

Flexible parametric multi-state modelling of employment history

Ardo van den Hout¹, and Wenhui Tan¹,

¹ Department of Statistical Science, University College London, United Kingdom

Address for correspondence: Ardo van den Hout, Department of Statistical Science, University College London, Gower Street, London WC1E 6BT, United Kingdom.

E-mail: ardo.vandenhout@ucl.ac.uk.

Phone: (+44 20) 3108 3243.

Fax: (+44 20) 3108 3105.

Abstract: A multi-state model is used to describe employment history. Transition-specific rates are defined using generalised gamma distributions and Gompertz distributions. This flexible parametric modelling of the rate of change is combined with latent classes for unobserved propensity to change jobs. The propensity is described by two latent classes which can be interpreted as consisting of *movers* and *stayers*. The modelling is illustrated by analysing longitudinal data from the German Life History Study.

Key words: demography; frailty model; generalised gamma distribution; Gompertz

distribution; Markov process; survival analysis

1 Introduction

Multi-state models are used to describe stochastic processes where the change of status is of interest. Many applications can be found in biostatistics with the illness-death model as the quintessential example. An illness-death model can be defined by three states: a healthy state, an ill state, and the dead state. Of interest can be risk factors for the onset of the illness, or expected duration in the ill state. In social statistics and in demography, multi-state models are used to study processes such as changes in region of residence, employment history, or changes in marital status.

Statistical methods for multi-state models are typically not discipline specific. As an example, a three-state illness-death model for medical data can be quite similar to a three-state model for employment history when the states in the latter are defined as employed, unemployed, and retired.

The aim of this paper is to introduce the generalised gamma distribution ([Stacy, 1962](#)) for flexible parametric continuous-time multi-state modelling in demography. The application is with respect to employment history, where states are defined corresponding to the number of past jobs, and intermediate spells without a job. The time scale in our model for employment history is *time since entry into the labour market*. We use the generalised gamma distribution to model transition-specific time dependence and we show that using this distribution improves the statistical inference when combined with transition-specific exponential and Gompertz distributions. In addition, we show that parametric models can be extended with a definition of latent

classes.

For continuous-time multi-state models, the generalised gamma distribution is discussed in [Jackson \(2016\)](#). The parametric modelling is within a wider methodological framework for multi-state models described by, for example, [Kalbfleisch and Lawless \(1985\)](#), [Kay \(1986\)](#), [Hougaard \(2000\)](#), [Jackson \(2011\)](#), and [Van den Hout \(2017\)](#). The latent-class model that we define, can be seen as a random-effects model with a discrete distribution for the random effects. Latent-class models for discrete-time multi-state models are discussed in, for example, [Vermunt et al. \(1999\)](#) and [Bartolucci et al. \(2012\)](#). An example in demography is [Dias and Willekens \(2005\)](#), who discuss determining the number of latent classes. For continuous-time multi-state models which include random effects see, for example, [Hougaard \(2000\)](#), [Putter and Van Houwelingen \(2015\)](#), and [Van den Hout \(2017\)](#). For multi-state modelling in demography, an overview with a wide range of random-effects structures is given by [Bijwaard \(2014\)](#).

When multi-state models are applied in biostatistics, we can see the latent-class model as a frailty model. For example, if progression through a set of states denotes a deterioration of health, a model with two latent classes can distinguish individuals who move quickly through the states (the frail individuals) from those who move less quickly (the more healthy ones). In the current paper, the latent classes are defined with respect to employment history and will allow us to distinguish individuals who tend to change job more quickly (the *movers*) from those who tend to stay put (the *stayers*). In the application, we will show that such a distinction can lead to a model that fits the data better.

Our multi-state model for employment history combines the generalised gamma distri-

bution with the latent-class approach. Although this distribution and the latent-class approach have been discussed in the literature separately (see the references above), our contribution is to combine these two concepts and thus defining a very flexible statistical modelling framework. The paper will illustrate this by defining a series of models, and by model comparison, validation, and interpretation. We show that the modelling framework is general and allows for a wide range of applications in demography.

2 The German Life History Survey

The German Life History Study (1980-2005) provides retrospective life course information for Germans born between 1919 and 1951 ([Mayer, 2015](#)). The study is often used to investigate education, employment history, and family formation.

For the current paper, we use the Blossfeld-Rohwer subsample of the German Life History Study ([Blossfeld and Rohwer, 2002](#)). These data are available in the `Biograph` package in the R software. This package is introduced by [Willekens \(2014\)](#), who uses the Blossfeld-Rohwer subsample in a three-state model for employment history.

This subsample contains data for 201 individuals on job episodes and spells without a job. The start and end of each job episode are available on a time scale in months. Individual background information is given by covariates such as age, gender, and education.

Interviews for GLHS were conducted in 1981. In what follows, we propose a statistical model for employment history as known at the time of the interview. This history

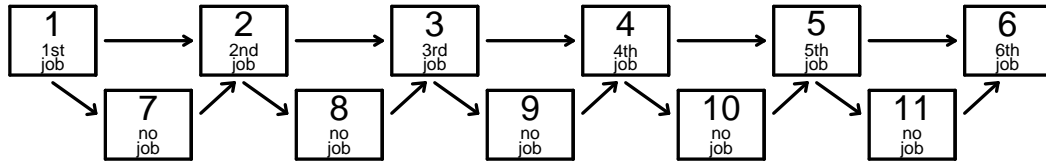


Figure 1: Eleven-state process for employment history in GLHS.

is defined by a series of mutually exclusive states. Because not many individuals in GLHS have had more than six jobs, we restrict the modelling of employment history up to (and including) the sixth job.

Figure 1 shows the diagram for the process that we will model in this paper. States 1 up to 6 denote the first job up to the sixth job. State 7 up to 11 denote episodes between jobs; that is, state 7 denotes having no job after the first job, state 8 denotes having no job after the second job, etcetera. With the imposed restriction to a maximum of six jobs, state 6 acts as an absorbing state.

It is possible to define a process with fewer states by collapsing the five no-job states into one state. However, the model comparison in Section 6 shows that it is worthwhile to distinguish the five no-job states as illustrated in Figure 1.

In the subsample (denoted *GLHS* from now on), all individuals start in state 1, but they do so at different times and different ages. The time scale in GLHS is century



Figure 2: For the 201 individuals in GLHS, a two-dimensional representation of the start time of the first job and the age at the start of the first job.

months. For example, the first individual in the data is a man who starts his first job at century month 555. In calendar years, this is $(555 - 1)/12 + 1900 = 1946$. In GLHS, three birth cohorts are represented: 1929-31, 1939-41, and 1949-51.

For the 201 individuals in GLHS, Figure 2 shows the start time of the first job (in calendar years) and the age at which the individuals started their first job. The graph shows a cohort effect in the sense that the age at the start of the first job is increasing with increasing calendar year for the start of the first job. Therefore, we will include cohort effects in the statistical model for employment history.

Table 1 is the state table for the eleven-state employment history in GLHS. This

Table 1: State table for GLHS: number of times for each successive pair of states.

The states are defined by employment history; see Figure 1.

<i>From</i>	<i>To</i>										
	1	2	3	4	5	6	7	8	9	10	11
1	16	96	0	0	0	0	89	0	0	0	0
2	0	44	78	0	0	0	0	40	0	0	0
3	0	0	39	52	0	0	0	0	16	0	0
4	0	0	0	27	24	0	0	0	0	11	0
5	0	0	0	0	11	18	0	0	0	0	3
7	0	66	0	0	0	0	23	0	0	0	0
8	0	0	29	0	0	0	0	11	0	0	0
9	0	0	0	10	0	0	0	0	6	0	0
10	0	0	0	0	8	0	0	0	0	3	0
11	0	0	0	0	0	2	0	0	0	0	1

table shows the frequencies for successive pair of states. Right-censored histories are presented in the (r, r) entries with $r = 1, \dots, 5, 7, \dots, 11$. To give an example, of the 201 individuals who started in state 1, 16 are still in their first jobs at the time of the interview, 96 went straight from their first jobs to their second job, and 89 had a period without a job after their first job. The table shows that data are sparse for state 5, 6, 10 and 11. In total there are 564 job episodes (sum of the frequencies in the first five rows) and 159 spells without a job (sum of the frequencies in the last five rows).

To model potential time dependence in employment history, we need to specify a time scale for the process. Possible options are *time in current job*, *age*, and *time since entry into the labour market*. For our statistical model, we use the third option following the choice of Blossfeld and Rohwer (2002, Chapter 2) for their model for the rate of leaving the current job. This option implies, for example, that our modelling allows two individuals who have been in their third job equally long, to have a different distribution for moving to the fourth job due to a difference in time since entry into the labour market. It allows us to investigate the effects of gender and education while controlling for time spent in the labour market.

3 Distributions for transition times

To model change of state in continuous time, we specify continuous parametric distributions for transition time $T \geq 0$. This section starts with the definition of transition-specific hazard functions and state-specific survivor functions, and will then discuss the exponential distribution, the Gompertz distribution, and the generalised gamma distribution.

Our modelling is based on continuous-time stochastic processes, which are defined as $\{Y_t | t \in (0, \infty)\}$ with Y_t denoting the state at time t . The range of Y_t is defined by the state space \mathcal{S} .

In our application, we have finite $\mathcal{S} = \{1, 2, \dots, 11\}$, with state 6 being the only absorbing state. Let $(t_1, t_2]$ denote a generic time interval, and define $Y_j = Y_{t_j}$, for $j = 1, 2$. Specific for GLHS, the following type of observations are possible for a time interval $(t_1, t_2]$. If there is no transition, then $(y_1, y_2) = (r, r)$, for $r \in \mathcal{S} \setminus \{6\}$. If there

is a transition, then $(y_1, y_2) = (r, s)$ with $s = r + 1$ or $s = r + 6$ if $r \in \{1, 2, 3, 4, 5\}$, and $s = r - 5$ for $r \in \{7, 8, 9, 10, 11\}$. For example, if the individual moves out of state 1, then he or she moves to state 2 or state 7. Given state 7, the next possible state is 2.

Let $h_{rs}(t)$ denote the hazard for moving from state r to state s at time t ; that is,

$$h_{rs}(t) = \lim_{\Delta \downarrow 0} \frac{P(Y_{t+\Delta} = s | Y_t = r)}{\Delta}.$$

If a transition is not possible from r to s , we define $h_{rs}(t) = 0$. Let $h_{r\bullet}(t)$ denote the hazard for moving out of state r . For example, for state 1 we have $h_{1\bullet}(t) = h_{12}(t) + h_{17}(t)$.

For $(t_1, t_2]$, the conditional survivor function for state r at t_1 is given by

$$S_r(t_2|t_1) = \exp\left(-\int_{t_1}^{t_2} h_{r\bullet}(u) du\right).$$

This definition of $S_r(t_2|t_1)$ implies that given state r at t_1 , the hazard function $h_{r\bullet}$ defines the probability of a transition out of state r . In the application, if all five hazard functions are defined by exponential distributions, then the corresponding transition probabilities define a time-homogeneous first-order Markov chain. If the hazard functions are time-dependent or contain covariate effects, then the process is not a Markov chain in the strict sense; see also [Commenges \(1999\)](#).

3.1 Exponential and Gompertz distribution

The transition-specific hazard function for the exponential distribution is given $h_{rs}(t) = \exp(\mu_{rs})$, where $\mu_{rs} \in \mathbb{R}$. The exponential model is a very common model for multi-state processes and is readily available in software for multi-state models. Note that

this hazard function does not depend on time t , which may be too restrictive in some applications.

The transition-specific hazard function for the Gompertz distribution is $h_r(t) = \exp(\mu_{rs} + \xi_{rs}t)$, where $\mu_{rs}, \xi_{rs} \in \mathbb{R}$. As a distribution for a time to event, the Gompertz distribution is used in biostatistics and sociology; see, for example, [Hougaard \(2000\)](#) and [Blossfeld and Rohwer \(2002\)](#), respectively. For usage in demography; see, for example, [Mueller et al. \(1995\)](#), [Olshansky and Carnes \(1997\)](#), and [Bongaarts \(2005\)](#).

Models with exponential and Gompertz distributions can be extended in the usual way by log-linear regression: $\mu_{rs} = \boldsymbol{\beta}_{rs}^\top \mathbf{x}$, where \mathbf{x} is a vector of covariates values including an intercept.

3.2 Generalised gamma distribution

The generalised gamma distribution is a very flexible continuous parametric distribution for event time $T \geq 0$. The distribution includes the exponential, Weibull, log-normal, and gamma distributions as special cases; that is, these cases can be defined by restricting certain parameters in the generalised gamma distribution. We will first review the generalised gamma distribution, and then use this distribution to define the transition-specific hazards.

Following the presentation and terminology in [Cox et al. \(2007\)](#), the generalised gamma distribution $T \sim GG(\mu, \sigma, q)$ has probability density function

$$f(t|\mu, \sigma, q) = \frac{|q|}{\sigma t \Gamma(q^{-2})} (q^{-2}(\exp(-\mu)t)^{q/\sigma})^{q^{-2}} \exp(-q^{-2}(\exp(-\mu)t)^{q/\sigma}), \quad (3.1)$$

for location μ , scale $\sigma > 0$, and shape q . The survivor function is given by

$$S(t) = \begin{cases} 1 - \Gamma [q^{-2}(\exp(-\mu)t)^{q/\sigma}; q^{-2}] & \text{if } q > 0 \\ \Gamma [q^{-2}(\exp(-\mu)t)^{q/\sigma}; q^{-2}] & \text{if } q < 0 \end{cases}, \quad (3.2)$$

where $\Gamma [t; \gamma] = \int_0^t x^{\gamma-1} e^{-x} dx / \Gamma(\gamma)$. [Prentice \(1974\)](#) introduced the above parameterisation for the generalised gamma by [Stacy \(1962\)](#). This choice of parameterisation allows $q \leq 0$. The generalised gamma distribution is available in R via the `flexsurv` package ([Jackson, 2016](#)).

The log-normal distribution is defined as $GG(\mu, \sigma, q = 0)$; the Weibull distribution as $GG(\mu, \sigma, q = 1)$; and the gamma distribution as $GG(\mu, \sigma, q = \sigma)$. To illustrate this for the Weibull distribution, for $q = 1$ density [\(3.1\)](#) reduces to

$$f(t|\mu, \sigma, q = 1) = \frac{1}{\sigma} t^{1/\sigma-1} (\exp(-\mu))^{1/\sigma} \exp(-(\exp(-\mu)t)^{1/\sigma}).$$

Re-parameterisation by $\tau = 1/\sigma$ and $\lambda = \exp(\mu)$ gives

$$f(t|\tau, \lambda) = \frac{\tau}{\lambda} \left(\frac{t}{\lambda}\right)^{\tau-1} \exp\left(-\left(\frac{t}{\lambda}\right)^\tau\right),$$

which is a standard representation of the Weibull density. The median of this distribution is $\lambda (\log(2))^{1/\tau} = \exp(\mu) (\log(2))^\sigma$. This illustrates that the location parameter μ determines the median given a fixed value for scale σ . This also holds for the unrestricted generalised gamma distribution. Scale σ and shape q together determine the type of hazard function—as stated and illustrated in [Cox et al. \(2007\)](#); see also the data analysis in [Section 6](#).

For the multi-state model, we define transition-specific hazard as $h_{rs}(t) = f_{rs}(t)/S_{rs}(t)$, where the density function $f_{rs}(t)$ and the survivor function $S_{rs}(t)$ are defined by the generalised gamma distribution $GG(\mu_{rs}, \sigma_{rs}, q_{rs})$. If there are two possible transitions

out of state r , then we are in a competing-risks setting. For example, the function $S_{12}(t)$ as defined above should not be interpreted as the survivor function with respect to being in state 1; see also (Putter et al., 2007, Section 3.2).

4 Latent-class models

A latent-class model is defined by specifying K classes, probabilities $\pi_{ik} = P(i \in \text{class } k)$ for individual $i \in \{1, \dots, N\}$ and class $k \in \{1, 2, \dots, K\}$, and class-specific model parameters. The probabilities π_{ik} model the class membership with the restriction $\sum_{k=1}^K \pi_{ik} = 1$. Class-specific model parameters can refer to location, shape or scale parameters that vary across the K classes.

The class-specific parameters induce class-specific distributions for the transition times. As an example, consider the exponential model and $K = 2$. Say we want to explore the latent-class modelling for direct transitions from one job to the next job. We can define $\pi_{i1} = \pi_1$ and $\pi_{i2} = 1 - \pi_1$, for all i , and make the relevant location parameters class-specific by defining $\mu_{rs,k} = \beta_{rs} + b_k$, where $(r, s) \in \{(1, 2), (2, 3), (3, 4), (4, 5), (5, 6)\}$ and $b_1 = -b_2$. The two latent classes induce two extra parameters to be estimated: π_1 and b_1 . The first is the probability to belong to class 1, and the second is the effect on the location parameter due to being in class 1. For example, if $b_1 < 0$, then $\exp(\mu_{rs,1}) < \exp(\mu_{rs,2})$, which implies that individuals in class 2 are more likely to move to state s than individuals in class 1. A model like this allows us to distinguish so-called *movers* from *stayers*: unobserved heterogeneity in the propensity to move to a next state.

The distinction between individuals who move quickly through the states and individ-

uals who tend not to move so quickly is described as *tracking* by [Satten \(1999\)](#): there is a correlation between the transition times within individuals, and rapid progressors can be distinguished from slow progressors. In the literature, a mover/stayer model can also denote a stochastic process that is a mixture of two processes, one of which has transition hazards equal to zero; see, for example, [Frydman \(1984\)](#) and [Cole et al. \(2005\)](#). This option, however, will not be explored in the current paper.

Latent-class models can be seen as random-effects models where the random effects have a discrete distribution; see, for example, the generalised linear mixed models as discussed in [Aitkin \(1999\)](#). When fitting the latent-class as a random-effects model, the classes that are created are primarily aimed at capturing unobserved heterogeneity—interpretation of the classes is not of primary interest.

5 Estimation

For individual i , for $i \in \{1, 2, \dots, N\}$, GLHS data were formatted such that the data are given by times $\mathbf{t}_i = (t_{i1}, t_{i2}, \dots, t_{in_i})$, states $\mathbf{y}_i = (y_{i1}, y_{i2}, \dots, y_{in_i})$, and baseline covariates values $\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{im})$. We also define transition indicators $\boldsymbol{\delta}_i = (\delta_{i2}, \delta_{i3}, \dots, \delta_{in_i})$ to maintain the standard notation in survival analysis. Note, however, that in the GLHS data, times $t_{i1}, \dots, t_{in_i-1}$ are all transition times, and t_{in_i} is either a right-censored time, or a transition time for moving to state 6.

Consider a time interval $(t_{ij}, t_{ij+1}]$, $j = 1, \dots, n_i - 1$. When $y_{ij} = y_{ij+1}$, there is no transition and the conditional survivor function for observing state Y_{ij+1} at t_{ij+1} is $S_{y_{ij}}(t_{ij+1}|t_{ij})$. When $y_{ij} \neq y_{ij+1}$, time t_{ij+1} denotes a time of a transition, and the density function for moving to observed state Y_{ij+1} at t_{ij+1} is $S_{y_{ij}}(t_{ij+1}|t_{ij})h_{y_{ij}y_{ij+1}}(t_{ij+1})$.

The former case is labelled by $\delta_{ij+1} = 0$ and the latter by $\delta_{ij+1} = 1$.

Individual contributions to the likelihood function for the fixed-effects models are given by

$$L_i(\boldsymbol{\theta}|\mathbf{t}_i, \boldsymbol{\delta}_i, \mathbf{y}_i, \mathbf{x}_i) = \prod_{j=1}^{n_i-1} S_{y_{ij}}(t_{ij+1}|t_{ij})h_{y_{ij}y_{ij+1}}(t_{ij+1})^{(\delta_{ij+1})},$$

for $i = 1, 2, \dots, N$, with $\boldsymbol{\theta}$ the vector containing all the model parameters. The overall likelihood function is $L(\boldsymbol{\theta}) = \prod_{i=1}^N L_i(\boldsymbol{\theta}|\mathbf{t}_i, \boldsymbol{\delta}_i, \mathbf{y}_i, \mathbf{x}_i)$.

Individual contributions to the likelihood function for the latent-class models are given by

$$L_i(\boldsymbol{\theta}|\mathbf{t}_i, \boldsymbol{\delta}_i, \mathbf{y}_i, \mathbf{x}_i) = \sum_{k=1}^K \left(\prod_{j=1}^{n_i-1} S_{y_{ij}}(t_{ij+1}|t_{ij}, k)h_{y_{ij}y_{ij+1}}(t_{ij+1}|t_{ij}, k)^{(\delta_{ij+1})} \right) \pi_{ik},$$

where $\boldsymbol{\theta}$ includes the parameters for the latent-class modelling. Probabilities π_{ik} can be linked to individual covariate values using a multinomial regression model. However, in the GLHS data analysis in Section 6 we assume $\pi_{ik} = \pi_k \forall i$.

The likelihood function for the fixed-effects models is the standard format for survival data extended to multiple event times within individuals. Note that possible association between observations within an individual is not modelled. In contrast, the likelihood function for the latent-class models allows a correlation between observations via the latent-class parameters.

For the maximum likelihood estimation we use the general-purpose optimiser in the R software `optim` (R Core Team, 2013). This versatile optimiser can maximise the log-likelihood function without the need to provide explicit expressions for the derivatives. The optimiser will return the numerically differentiated Hessian matrix if requested.

Covariance of a function of estimated model parameters can be derived by using

simulation. An important example of such a function are transition probabilities for a specified time interval. For the simulation, parameter vectors $\boldsymbol{\theta}^{(s)}$, for $s = 1, \dots, S$, are drawn from $N(\widehat{\boldsymbol{\theta}}, \widehat{\mathbf{V}}_{\boldsymbol{\theta}})$, where $\widehat{\mathbf{V}}_{\boldsymbol{\theta}}$ is the estimated covariance matrix derived from the estimated Hessian. Next, for any function $g(\boldsymbol{\theta})$, summary statistics such as mean and covariance are derived from $g(\boldsymbol{\theta}^{(1)}), \dots, g(\boldsymbol{\theta}^{(S)})$; see [Mandel \(2013\)](#) for a justification of this method.

6 Data analysis

We start the data analysis for GLHS with the basic exponential model. This model can be seen as an intercepts-only model for the location parameters, and has as many parameters as transitions in the process: $\mu_{rs} = \beta_{rs,0}$, for the pairs (r, s) as defined by the process in [Figure 1](#). This model has Akaike information criterion (AIC) equal to 6605.4.

Dummy variables for the cohort effects are defined as $(x_1, x_2) = (0, 0)$ for birth cohort 1929-31, $(x_1, x_2) = (1, 0)$ for cohort 1939-41, and $(x_1, x_2) = (0, 1)$ for cohort 1949-51; see also [Blossfeld and Rohwer \(2002\)](#). Let x_3 and x_4 denote the dummy variables for the effect of gender and education. Women are coded by $x_3 = 0$, men by $x_3 = 1$. Education is defined by years of education before entry into labour market: $x_4 = 0$ for less than 12 years, $x_4 = 1$ for 12 years or more. The threshold 12 years is equal to middle school qualification with vocational training.

For the modelling of the effects of the dummy variables, we distinguish three types of transitions: from one job to the next (A), from a job to no job (B), and from no job to the next job (C). The model for the location parameters with the covariates

is thus given by

$$\begin{aligned}
 A) \mu_{rs} &= \beta_{rs.0} + \beta_{A.1}x_1 + \beta_{A.2}x_2 + \beta_{A.3}x_3 + \beta_{A.4}x_4 & \left\{ \begin{array}{l} \text{From a job to the next:} \\ (r, s) = (r, r + 1), r \in \{1, \dots, 5\} \end{array} \right. \\
 B) \mu_{rs} &= \beta_{rs.0} + \beta_{B.1}x_1 + \beta_{B.2}x_2 + \beta_{B.3}x_3 + \beta_{B.4}x_4 & \left\{ \begin{array}{l} \text{From a job to no job:} \\ (r, s) = (r, r + 6), r \in \{1, \dots, 5\} \end{array} \right. \\
 C) \mu_{rs} &= \beta_{rs.0} + \beta_{C.1}x_1 + \beta_{C.2}x_2 + \beta_{C.3}x_3 + \beta_{C.4}x_4 & \left\{ \begin{array}{l} \text{From no job to the next job:} \\ (r, s) = (r, r - 5), r \in \{7, \dots, 11\} \end{array} \right.
 \end{aligned} \tag{6.1}$$

and has $\text{AIC} = 6437.2$, which shows a clear improvement over the intercepts-only model.

It is possible to restrict the intercepts $\beta_{rs.0}$ in such a way that the hazards for moving to a no-job state are the same: $\beta_{r,r+6.0} = \beta_{B.0}$ for $r \in \{1, \dots, 5\}$, and the hazards for leaving a no-job state are the same: $\beta_{r,r-5.0} = \beta_{C.0}$ for $r \in \{7, \dots, 11\}$. This defines 7 independent intercepts. Adding the 12 parameters for the dummy variables, the resulting model has $\text{AIC} = 6453.9$. The increase of the AIC compared to the previous model shows that it is worthwhile to distinguish the transition hazards for the five no-job states. This implies that we should not merge the five no-job states into one general no-job state.

To investigate non-constant transition hazards, we discuss two models. For the first model, we add Gompertz ξ -parameters to the first four transitions from one job to the next; that is, we add ξ_{rs} for $(r, s) = (r, r + 1)$, $r \in \{1, 2, 3, 4\}$. Because data are scarce for the transition from state 5 to state 6, we do not extend the model for this transition. The Gompertz model yields an $\text{AIC} = 6400.9$, which shows that allowing hazards to change over time can lead to a model with better fit.

Table 1 shows that there is good information for moving out of states 1 and 2. To allow for more flexible parametric shapes, we fit generalised gamma distributions for $(r, s) \in \{(1, 2), (1, 7)\}$. The shape parameters are q_{12} and q_{17} , and the scale parameters are

$$\sigma_{12} = \exp(\psi_{12}) \quad \sigma_{17} = \exp(\psi_{17}).$$

For $(r, s) \in \{(2, 3), (2, 8), (3, 4), (4, 5)\}$, we use Gompertz distributions. Note that this also implies a non-constant hazard for moving from state 2 to no-job state 8. For the remaining transitions we fit exponential distributions. The covariate effects are again as specified in (6.1). We call this Model \mathcal{A} . It has $\text{AIC} = 6379.6$, and shows a further improvement over the earlier models.

Next, we extend Model \mathcal{A} with two latent classes ($K = 2$). We define class-specific distributions via the location parameters for the first four direct job changes. To impose the mover/stayer distinction, the class effect for the four job changes is assumed to be the same. For $k \in \{1, 2\}$, this implies the following change for the location parameters:

$$\mu_{rs.k} = \beta_{rs.0} + b_k + \beta_{A.1}x_1 + \beta_{A.2}x_2 + \beta_{A.3}x_3 + \beta_{A.4}x_4, \quad (6.2)$$

for $(r, s) = (r, r + 1), r \in \{1, \dots, 4\}$. The model is identified by the sum-to-zero restriction for the class-specific effects; that is, $b_1 + b_2 = 0$. We call this Model \mathcal{B} . It has two more parameters (π_1 and b_1) compared to Model \mathcal{A} , and has $\text{AIC} = 6366.8$.

Model \mathcal{B} is a better model than \mathcal{A} given the improvement in AIC. In the remainder of this section we present statistical inference for Model \mathcal{B} .

Estimated parameters are presented in Table 2. The probability π_1 to be in class 1 is estimated at 0.276 (with estimated standard error 0.083). For this class, the additive

effect b_1 on the location parameters in (6.2) is 0.641 (0.108). Given that $b_1 > 0$, class 1 represents the movers—those individuals who are more likely to move directly to a next job when compared to individuals in class 2.

For women born in 1939-41 and with less than 12 years of education, Figure 3 depicts the hazards in Model \mathcal{B} that are time dependent. The difference between the two classes is illustrated with the black lines representing class 1, and the grey representing class 2. Confidence bands in this figure are derived by simulation; see end of Section 5. For $1 \rightarrow 7$ and $2 \rightarrow 8$, the hazards are restricted to be the same for both classes. Hazards are modelled conditional on current state. For example, if the women are in state 3 quickly after having entered the labour market, then the model implies a high hazard for moving the state 4.

The flexibility of the generalised gamma distribution is illustrated in Figure 3 by the arc-shape hazards transitions $1 \rightarrow 2$ and $1 \rightarrow 7$. For example, the hazard for moving directly to the second job increases in the first four to five years, and shows a decrease afterwards. This can be linked to the estimated scale $\hat{\sigma}_{12} = \exp(\hat{\psi}_{12})$ and shape $\hat{q}_{12} < 0$ in Table 2. This combination of values is in the set $\{(\sigma, q) | q < \min\{\sigma, 1/\sigma\}\}$ that defines arc-shape hazards, starting and ending at zero (Cox et al., 2007). For the Gompertz model, the estimated ξ -parameters for transitions $2 \rightarrow 3$, $2 \rightarrow 8$, $3 \rightarrow 4$ and $4 \rightarrow 5$ are negative and illustrate a negative time effect on the hazard.

Looking at the estimated standard errors in Table 2, we see that there are clear cohort effects. The younger cohorts have higher hazards for all transitions. The effects for gender illustrate that men (coded by $x_3 = 1$) are more likely to move directly to a new job and also more likely to move from having no job to the next job. Women have a higher hazard for moving to a no-job state. Given that the population in GLHS

Table 2: Parameter estimates for the mover/stayer model for the GLHS data on employment history (Model \mathcal{B}). Time scale is months since entry into the labour market. Estimated standard errors in parentheses. Estimated intercepts are not included.

Location				Scale and shape			
$\beta_{A.1}$	0.417	(0.201)	$\beta_{A.3}$	0.200	(0.160)	ψ_{12}	0.044 (0.197)
$\beta_{B.1}$	0.497	(0.205)	$\beta_{B.3}$	-0.631	(0.177)	ψ_{17}	0.311 (0.086)
$\beta_{C.1}$	0.539	(0.231)	$\beta_{C.3}$	2.107	(0.202)	q_{12}	-2.038 (0.760)
$\beta_{A.2}$	0.228	(0.211)	$\beta_{A.4}$	0.003	(0.188)	q_{17}	-0.439 (0.459)
$\beta_{B.2}$	0.599	(0.206)	$\beta_{B.4}$	-0.160	(0.202)		
$\beta_{C.2}$	0.429	(0.254)	$\beta_{C.4}$	-0.717	(0.262)	ξ_{23}	-0.004 (0.002)
						ξ_{28}	-0.007 (0.002)
Class 1						ξ_{34}	-0.003 (0.002)
π_1	0.276	(0.083)	b_1	0.641	(0.108)	ξ_{45}	-0.004 (0.002)

is born before 1951, this higher hazard can be explained by women dropping out of the labour market due to marriage or pregnancy. The estimated effects for years of education have high standard errors. More years of education seems to be negatively associated with moving out of a no-job state when controlling for the effects of time and gender. We do not have a good explanation for this effect. Note that when someone is in the no-job state, this does not imply that this individual is looking for a new job. For example, when people are retired, they are also classified in this state.

Goodness of fit for Model \mathcal{B} can be checked visually by comparing model-based survival with Kaplan-Meier survivor curves; see Figure 4. The solid lines are the model-

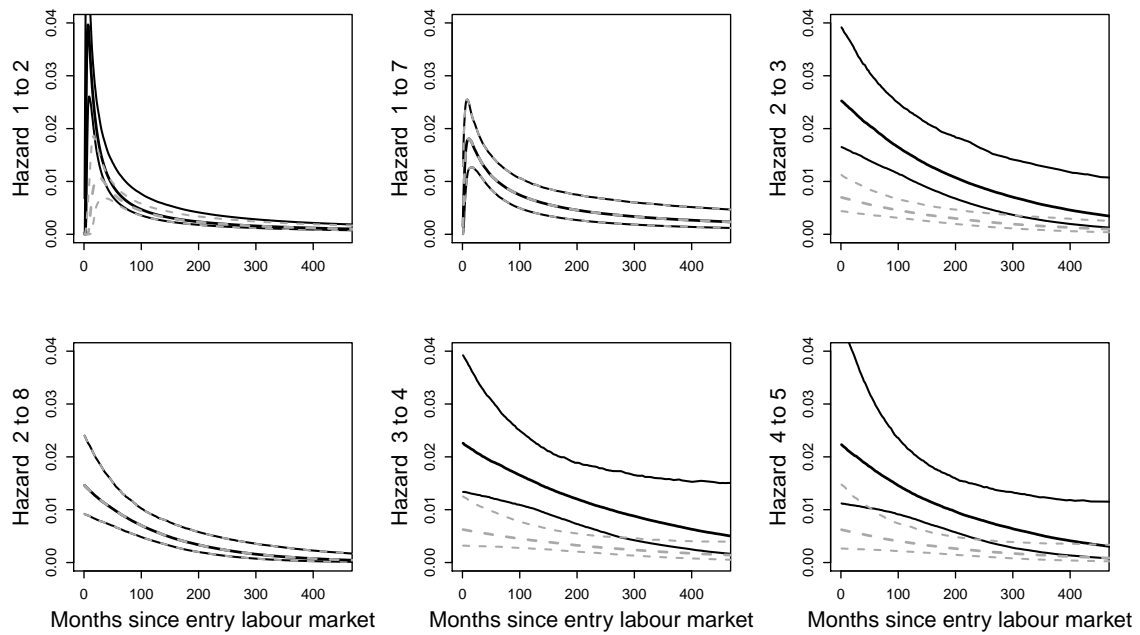


Figure 3: Transition-specific hazards for women born in 1939-41 and with less than 12 years of education. Black curves for class 1, and grey curves for class 2 (with 95% confidence bands).

based survival predicted using the baseline data only. The survival is defined for a progressive series of sets of states. The first graph is for survival prior to state 2; that is, survival in state 1. The second graph is for survival prior to state 3; that is, survival in states 1, 7, or 2. The graphs for the higher states are defined analogously. The model-based prediction is derived as the mean of the survivor curves for the 201 individuals in the data. The latent-class structure is taken into account by predicting survival for each individual twice: once assuming the individual is in class 1, once assuming the individual is in class 2. The individual survivor curve is then defined as a weighted sum of the class-specific survival where the weights are the estimated probabilities π_k for $k \in \{1, 2\}$. The dashed lines are the Kaplan-Meier survivor curves (with 95% confidence intervals), where the plus-symbols depict right-censored times.

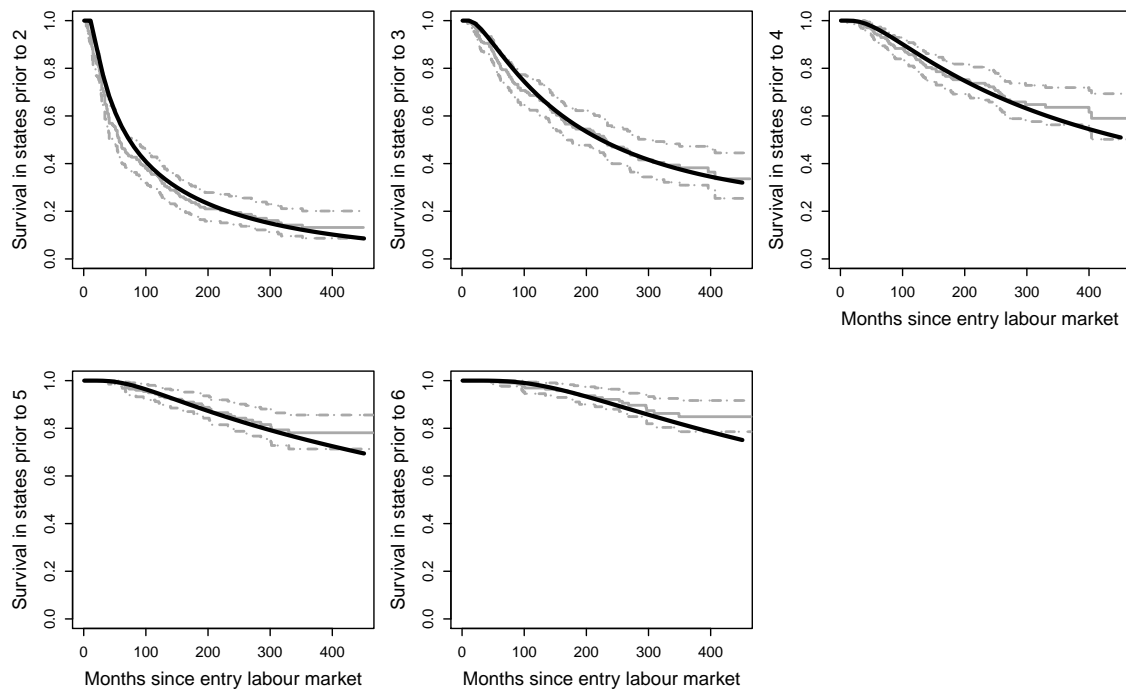


Figure 4: Comparison of predicted model-based survival and Kaplan-Meier survivor curves. Solid lines for mean of predicted survival given baseline GLHS data for the 201 individuals. Dashed lines for the Kaplan-Meier curves (with 95% confidence intervals).

For survival prior to state 2 and for survival prior to state 3, the model-based curves are close to the Kaplan-Meier curves. For the higher states, Figure 4 shows some lack of agreement at the later times—part of which is due to the right-censoring. Using the non-parametric Kaplan-Meier estimation as a data summary, the overall similarity with the model-based prediction validates the main features of the fitted model.

Hazards such as the ones in Figure 3 inform about employment history in the sense that they show the differences between the classes and the rates of moving to a next job. Transition probabilities can be used for additional inference. For example, to compare the two classes, we can look at the probabilities for changing states within a

five-year period. Say we do this for a man in birth cohort 1949-51, with more than 12 years of education, who has been in the labour market for five years, and is currently in his second job. The probability to move to (or stay in) state $s \in \{2, 3, 4, 5, 6\}$ is given by

$$p_{2s}(t_1 = 60, t_2 = 120 \mid \text{class} = k, (x_1, x_2) = (0, 1), x_3 = 1, x_4 = 1),$$

which we will denote p_{2s} , for short. If this man is in class $k = 1$, then we get

$$p_{22} = 0.231, \quad p_{23} = 0.343, \quad p_{24} = 0.226, \quad p_{25} = 0.097, \quad \text{and} \quad p_{26} = 0.025.$$

If this man is in class $k = 2$, then we get

$$p_{22} = 0.572, \quad p_{23} = 0.283, \quad p_{24} = 0.052, \quad p_{25} = 0.005, \quad \text{and} \quad p_{26} = 0.001.$$

For the computation of these probability, we used a piecewise-constant approximation (defined by a one-month grid) to take the non-constant hazards into account. These transition probabilities illustrate the differences between the classes on the scale of probabilities. For example, we see that for staying in the second job, the difference between the classes already quite large (0.231 vs. 0.572 for class 1 and 2, respectively). When in class 1, the man has a consistently higher chance to be in another job after five years then when he is in class 2, illustrating that class 1 is the class of the movers.

If we want to allocate the above specified man (with identifier i , say) to class $k \in \{1, 2\}$ we can compute the probability

$$\begin{aligned} & P(i \in \text{class } k \mid \mathbf{t}_i, \boldsymbol{\delta}_i, \mathbf{y}_i, \mathbf{x}_i; \boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}) \\ &= \frac{g(y_{i1}, \dots, y_{i2} \mid y_{i1}, i \in \text{class } k; \boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}) P(i \in \text{class } k \mid \boldsymbol{\theta} = \widehat{\boldsymbol{\theta}})}{\sum_{\ell=1}^{K=2} g(y_{i1}, \dots, y_{i2} \mid y_{i1}, i \in \text{class } \ell; \boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}) P(i \in \text{class } \ell \mid \boldsymbol{\theta} = \widehat{\boldsymbol{\theta}})}, \end{aligned}$$

where $g(y_{i1}, \dots, y_{i2} \mid y_{i1}, i \in \text{class } k; \boldsymbol{\theta} = \widehat{\boldsymbol{\theta}})$ is short-hand notation for the likelihood contribution for individual i conditional on being in class k ; see Section 5. For

example, if i is a job hopper in the sense that he changes job every year directly with $\mathbf{t}_i = (0, 12, 24, 36)$ and $\mathbf{y}_i = (1, 2, 3, 4)$, then the probability to be in class 1 is estimated at 0.90 and we would allocate i to class 1. However, if i changes job every 5 years, that is, $\mathbf{t}_i = (0, 60, 120, 180)$ with the same \mathbf{y}_i , then this probability is estimated at 0.36 and we would allocate i to class 2.

To summarise the GLHS data analysis, Model \mathcal{A} shows that inference can improve by fitting generalised gamma distributions in a model for employment history with *time since entry into the labour market* as the time scale. Effects of cohort, gender and education are important and have to be taken into account in the statistical inference. In addition, we illustrated that the latent-class Model \mathcal{B} is an interesting extension of Model \mathcal{A} , and allows to explore an assumed latent-class structure. Fitting a two-class structure shows a clear distinction between stayers and movers with respect to employment history. One possible way to extend the analysis would be to investigate covariate effects for class membership. For example, the probability to be a mover (π_1) could be modelled using logistic regression with gender as a covariate.

7 Conclusion

This paper illustrates the use of the generalised gamma distribution for a progressive multi-state process in demography. In addition, it is shown that an extension to a latent-class structure is possible and can lead to improved statistical inference.

Because the multi-state process in the application is progressive, the likelihood function for the time-dependent model can be constructed using transition-specific hazard functions and state-specific survivor functions. For multi-state processes where back-

and-forth transitions between the states are possible, the distributional options are the same, but the estimation becomes more complex. For such a case, we would propose to use a piecewise-constant approximation to the continuous-time parametric shape. The approximation would consist of a series of exponential models with changing hazard specifications; see, for example, [Blossfeld and Rohwer \(2002\)](#) and [Van den Hout \(2017\)](#).

This paper combines parametric hazard models with a discrete distribution for unobserved heterogeneity. Other options are possible. [Putter et al. \(2007\)](#) discuss a semi-parametric alternative for the fixed-effects models, and [Hougaard \(2000\)](#) discusses parametric choices for the distribution for unobserved heterogeneity. Advantages of parametric hazard models are efficiency and the option to predict the process outside the range defined by the data. With respect to the discrete distribution for unobserved heterogeneity, as long as this distribution is defined by a few classes, this option is computationally advantageous and, of course, it circumvents the need to specify a parametric shape for a latent characteristic.

Acknowledgements

The creation of Figure 2 is based on R code on www.everydayanalytics.ca, which is run by Myles Harrison (assessed on 24 September 2017). The authors would like to thank two anonymous reviewers for their detailed feedback.

References

- Aitkin, M. (1999). A general maximum likelihood analysis of variance components in generalized linear models. *Biometrics*, **55**, 117–128.
- Bartolucci, F., Farcomeni, A., and Pennoni, F. (2012). *Latent Markov Models for Longitudinal Data*. London: Chapman & Hall/CRC.
- Bijwaard, G. E. (2014). Multistate event history analysis with frailty. *Demographic Research*, **30**, 1591–1620.
- Blossfeld, H.-P. and Rohwer, G. (2002). *Techniques of Event History Modeling. New Approaches To Causal Analysis. Second Edition*. Mahwah: Lawrence Erlbaum Associates.
- Bongaarts, J. (2005). Long-range trends in adult mortality: Models and projection methods. *Demography*, **42**, 23–49.
- Cole, B. F., Bonetti, M., Zaslavsky, A. M., and Gelber, R. D. (2005). A multistate Markov chain model for longitudinal, categorical quality-of-life data subject to non-ignorable missingness. *Statistics in Medicine*, **24**, 2317–2334.
- Commenges, D. (1999). Multi-state models in epidemiology. *Lifetime Data Analysis*, **5**, 315–317.
- Cox, C., Chu, H., Schneider, M., and Muñoz, A. (2007). Parametric survival analysis and taxonomy of hazard functions for the generalized gamma distribution. *Statistics in Medicine*, **26**(23), 4352–4374.

- Dias, J. G. and Willekens, F. (2005). Model-based Clustering of Sequential Data with an Application to Contraceptive Use Dynamics. *Mathematical Population Studies*, **12**, 135–157.
- Frydman, H. (1984). Maximum likelihood estimation in the mover-stayer model. *Journal of the American Statistical Association*, **79**, 632–638.
- Hougaard, P. (2000). *Analysis of Multivariate Survival Data*. New York: Springer.
- Jackson, C. H. (2011). Multi-state models for panel data: the msm package for R. *Journal of Statistical Software*, **38**.
- Jackson, C. H. (2016). flexsurv: A platform for parametric survival modeling in r. *Journal of Statistical Software*, **70**.
- Kalbfleisch, J. and Lawless, J. F. (1985). The analysis of panel data under a Markov assumption. *Journal of the American Statistical Association*, **80**, 863–871.
- Kay, R. (1986). A Markov model for analysing cancer markers and disease states in survival studies. *Biometrics*, **42**, 855–865.
- Mandel, M. (2013). Simulation-based confidence intervals for functions with complicated derivatives. *The American Statistician*, **67**, 76–81.
- Mayer, K. U. (2015). The german life history study – an introduction. *European Sociological Review*, **31**(2), 137–143.
- Mueller, L. D., Nusbaum, T. J., and Rose, M. R. (1995). The gompertz equation as a predictive tool in demography. *Experimental Gerontology*, **30**, 553–569.
- Olshansky, S. J. and Carnes, B. A. (1997). Ever since gompertz. *Demography*, **34**, 1–15.

- Prentice, R. L. (1974). A log gamma model and its maximum likelihood estimation. *Biometrika*, **61**(3), 539–544.
- Putter, H. and Van Houwelingen, H. C. (2015). Frailties in multi-state models: Are they identifiable? Do we need them? *Statistical Methods in Medical Research*, **24**, 675–692.
- Putter, H., Fiocco, M., and Geskus, R. B. (2007). Tutorial in biostatistics: competing risks and multi-state models. *Statistics in Medicine*, **26**, 2389–2430.
- R Core Team (2013). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. URL <http://www.R-project.org>.
- Satten, G. A. (1999). Estimating the extent of tracking in interval-censored chain-of-events data. *Biometrics*, **55**, 1228–1231.
- Stacy, E. W. (1962). A generalization of the gamma distribution. *The Annals of Mathematical Statistics*, **33**(3), 1187–1192.
- Van den Hout, A. (2017). *Multi-State Survival Models for Interval-Censored Data*. Boca Raton, FL: Chapman & Hall/CRC.
- Vermunt, J. K., Langeheine, R., and Böckenholt, U. (1999). Discrete-time discrete-state latent Markov models with time-constant and time-varying covariates. *Journal of Educational and Behavioral Statistics*, **24**, 179–207.
- Willekens, F. (2014). *Multistate Analysis of Life Histories with R*. New York: Springer.