long-run family gap. Giving birth late in life may lead very diverse outcomes: on the one hand, it increases the odds to drop out of labor force, and on the other hand, it increases the odds to reach a high-wage career track.

Keywords: fertility, timing of birth, family gap, Transition Data, Markov Chain Monte Carlo, Multinomial Logit, Panel Data

*Department of Finance, Accounting, and Statistics, Vienna University of Economics and Business
†Department of Applied Statistics, Johannes Kepler University Linz
‡Department of Economics, Universität Mannheim and WIFO, Vienna
§Department of Economics, Johannes Kepler University Linz and IHS, Vienna

1 Introduction

Childbearing is typically associated with substantial career costs for women, either in terms of lower wages, periods of parental leave or even longer career breaks. Many studies have documented the so-called 'family gap': differential earnings paths between mothers and childless women with significant gaps opening up right after the birth of the first child. A smaller part of the literature has examined the role of the mother's age at birth focusing on specific advantages or disadvantages of giving birth early or late in life or in the professional career.

In this study we focus specifically on the timing of the first birth of a women. This is an important topic, given the demographic trends towards longer pre-labor-market educational spans and increasingly volatile and uncertain career paths of young women.

Unlike most of the literature, our main analysis is not restricted to the impact of childbearing on earnings profiles, but we are interested in general career paths after childbearing. We focus on a sample of women at the birth their first child and classify their subsequent labor market behavior. Using detailed data from administrative registers, we can follow their careers between eight and nineteen years. Overall, careers of young mothers are characterized by frequent transitions in and out of employment, maternity leave, or unemployment, as well as high levels of mobility between earnings groups.

Our goal in this paper is twofold. First, we want to identify specific career patterns that characterize the employment, and earnings paths of mothers after the birth of their first child. Thereby we consider transitions from parental leave back into employment, mobility between different earnings groups, and subsequent career interruptions either due to unemployment or to maternity spells. Our specific interest is in heterogeneity of career transition patterns across mothers. We use Bayesian Markov chain clustering analysis, a purely data-driven method, which isolates five different clusters of mothers with similar career paths: a cluster of women who return to their jobs quickly after childbirth and move to high-paying jobs on a steep earnings profile, a cluster of women who decide to work in part-time jobs, a cluster of mothers who take an extended family break before returning to work, a cluster of women who drop out of the labor force after the birth of the first child, and a cluster of mothers with highly mobile careers switching in and out of employment multiple times.

Our second goal is to find out whether characteristics of the mother determine the type of career pattern she follows after the birth of her first child. We will focus on the correlation with education, earnings and the type of job prior to birth, but concentrate especially on the mother's age and level of labor market experience at the first birth. To model these correlations, we use a clustering approach based on finite mixture models, which models the prior probability to belong to a certain cluster through a multinomial logit model as being dependent on individual characteristics following the method developed in Frühwirth-Schnatter et al. (2012).

Our paper links to the literature that has looked at a general family gap, by comparing earnings profiles of mothers and non-mothers after birth (Korenman and Neumark (1992), Waldfogel (1998) for the US, Ejrnaes and Kunze (2013), Schönberg and Ludsteck (2014) for Germany or Simonsen and Skipper (2006) for Denmark) and typically finds wage gaps of around 6% for one child and up to 15% for two children. Corresponding family gaps are much smaller in Nordic countries, e.g. Denmark. [1] In contrast to this literature, we only focus on women, who have decided to have a child. Our cluster approach classifies different career patterns that start with the birth of the first child and are modeled as Markov transition processes. Instead of comparing mothers with non-mothers, our approach is thus comparing mothers across different cluster-groups. The steady states of the cluster-specific transition processes can be interpreted as the long-run career potential, which women in the respective cluster group are able to reach after the birth of her child. Therefore we analyze the convergence to the steady state in the different cluster groups and interpret the time it takes to reach the full potential as the cluster-specific "family gap".

We further link to studies examining the effects of the timing of childbearing and the change in family-gap with the age of the mother. One strand of this literature has traditionally concentrated on presumed disproportionate difficulties for teenage mothers (Geronimus and Korenman (1992) or Chevalier and Viitanen (2003)). The most recent studies (Hotz et al., 2005) find that differences between teen parents and older parents are minor.

Studies explicitly considering the timing of childbearing (Amuedo-Dorantes and Kimmel (2005), Miller (2011), Taniguchi (1999), Herr (2012), Wilde et al. (2010)) typically find that

[1] See Bronars and Grogger (1994), Angrist and Evans (1998), Cristia (2008) or Fitzenberger et al. (2013) for effects on labor supply.

delaying childbirth reduces the cost of childbearing, in particular for more educated women. In a study on Germany, Fitzenberger et al. (2013) find that women who are older at first birth suffer larger long term employment losses. Herr (2012) points out that the measure of age at birth plays a crucial role in the comparison of after birth wage profiles of mothers. She proposes the level of labor market experience as an alternative measure. We contribute to this literature by investigating the relationship between both age and the level of labor market experience and the probability of belonging to a certain career profile.

The approach chosen in our statistical analysis is purely descriptive. Fertility decisions are closely linked with many other career decisions such as marriage, human capital accumulation, or labor supply. This makes it very difficult to credibly isolate causal pathways. The literature has proposed sources of exogenous variation in the timing of birth, based on instruments such as incidence of miscarriage, contraceptive use or failures (Miller (2011)). However, some of these instruments may be problematic, see discussion in Wilde et al. (2010), because incentives for birth control are strongest for women with the highest economic costs of childbearing; on the other hand, the probability to have a miscarriage tends to be correlated with health and social outcomes (Fletcher and Wolfe, 2008). Therefore, we concentrate on the statistical classification of career patterns for women.

2 Data

Our empirical analysis is based on data from the Austrian Social Security Data Base (ASSD), which combines detailed longitudinal information on employment and earnings of all private sector workers in Austria since 1972 (Zweimueller et al., 2009). The data also record births and spells of parental leave of mothers who have entered the labor market before their first child is born.

Our sample consists of female workers, whose labor market careers we follow after the birth of their first child. We concentrate on women which a certain attachment to the labor market before the birth of their first child, by restricting the sample to women who were employed for at least 100 days in the last year before giving birth. We focus on births between 1990 and 2000 and we restrict the age of the mother at first birth to be between 16 and 35 years old. Further, we exclude women who were civil servants or self-employed before giving birth, because for civil servants different job security provisions apply and for self-employed workers employment spells are often difficult to measure due to free time arrangements. We exclude non-Austrian citizens because for these earlier working careers might be censured in our data set. The final sample consists of $N = 231\,095$ female workers.

To characterize long-run employment careers after childbirth we organize the data into a panel of yearly observations: starting the first time period six months after the birth of the first child we track the labor market status of a women in annual intervals: from the sixth month (t=0) to the eighteenth month, etc.[2] Given the time frame of the data, we observe women in our sample for 8 to 19 time periods; the median number of observations per women is 14 years. As our employment data are from social security records, whose aim is to document claims towards old age pensions and other social security benefits, data quality is exceptionally high: all employment spells with corresponding wages are precisely recorded. The downside is that information not relevant for social insurance issues is sparse. Most importantly, we lack data on working time, and monthly earnings are top-coded, which applies to roughly three percent of the data points.

To model employment careers we proceed by constructing for each person a time series of their employment and earnings status. Specifically the annual values are categorized to take the following five values: Category 'K' represents periods of parental leave benefit receipt following the birth of a child. Category '0' corresponds to economic inactivity with zero labor earnings, i.e. unemployment or out-of-labor force. Employment spells are coded by three distinct categories representing tertiles of the earnings distribution of females in the corresponding calendar year. Using this strategy we can differentiate between very low incomes, presumably part-time work, (category '1'), medium-wage employment, potentially full-time, (category '2') and high-wage employment (category '3'). This crude classification, while not necessarily accurate in all cases – i.e. category 1 might be full-time, but very low paid employment – allow us to overcome both the problem of missing hours of work and top-coded earnings. Based on these time series with five employment and earnings categories, we are going to model labor market careers by analyzing transitions between these discrete states.

[2]The choice of the timing is due to generous parental leave regulations in Austria. For details see below.

To study factors that have an impact on the different career patterns after the birth of the first child, we focus on variables which are pre-determined at the time of the first birth, see Table 3. Specifically, control variables include the mother's age at first birth, her years of labor market experience, education, and marital status. Moreover, we control for the job-type, tenure, and the monthly wage in the last job before birth, as well as average monthly earnings during the last 5 years before birth.[3]

The *Austrian family policy* provides fairly generous government transfers for parents of young children. To protect the health of the mother and the child women are not allowed to work over 16 weeks around birth but they are eligible for a benefit equal to their wages. Parental leave sets in after the maternity protection. During this periods mothers receive provides a flat rate benefit and job protection. The duration of the maternity leave was extended in several reforms over the 1990's from one up to three years. Take-up of parental leave is very high in Austria, which results in long employment gaps of young mothers. See Lalive et al. (2013) for a detailed analysis of labor supply responses to the parental leave reforms.

3 Method

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 of five states, labeled by $\{K, 0, 1, 2, 3\}$, and is observed for N individuals $i = 1, \dots, N$ over T_i discrete time periods. For each individual i, we model the state of y_{it} in period t to depend on the past state $y_{i,t-1}$ of the outcome variable in a first order model. To capture the presence of unobserved heterogeneity in the dynamics in our discrete-valued panel data, we follow Frühwirth-Schnatter et al. (2012) who introduced mixtures-of-experts Markov chain clustering for this type of time series.

3.1 Mixtures-of-Experts Markov Chain Clustering

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 recent review. Within each class,

[3]Top-coded wages are recorded at the top-coding limit. For computing average earnings in the last 5 years, only periods with positive wages are considered.

say h, all time series can be characterized by the same data generating mechanism, also called a clustering kernel, which is defined in terms of a probability distribution for the time series $\mathbf{y}_i = \{y_{i0}, \dots, y_{i,T_i}\}$, depending on an unknown class-specific parameter $\boldsymbol{\xi}_h$. A latent group 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. $p(\mathbf{y}_i|S_i, \boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_H) = p(\mathbf{y}_i|\boldsymbol{\xi}_{S_i})$.

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 $p(\mathbf{y}_i|\boldsymbol{\xi}_h) = \prod_{t=1}^{T_i} p(y_{it}|y_{i,t-1}, \boldsymbol{\xi}_h)$ is formulated conditional on the first observation y_{i0}. For discrete-valued time series, persistence is captured by assuming that $\mathbf{y}_i$ follows a time-homogeneous Markov chain of order 1. Hence, Markov chain clustering uses a Markov chain model with class-specific transition matrix $\boldsymbol{\xi}_h$ as clustering kernel, i.e.:

$$p(\mathbf{y}_i|\boldsymbol{\xi}_h) = \prod_{j=1}^{5} \prod_{k=1}^{5} \xi_{h,jk}^{N_{i,jk}}, \tag{1}$$

where $\xi_{h,jk} = \Pr(y_{it} = k|y_{i,t-1} = j, S_i = h)$ and $N_{i,jk} = \#\{t \in \{1, \dots, T_i\}|y_{i,t-1} = j, y_{it} = k\}$ is the number of transitions from state j to state k observed in time series $\mathbf{y}_i$ for $j, k = 1, \dots, 5$. Each row $\boldsymbol{\xi}_{h,j\cdot} = (\xi_{h,j1}, \dots, \xi_{h,j5})$ of the matrix $\boldsymbol{\xi}_h$ represents a probability distribution over the states $\{K, 0, 1, 2, 3\}$, i.e. $\sum_{k=1}^{5} \xi_{h,jk} = 1$. Previous applications of this approach to clustering individual wage careers in the Austrian labor market include Pamminger and Frühwirth-Schnatter (2010), Pamminger and Tüchler (2011), and Frühwirth-Schnatter et al. (2012).

In standard model-based clustering it is assumed that each individual i has the same prior probability to belong to a certain latent class, regardless of its specific characteristics. Since this assumption seems to be unrealistic for labor market data, Frühwirth-Schnatter et al. (2012) allowed exogenous factors or covariates $(x_{i1}, \dots, x_{ir})$ to influence the prior class assignment distribution which is modeled as a multinomial logit (MNL) model:

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

The row vector $\mathbf{x}_i = (x_{i1} \cdots x_{ir}\, 1)$ includes a constant for the intercept, in addition to the exogenous factors or 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.

Finite mixture models with prior class assignment according to (2) have been introduced in the machine learning literature as mixture-of-experts models (Peng et al., 1996) and have been applied to model-based clustering of economic time series in Frühwirth-Schnatter and Kaufmann (2008) and Frühwirth-Schnatter et al. (2012).

3.2 Bayesian Inference

For estimation, we vary the number of clusters from $H = 2, \ldots, 6$ and use statistical criteria as well as economic interpretability to select the final cluster solution, see Subsection 4.1.

For a fixed number H of clusters, the latent group indicators $\mathbf{S} = (S_1, \ldots, S_N)$ are estimated along with the unknown group-specific parameters $\boldsymbol{\theta}_H = (\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_H, \boldsymbol{\beta}_2, \ldots, \boldsymbol{\beta}_H)$ from the data using a Bayesian approach. Practical Bayesian inference is carried out by means of the R package `bayesMCClust` which implemented the Markov chain Monte Carlo (MCMC) sampler introduced in Frühwirth-Schnatter et al. (2012), where all necessary computations are discussed in full detail.

Concerning prior choices, we assume prior independence between $\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_H$ and $\boldsymbol{\beta}_2, \ldots, \boldsymbol{\beta}_H$. All regression coefficients β_{hj} are assumed to be independent a priori, each following a standard normal distribution. The five rows $\boldsymbol{\xi}_{h,1\cdot}, \ldots, \boldsymbol{\xi}_{h,5\cdot}$ of $\boldsymbol{\xi}_h$ are independent a priori each following a Dirichlet distribution $\mathcal{D}(e_{0,j1}, \ldots, e_{0,j5})$ with (uninformative) prior parameter $e_{0,jk} = 5$.

4 Results

To identify groups of individuals with similar career patterns after the first births, we apply Markov chain clustering for 2 up to 6 groups. For each number H of groups we simulated 5 000 MCMC draws after a burn-in of 5 000 draws with a thinning parameter equal to 5 and used the remaining 1000 draws for posterior inference.[4] We started MCMC estimation by choosing initial values for the group-indicators $\mathbf{S}$ through random initial clustering by sampling S_i from $(1, \ldots, H)$ with replacement. We repeatedly use this strategy to verify that all chains converge

[4]The computing time for all 10 000 draws is approx. 50 hours for $H = 2$, 93 hours for $H = 3$, 117 hours for $H = 4$, 162 hours for $H = 5$ and 216 hours for $H = 6$ on an Intel® Core™ 2 CPU E8400 @ 3.00 GHz 2.98 GHz.

to the same posterior distribution. The results indicate that there are no remarkable differences between the different starting strategies.

4.1 Model Selection and Posterior Classification

The various model selection criteria discussed in Frühwirth-Schnatter et al. (2012) are applied to the present data to select the number H of clusters, see Figure 1. However, as expected, these criteria are not unambiguous; the AIC and BIC criterion favor the six group solution, whereas the CLC or ICL criterion favor a rather small number of clusters. On the other hand, the AWE criterion refers to a five-group solution.

As these statistical criteria do not give a clear answer, we select the number of groups based on the economic interpretation. 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 groups of career-patterns, which are characterized by level and variability of earnings as well as the frequency of transitions into and out of the labor force: a "low-wage" and a "high-wage" group characterized by quick returns to the labor market, a group of women with "late return" to the labor force as well as a "out-of-labor-force" group (OLF) and a "mobile" group. In the six-group model, the distinctions between different groups are less clear. Therefore, in the following, we concentrate on the five-cluster solution, mainly, because this solution led to more meaningful interpretations from an economic point of view.

Individuals are assigned to the five groups of career-patterns using the posterior classification probabilities $t_{ih}(\boldsymbol{\theta}_5) = \Pr(S_i = h|\mathbf{y}_i, \boldsymbol{\theta}_5)$. 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 the 1 000 thinned MCMC draws of $\boldsymbol{\theta}_5$. Each female worker is then allocated to that cluster 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. Table 2 analyzes the segmentation power by reporting the quartiles and the median of the classification probabilities $\hat{t}_{i,\hat{S}_i}$ within the various groups. Note that one minus these numbers corresponds to the misclassification risk in each group (Binder, 1978), hence the closer to one, the smaller the misclassification risk. Segmentation power varies between the clusters and is the highest for the “high-wage” cluster and the lowest for the “low-wage” and “mobile” cluster.

Furthermore, Table 2 reports the average segmentation power over all individuals which is comparably high. 3 out of 4 individuals are assigned with at least 61.56 % to their respective groups. For 1 out of 4 individuals the assignment probability amounts to at least 93.19 %, leading to a misclassification risk of at most 6.81 %.

4.2 Estimation Results

In the following, we first describe the career patterns of young mothers that are implied by the estimated transition processes for each cluster group. Then we investigate the convergence to the steady state of the Markov process, which we will interpret as the cluster-specific career potential. Second, we will describe the correlation between group membership and mother’s characteristics that are pre-determined at birth of the first child, focusing especially on the role of age and labor market experience of the mother.

4.2.1 Analyzing Career Mobility

To analyze career mobility patterns in the five different clusters we investigate for each cluster $h = 1, \dots, 5$ the posterior expectation of the group-specific transition matrix $\boldsymbol{\xi}_h$. The five group-specific transition matrices are visualized in Figure 2 using “balloon plots”[5]. Full numerical results together with standard deviations are given in Table 1. These results are based on the prior distributions introduced in Subsection 3.2.

The circles in Figure 2 are proportional to the size of the corresponding entry in the transition matrix and each row is summing to one. Based on the posterior classification probabilities of group membership, we can also compute the size of each cluster. The share of individuals in each cluster are also shown in Figure 2. Observations in our sample are relatively evenly distributed across the five clusters: 20.5 % of the persons belong to the “low-wage” cluster, 28.4 % to the “late return” group, and 14.1 % to the “out-of-labor-force” cluster, 18.4 % to the “high-wage” group and 18.7 % to the “mobile” cluster.

Graphical evidence from the balloon plots (Figure 2) highlights remarkable differences in the transition patterns across the different cluster-groups. We will now present our interpretation of

[5] They are generated with the function `balloonplot()` from the `R` package `gplots` (Jain and Warnes, 2006).

the career transition patterns that evolve after the birth of the first child in each cluster-group in turn.

Group 1 – the "low-wage"group – is the second largest group with about 20% of observations. Women in this group return to employment from maternity leave at a relatively high rate. Predominantly they transit into lower wage categories, characterized by part-time jobs or low-wage full-time jobs. The transition matrix also reveals a relatively high rate of return to maternity leave from employment, indicating multiple maternity breaks.

Group 2, the largest group covering about 28% of the sample is labeled as "late return" group, because the transition matrix indicates extended maternity breaks and delayed returns to employment. The predominant exit state from maternity leave is non-employment, which indicates that mothers in this group extend their maternity break beyond the government subsidized maternity leave period. Eventually, mothers return from the non-employment state to employment and there is indication that they move up in the earnings distribution. The transition pattern does not show high rates of return to maternity leave, indicating that mothers in this group take a long career break after having a single child but eventually catch up with their careers.

This transition pattern distinguishes group 2 from group 3 – the "out-of-labor-force" cluster and the smallest group including 14% of observations. In this group, mothers exiting maternity leave are most likely to enter non-employment and this state has a very high persistency. The probability of entering employment from any state is extremely low and if a mother manages to return to work, persistence in employment is low as well. Women in this group return to maternity leave or non-employment at a higher rate than staying employed. Thus, the transition pattern indicates that mothers in group 3 choose to become housewives with one or more kids.

Group 4 – the "high-wage" group – is characterized by high mobility towards the upper earnings groups. The transition pattern indicates that mothers in this group leave maternity relatively quickly. Their return to employment is characterized by either returning to high-paid jobs immediately or following a high-growth career trajectory.

The final group 5 labeled the "mobile" group is characterized by a high rate of transitions across all states. Group members change their status frequently between employment, maternity leave, and low-earnings employment.

The transition matrices, graphically shown for each cluster group in Figure 2, characterize career transitions. In our sample the transition process always starts with the birth of the first child and eventually converges to the steady state of the corresponding Markov chain. Figure 3 illustrates the convergence of the process and the steady state distribution for each cluster-group. The first bar in the figures for each cluster h corresponds to the initial distribution $\boldsymbol{\pi}_{h,0}$ at $t = 0$ which is estimated from observations y_{i0} for all individuals i being classified to group h. In our sample, almost all women are still on maternity leave in period 0, i.e. six months after the birth of the child. The remaining bars show posterior expectations $\mathrm{E}(\boldsymbol{\pi}_{h,t}|\mathbf{y}, \boldsymbol{\pi}_{h,0})$ of the cluster-specific distribution $\boldsymbol{\pi}_{h,t}$ after t years ($\boldsymbol{\pi}_{h,t} = \boldsymbol{\pi}_{h,0}\boldsymbol{\xi}_h^t$) as well as for the steady state.[6]

Conditional on the choice of having a child our statistical method sorts mothers into cluster groups characterized by different career patterns. The long-run distribution of earnings and labor market states in the cluster group indicates the steady state to which the labor careers of mothers converge after childbirth. Thus, the steady state of the Markov process has a natural interpretation of the "career potential" of mothers in the respective cluster-group. In order to assess the effects of having a child on labor market outcomes, we are interested in this steady state distribution as well as in the speed of convergence which tells us how long mothers take to reach their "career potential" and thus can be seen as a measure of the family gap.

Figure 3 shows that convergence to the group-specific steady states is achieved within a few years for groups 1, 3, 4, and 5. Five years after the birth of the first child the group-specific distributions are relatively close to the respective steady states and there are almost no changes in the distributions between 10 and 20 years which is the average horizon of our sample. The "low-wage" group 1 converges to employment in low- or medium- earnings category; the "out-of-labor-force" group 3 converges to non-employment; the majority of "high-wage" group 4 members are employed in high-earnings jobs in the steady state; and the "mobile" group 5 is equally split between non-employment and low- or medium-earnings jobs.

The convergence pattern in the "late return" group 2, differs from the remaining groups significantly. In the steady state the vast majority of members of this group would be employed with almost equal probabilities in low-, medium, or high-wage jobs. The convergence to this

[6] The posterior expec tation is estimated by averaging the MCMC draws of $\boldsymbol{\pi}_{h,t}$ obtained by computing $\boldsymbol{\pi}_{h,t}$ for $t = 1, \dots, 50$ for all 1 000 draws of the thinned MCMC sample of $\boldsymbol{\xi}_h$.

steady state is very slow, however. Five years after the birth of the first child, the majority of group members are still either on maternity leave or in non-employment. The distribution shifts towards higher rates of employment over the time frame of our sample, i.e. between years 10 and 20 after the first birth. But the steady state is not achieved before about 50 years, which is clearly beyond the mothers' career horizon.

If we interpret the convergence rates toward the group-specific steady state distribution in terms of the family gap in reaching the career potential, we see a clear distinction between the groups. While for the other groups (maybe least for the "high-wage"group) maternity can be seen as a temporary career interruption and the gaps close after about 5 years, mothers in group 2 suffer a permanent loss from the maternity break with respect to their potential. These mothers would have the potential to return to stable employment, but the extended maternity breaks lead to very slow return rates and convergence to the potential is not achieved within a mother's career horizon.

4.2.2 The Impact of Observables on Group Membership

After having established differences in labor market careers after the birth of the first child across the five different cluster groups of mothers, we are setting out to investigate how individual characteristics correlate with group membership. From a social policy point of view, it is interesting to understand, what characteristics of a particular women makes her more prone to fall into one or the other cluster. Moreover, our interest centers on the timing of birth: is the career adjustment after birth easier for young mothers, who have most of their labor market careers in front of them, or for women with an established career, who resume their "regular working life" after the maternity break?

To answer these questions we model the prior probability of an individual to belong to a certain cluster by the multinomial logit model specified in equation (2). The estimation results are presented using the "high-wage" cluster as baseline. Specifically, our regression framework controls for impacts of education, the type of last job, and earnings in the last job as well as average earnings over the last five years before birth. In addition, we control for changes in the institutional framework by including a set of year of birth dummies and we control for changes in preferences for maternity and labor supply across cohorts by including 6 dummies for birth cohorts of the mothers. The effect of the age of the mother at first birth is thus captured by a difference-in-difference type of setup that abstracts from cohort and time effects. Specifically, we model the age of the mother at the time of confinement by 4 age categories: below age 20, 21-25 years, 26-30 years and 31-35 years of age. Similarly we control for labor market experience in 4 categories: 0-1 years, 2-4 years, 5-10 years, and more than 10 years.

Bayesian inference for the regression parameters in this multinomial logit model is summarized in Table 4, which reports the posterior expectations and the posterior standard deviations of all regression parameters relative to the baseline.

The results show that, giving birth early in the career is to be preferred over a situation where the potential mother has already a long professional experience: a higher professional experience increases the odds that the women will end up in a "low-wage" cluster or be a "late returner" – relative to the baseline of a "high-wage" career. The length of a previous career has almost no impact on the odds to be classified as "out-of-labor-force", but reduces the odds to be classified as a "mobile"worker. This pattern is consistent with the impact of job tenure in the last job. Given professional experience, age at birth is associated with a bifurcation: on the one hand, a higher age at birth increases the odds to end up out of labor force, on the other hand, it increases the odds to be in a high-wage cluster – relative to being low-wage or a late returner.

This pattern is particularly interesting, because it shows the value of the clustering approach: higher age at birth has no linear relation to future career outcomes; it is both more likely that a women who gives birth later in life will end up in a "high-wage"cluster or even leave the labor force.

The other results are according to expectations. Higher education almost uniformly increases the probability to be classified in the "high-wage"cluster – relative to all other clusters; the same if a women had a high-wage job or a white-collar job before the pregnancy. Interestingly, single mothers tend to be clustered less likely in "low-wage"or "mobile"clusters –relative to "high-wage"clusters. Unsurprisingly, in particular their odds to "return late"or drop "out of the labor force"are significantly reduced.

To visualize our results we show in Figure 4 the influence of the mother's age at the time of first birth as well as the influence of the mother's professional experience before giving birth

(Figure 5) on the prior probabilities to belong to each of the five groups. For this exercise all other control variables are set to their mean values. All computations are based on the last 5000 MCMC draws and a thinning parameter of 5 was applied. The prior probability that a women with certain pre-birth characteristics belongs to each of the labor market career groups after birth is computed for all MCMC draws. The plotted values are the average over all MCMC draws. The graphs can, therefore, be interpreted as giving a probability that a women with given characteristics is clustered into one of these five groups.

In the bottom right graph of Figure 4, we see that mothers who are older when the first child is born, are typically classified more often as being "out of labor force"; e.g. being above thirty results in a probability to be in this group, which is twice as high as when the birth happened in the teenage years. On the other hand, the probability to be coded as a "late returner" increases up to age 30. These shifts out of the labor market are mostly at the expense of women who would otherwise be in a "low wage" track. The probability to be in a "high wage" track is not much affected by the age of giving first birth. If we additionally stratify the sample of women into three specific educational groups (compulsory education, vocational school and college education[7]), it turns out that for those with compulsory schooling "mobile"careers are most important; in particular teenage mothers fall into the "mobile"group while those in the "late return"group come mostly from mothers giving birth in their twenties. Workers with vocational schooling are predominantly found in the "late return" group with the highest prevalence again for mothers in their twenties. The majority of mothers with college degrees are found in the "high wage"group. Concerning the age of the mother, it turns out that teenage mothers are more often clustered into the "mobile" group whereas mothers above age 25 are more often classified as "out of labor force".

As can be seen in Figure 5, having more work experience, on the other hand, leads to a somewhat tighter connection with the labor market. Women who give birth after having more than ten years of work experience are much less likely to drop "out of labor force", but they tend to return later to their jobs. In terms of wages, the probability to end up in a "low wage"career increases substantially while the probability to reach a "high wage" career path diminishes. With

[7]Individuals with missing education, as well as those with high or middle school are not presented. For those with college white collar status is set to 1.

respect to different education groups, we can see that, both for women holding only compulsory or vocational schooling, the prevalence of a "mobile" career diminishes dramatically with work experience. On the other hand, the later in the career the women gives birth, the more likely is a "late return" to the labor market. For women with college education, with increasing work experience the most notable change is the increase in "late return" to the labor market, whereas the prevalence of "high wage" careers decreases only slightly.

5 Conclusions

Using a Bayesian clustering approach we have investigated career paths of women after the birth of their first child. This data-driven method allows to characterize long-term career paths over up to 19 years by transitions in and out of parental leave, non-employment and different forms of employment. Given both long-term trends as well as short-term transition rates, we can identify five groups of women of almost equal numbers: a "low-wage" cluster, a "late-return" cluster, an "out-of-labor-force" cluster, a "high-wage" cluster and a "mobile" cluster. The economic career costs of having a child are, thus, enormously heterogeneous, depending on the cluster the women belongs to.

The chosen method of transition processes has the big advantage, that we can gain insights from the career potential of these women by assessing the steady state of mothers in a respective cluster, which can be interpreted as a form of family gap. Moreover, the speed of convergence to such a steady state tells us how volatile these developments are. It turns out that for most groups a steady state – and thus a final assessment of the family gap – is reached almost after five years.

Which factors determine the career path after giving births? When we use indicators predetermined at birth, we see that both age and the length of the professional career determine cluster membership. While the previous literature mostly found that early child-bearing is detrimental to a further labor market career, we can give a more nuanced answer. Giving birth late in life may lead very diverse outcomes: on the one hand, it increases the odds to drop out of labor force, and on the other hand, it increases the odds to reach a high-wage career track.

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"). Thanks to helpful comments by Bernd Fitzenberger and seminar participants in Innsbruck.

References

Amuedo-Dorantes, C. and J. Kimmel (2005). The motherhood wage gap for women in the United States: The importance of college and fertility delay. *Review of Economics of the Household 3(1)*, 17–48.

Angrist, J. D. and W. N. Evans (1998). Children and their parents' labor supply: Evidence from exogenous variation in family size. *The American Economic Review 88*(3), pp. 450–477.

Binder, D. A. (1978). Bayesian cluster analysis. *Biometrika 65*, 31–38.

Bronars, S. G. and J. Grogger (1994). The economic consequences of unwed motherhood: Using twin births as a natural experiment. *The American Economic Review 84*(5), pp. 1141–1156.

Chevalier, A. and T. K. Viitanen (2003). The long-run labour market consequences of teenage motherhood in Britain. *Journal of Population Economics 16*, 323–343.

Cristia, J. (2008). The effect of a first child on female labor supply: Evidence from women seeking fertility services. *Journal of Human Resources 43*(3), 487–510.

Ejrnaes, M. and A. Kunze (2013). Work and wage dynamics around childbirth. *The Scandinavian Journal of Economics 115(3)*, 856–877.

Fitzenberger, B., K. Sommerfeld, and S. Steffes (2013, 1). Causal effects on employment after first birth - a dynamic treatment approach. *Labour Economics 25*, 49–62.

Fletcher, J. and B. L. Wolfe (2008). Education and labor market consequences of teenage childbearing: Evidence using the timing of pregancy outcomes and community fixed effects. *Journal of Human Resources 44*.

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 S. Kaufmann (2008). Model-based clustering of multiple time series. *Journal of Business and 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.

Geronimus, A. T. and S. Korenman (1992). The socioeconomic consequences of teen childbearing reconsidered. *Quarterly Journal of Economics 107(4)*, 1187–1214.

Herr, J. L. (2012, 11). Measuring the effect of the timing of first birth. University of California, Berkeley, mimeo.

Hotz, V. J., S. W. McElroy, and S. G. Sanders (2005). Teen childbearing and its life cycle consequences. *Journal of Human Resources 45(3)*, 683–715.

Jain, N. and G. R. Warnes (2006). Balloon plot. *R News 6*(2), 35–38.

Korenman, S. and D. Neumark (1992). Marriage, motherhood and wages. *Journal of Human Resources 27/2*, 233–255.

Lalive, R., A. Schlosser, A. Steinhauer, and J. Zweimüller (2013). Parental leave and mothers' careers: The relative importance of job protection and cash benefits. *The Review of Economic Studies*, rdt028.

Miller, A. (2011). The effects of motherhood timing on career path. *Journal of Population Economics 24*, pp. 1071–1100. 10.1007/s00148-009-0296-x.

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.

Schönberg, U. and J. Ludsteck (2014). Expansions in maternity leave coverage and mothers labor market outcomes after childbirth. *Journal of Labor Economics 32*(3).

Simonsen, M. and L. Skipper (2006). The costs of motherhood: an analysis using matching estimators. *Journal of Applied Econometrics 21*, pp. 919–934.

Taniguchi, H. (1999). The timing of childbearing and women's wages. *Journal of Marriage and the Family 61*, 1008–1019.

Waldfogel, J. (1998). Understanding the "family gap" in pay for women with children. *The Journal of Economic Perspectives 12(1)*, 137–156.

Wilde, E. T., L. Batchelder, and D. T. Ellwood (2010, December). The mommy track divides: The impact of childbearing on wages of women of differing skill levels. Working Paper 16582, National Bureau of Economic Research.

Zweimueller, 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, University of Linz, Austria, mimeo.

Tables

	K	0	1	2	3
			"low-wage"		
	K	0	1	2	3
K	0.561(.0037)	0.0867(.0037)	0.254(.0052)	0.091(.0050)	0.008(.0005)
0	0.079(.0077)	0.197(.0189)	0.515(.0179)	0.201(.0154)	0.008(.0012)
1	0.082(.0030)	0.022(.0017)	0.788(.0090)	0.106(.0057)	0.003(.0003)
2	0.064(.0029)	0.017(.0014)	0.059(.0037)	0.838(.0081)	0.021(.0017)
3	0.211(.0109)	0.037(.0045)	0.077(.0057)	0.359(.0134)	0.316(.0202)
			"late return"		
	K	0	1	2	3
K	0.613(.0042)	0.299(.0058)	0.077(.0062)	0.011(.0025)	0.000(.0002)
0	0.040(.0034)	0.758(.0122)	0.185(.0098)	0.016(.0012)	0.000(.0001)
1	0.035(.0016)	0.040(.0020)	0.872(.0035)	0.051(.0016)	0.002(.0001)
2	0.010(.0012)	0.010(.0010)	0.042(.0039)	0.886(.0078)	0.051(.0036)
3	0.002(.0008)	0.005(.0011)	0.003(.0008)	0.046(.0053)	0.943(.0060)
			"out-of-labor-force"		
	K	0	1	2	3
K	0.623(.0058)	0.314(.0056)	0.039(.0028)	0.016(.0012)	0.008(.0007)
0	0.029(.0030)	0.946(.0035)	0.019(.0017)	0.005(.0004)	0.001(.0002)
1	0.344(.0181)	0.258(.0185)	0.348(.0289)	0.043(.0044)	0.006(.0011)
2	0.244(.0111)	0.199(.0103)	0.075(.0059)	0.423(.0198)	0.059(.0059)
3	0.262(.0154)	0.226(.0108)	0.026(.0038)	0.074(.0077)	0.411(.0237)
			"high-wage"		
	K	0	1	2	3
K	0.511(.0031)	0.105(.0024)	0.079(.0021)	0.168(.0034)	0.139(.0024)
0	0.028(.0029)	0.508(.0170)	0.090(.0045)	0.214(.0092)	0.156(.0062)
1	0.084(.0034)	0.027(.0018)	0.463(.0109)	0.357(.0083)	0.068(.0026)
2	0.049(.0015)	0.019(.0008)	0.028(.0011)	0.717(.0069)	0.188(.0054)
3	0.037(.0005)	0.016(.0004)	0.004(.0002)	0.038(.0006)	0.905(.0009)
			"mobile"		
	K	0	1	2	3
K	0.595(.0036)	0.259(.0050)	0.100(.0047)	0.042(.0020)	0.003(.0003)
0	0.079(.0027)	0.547(.0103)	0.264(.0075)	0.099(.0036)	0.011(.0005)
1	0.072(.0024)	0.215(.0064)	0.596(.0097)	0.112(.0035)	0.007(.0003)
2	0.061(.0021)	0.156(.0044)	0.111(.0032)	0.614(.0084)	0.059(.0021)
3	0.040(.0029)	0.118(.0053)	0.031(.0024)	0.174(.0077)	0.636(.0133)

Table 1: Posterior expectation $\mathrm{E}(\boldsymbol{\xi}_h|\mathbf{y})$ and, in parenthesis, posterior standard deviations $\mathrm{SD}\,(\boldsymbol{\xi}_h|\mathbf{y})$ of the average transition matrix $\boldsymbol{\xi}_h$ in the various clusters. K=parental leave, 0=out of labor force, 1=low wage employment, 2=middle wage, 3=high wage.

	Markov chain clustering		
	1st Qu.	Median	3rd Qu.
"low-wage"	0.5930	0.7216	0.8482
"late return"	0.5987	0.7800	0.9236
"out-of-labor-force"	0.6620	0.7966	0.9063
"high-wage"	0.7865	0.9678	0.9984
"mobile"	0.5568	0.7352	0.9058
overall	0.6156	0.7885	0.9319

Table 2: Segmentation power of Markov chain clustering; reported are the lower quartile, the median and the upper quartile of the individual posterior classification probabilities $\hat{t}_{i,\hat{S}_i}$ for all individuals within a certain cluster as well as for all individuals.

Mother's educational achievement	
College	5.56%
High school	9.57%
Vocational school	28.04%
Middle school	10.47%
Compulsory school	22.48%
Education unknown	23.89%
Mother's age (in years)	
16-20	10.61%
21-25	41.36%
26-30	36.48%
31-35	11.56%
Mother's professional experience (in years)	
0-1	5.05%
2-4	34.74%
5-10	42.05%
>10	18.16%
Single mothers	19.47%
White-collar workers	68.84%
Monthly wage of last job (in 1000)	1.067
Avg 5 year wage (in 1000)	1.147
Tenure last job (in years)	3.53

Table 3: Descriptive statistics for the control variables in the multinomial logit model to explain group membership.

	"low-wage"	"late return"	"out-of-labor-force"	"mobile"
Intercept	2.683 (0.058)	4.405 (0.065)	3.097 (0.079)	4.118 (0.073)
Compulsory school (basis)				
College	−2.235 (0.066)	−2.528 (0.080)	−1.410 (0.058)	−2.345 (0.067)
High school	−0.607 (0.043)	−0.623 (0.047)	−0.805 (0.046)	−1.139 (0.049)
Vocational school	0.282 (0.046)	0.646 (0.042)	−0.003 (0.045)	−0.225 (0.044)
Middle school	−0.133 (0.044)	−0.234 (0.045)	−0.514 (0.048)	−0.711 (0.050)
Education unknown	−0.468 (0.045)	−1.190 (0.054)	−0.441 (0.041)	−6.866 (0.240)
Wage of last job in 1000	−0.413 (0.027)	−0.705 (0.026)	−0.527 (0.027)	−0.568 (0.029)
Avg 5 year wage in 1000	−1.807 (0.043)	−2.239 (0.050)	−1.533 (0.047)	−1.488 (0.049)
Tenure last job	0.052 (0.004)	0.096 (0.004)	0.038 (0.004)	−0.040 (0.006)
Mother's prof exp 0-1y	−0.612 (0.065)	−0.462 (0.067)	0.107 (0.061)	0.077 (0.060)
Mother's exp 2-4y (basis)				
Mother's prof exp 5-10y	0.358 (0.029)	0.531 (0.030)	0.035 (0.031)	−0.228 (0.033)
Mother's prof exp >10y	0.543 (0.043)	0.879 (0.045)	0.080 (0.049)	−0.312 (0.058)
Single	−0.492 (0.026)	−0.961 (0.030)	−0.915 (0.031)	−0.363 (0.029)
White-collar	−1.033 (0.030)	−1.288 (0.034)	−1.655 (0.033)	−1.292 (0.035)
Child born in 1990 (basis)				
Child born in 1991	0.197 (0.048)	0.094 (0.039)	0.092 (0.041)	0.091 (0.047)
Child born in 1992	0.450 (0.048)	0.236 (0.041)	0.299 (0.044)	0.274 (0.047)
Child born in 1993	0.750 (0.050)	0.404 (0.044)	0.486 (0.045)	0.357 (0.052)
Child born in 1994	0.986 (0.051)	0.446 (0.046)	0.626 (0.050)	0.486 (0.056)
Child born in 1995	1.248 (0.054)	0.535 (0.053)	0.809 (0.053)	0.722 (0.061)
Child born in 1996	1.531 (0.059)	0.595 (0.060)	0.912 (0.057)	0.833 (0.066)
Child born in 1997	2.116 (0.065)	1.153 (0.068)	1.392 (0.067)	1.385 (0.074)
Child born in 1998	2.487 (0.069)	1.370 (0.079)	1.601 (0.077)	1.724 (0.078)
Child born in 1999	2.586 (0.077)	1.428 (0.087)	1.517 (0.085)	1.934 (0.089)
Child born in 2000	2.966 (0.081)	1.371 (0.108)	1.754 (0.093)	2.136 (0.096)
Mother born in 1954-58	−0.679 (0.155)	−0.257 (0.103)	0.451 (0.097)	0.039 (0.139)
Mother born in 1959-63	−0.549 (0.058)	−0.190 (0.056)	0.276 (0.055)	−0.064 (0.069)
Mother born in 1964-68	−0.229 (0.033)	−0.062 (0.033)	0.080 (0.034)	−0.130 (0.039)
Mother b. 1969-1973 (basis)				
Mother born in 1974-78	0.016 (0.044)	−0.313 (0.054)	−0.096 (0.051)	0.164 (0.048)
Mother born in 1979-84	0.103 (0.171)	−1.315 (0.282)	0.172 (0.198)	0.513 (0.178)
Mother's age 16-20 y	0.209 (0.050)	0.099 (0.054)	0.089 (0.056)	0.287 (0.052)
Mother's age 21-25 y (basis)				
Mother's age 26-30 y	−0.156 (0.029)	−0.045 (0.031)	0.264 (0.032)	−0.023 (0.039)
Mother's age 31-35 y	−0.376 (0.058)	−0.167 (0.058)	0.463 (0.056)	−0.036 (0.071)

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

Figures

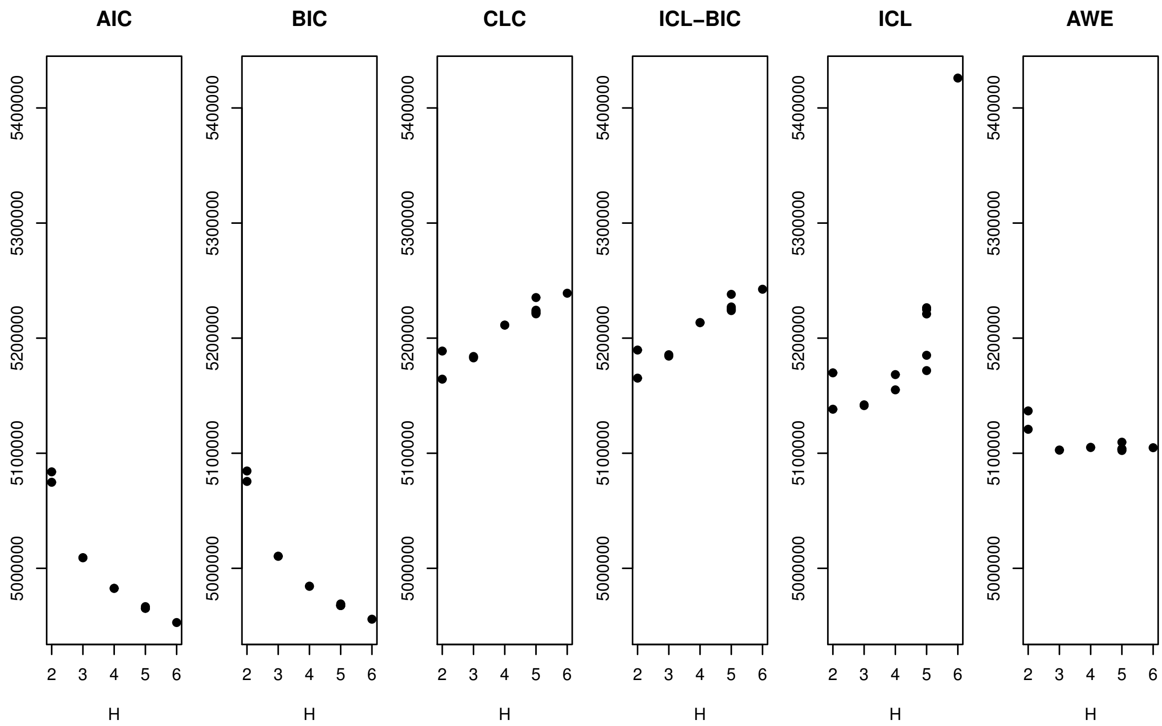

Figure 1: Model selection criteria for various numbers H of clusters and several independent MCMC runs.

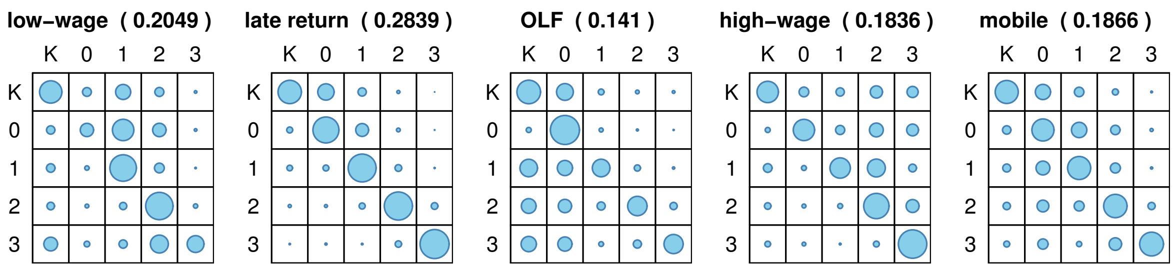

Figure 2: Visualization of posterior expectation of the transition matrices $\boldsymbol{\xi}_1$, $\boldsymbol{\xi}_2$, $\boldsymbol{\xi}_3$, $\boldsymbol{\xi}_4$, and $\boldsymbol{\xi}_5$ obtained by Markov chain clustering. The circular areas are proportional to the size of the corresponding entry in the transition matrix. The corresponding group sizes are calculated based on the posterior classification probabilities and are indicated in the parenthesis. K=parental leave, 0=out of labor force, 1=low wage employment, 2=middle wage, 3=high wage.

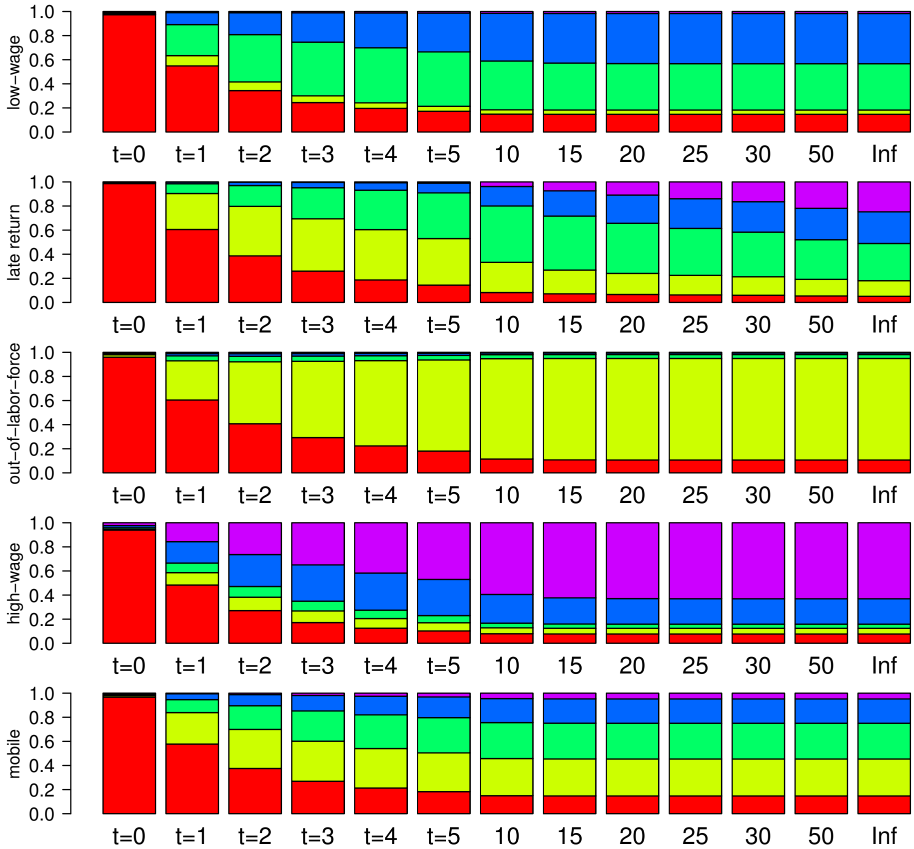

Figure 3: Posterior expectation of the distribution $\boldsymbol{\pi}_{h,t}$ over the 5 states (parental leave, out of labor force, low wage employment, middle wage, high wage) after a period of t years in the various clusters. `Inf` corresponds to the steady state in each cluster.

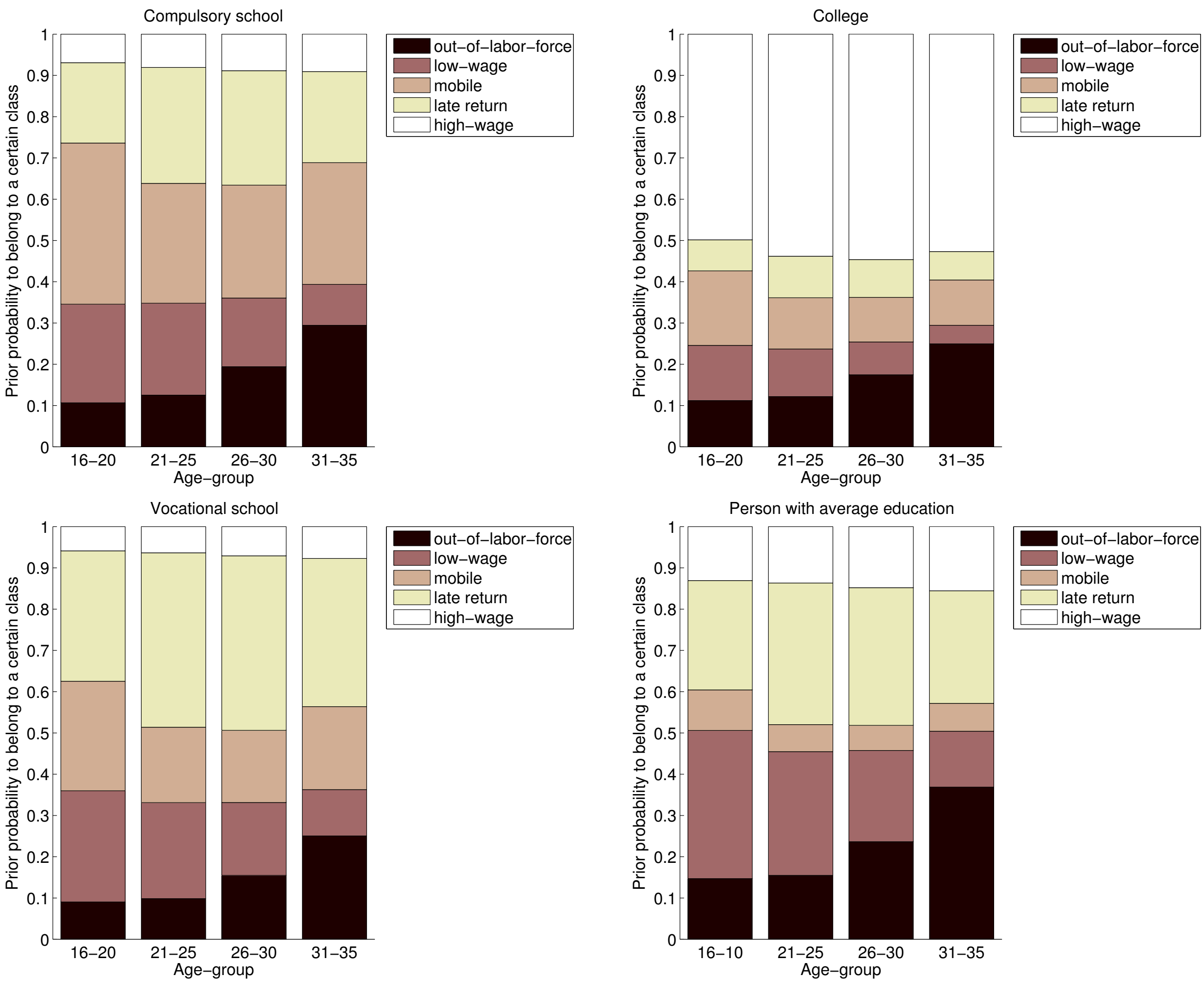

Figure 4: Influence of a mother's age (for specific educational groups and averaged over education)

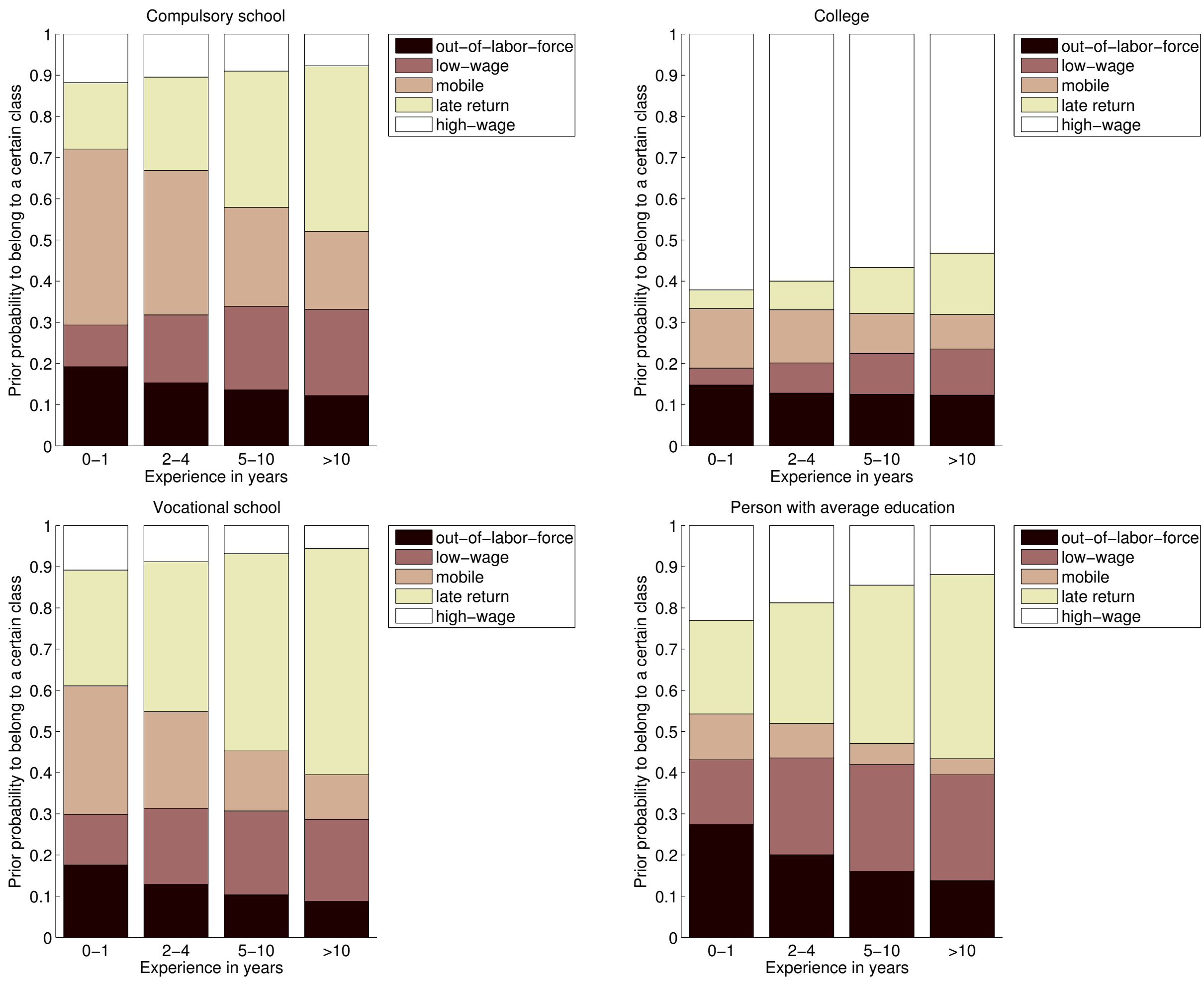

Figure 5: Influence of a mother's experience (for specific educational groups and averaged over education)