Chapter 7

Modeling the Genesis of Life Courses

Sapphire Yu Han and Cees H. Elzinga

7.1 Introduction

Life Course Models

Life course research as a paradigm started in the early seventies with the stepping stone work of Glen Elder and his collaborators. Since then, the methods and models used have been descriptive. Descriptive in the usual sense of describing life courses of particular social groups in particular epochs (e.g. Elder 1974) or describing frequent patterns seen in particular social strata (e.g. McVicar and Anyadike-Danes 2002) or age-groups (e.g. Brzinsky-Fay 2007). Descriptive in the wider sense also: by statistically inventorying the effects of certain variables on the occurrence and timing of particular events and outcomes (e.g. Eerola and Helske 2016). Descriptive research is fine, is necessary and often is very difficult. Descriptive research is where everything starts. If we agree on the descriptions, we can start modelling what we described. However, we should not confuse a list of variables that play a role in the genesis of data with a model of the genesis of those data. For example, when we model the risk of first parenthood, we might use covariates like education, religion, etc. and summarize their effects in a regression model that describes the rate changes of first parenthood as produced by these covariates; a hazard model (e.g. Kalbfleisch and Prentice 1980; Allison 1982; Vermunt 1997). It will then appear that some of the variables included do affect the odds and others do not. After some decades of research in many countries and using different scales for educational attainment, we will generally agree that education is relevant and some other variables are not. However, the fact that we include education in the inventory,

S. Y. Han (✉)

Directorate for Employment, Labour and Social Affairs, Organisation for Economic Co-operation and Development, Paris, France

C. H. Elzinga

Netherlands Interdisciplinary Demographic Institute (NIDI), The Hague, The Netherlands

perhaps even with an indicator of the relative strength of the effect of changes in the variable “education”, does not prove that we understand, that we modelled the mechanism - the process of becoming a parent - that uses “education” and produces the rate function. That we eventually include interaction of education and religion only means that we acknowledge that these variables cannot act in an additive (or multiplicative) way to generate the rate function – it does not specify the guise of this non-additivity. If we have an idea about the generating mechanism at all, such an idea is exogenous to the hazard function; at best, the (testing of the) hazard model is a consequence of the hypothesized, informally stated mechanism.

Today, we generally agree about the effects that many variables have on the shaping and outcomes of our life courses. So, it is time to start modelling the process of this shaping, the process that generates life courses. And this is quite a formidable task since it involves many correlated micro- and macro-variables that are expensive to measure and expensive to process in the calculations with such models. This paper is an attempt to use a versatile class of models as one of the central building blocks of a general life course model; the class of Hidden Markov Models (HMM’s for short).

Life-Course Data

Life courses are very complex narratives that live in our minds in a form that is hard to capture or communicate, even to those who are dear us. Here, we confine to the first, seemingly unavoidable simplification of such narratives; their encoding as multivariate time series. Multivariate, since the life course, understood as a series of events (e.g. marriage), spells (e.g. education) or states (e.g. parenthood), consists of “careers” in many aspects of life: family/household formation, education, housing/migration, employment, health/care, school-to-work, labor-to-pension, etc. And even these channels may be further factored or refined, for example in the guise of a fertility history or a series of incomes.

Seen in this way, the life course consists of a time series of the form

$$\mathbf{x} = \mathbf{x}_1 \mathbf{x}_2 \dots \mathbf{x}_t \dots \quad (7.1)$$

wherein each array $\mathbf{x}_t$ is of the guise

$$\mathbf{x}_t = \begin{pmatrix} v_{t1} \\ v_{t2} \\ \vdots \end{pmatrix} \quad (7.2)$$

and the v_{ti} denote observations expressed as numerical or categorical variables. For example, such an array could look like

$$\begin{pmatrix} M \\ P \\ 3 \\ R \end{pmatrix}$$

meaning that the person was, at that time t , Married, in Part-time employment, had 3 kids and was living in a Rural area.

We will say that a model is a general life course model, a GLCM for short, precisely when it generates time series of the form as specified above. We should however not forget that this scientific simplification is not totally congruent with our private experience of what we feel is our own life course.

Model Requirements

If one tries to think about building a GLCM, one has to make sure that it generates life courses in accordance with what 45 years of life course research has unveiled since the stepping stone work of Glen Elder (1974). This implies that GLCM's must at least satisfy a number of seemingly simple properties:

- (a) Individual agency is the capacity of individuals to independently make free choices (e.g. Hewson 2010). Any GLCM should explicitly recognize agency as one of the main driving forces of the mechanism that generates the life course. Free decisions require a decision-making process and such cognitive-emotional processes are essentially unobservable – only the results can be observed. Therefore, a GLCM should contain a latent process that represents individual decision making.
- (b) Structure is the complex of arrangements that affect or limit agency (e.g. Bourdieu 1977). Structure operates on the macro-level (Giddens 1984) as well as on the micro-level. Therefore, a GLCM should be able to recognize these structure effects as covariates that act on the decisions that the individuals take and on the effects that these decisions have on later outcomes.
- (c) Many choices we make and events we experience early in life have severe consequences in later life and this is true in most facets of the life course, be it a job career (e.g. Arulampalam 2001), family formation (e.g. Dronkers and Härkönen 2008) or the healthcare history (e.g. Walker et al. 1999). Therefore, a GLCM should have some sort of memory to account for time-lagged effects.
- (d) We know that, even within very specialized strata and narrow time windows, life courses show an enormous variability in the occurrence, timing and duration of many kinds of states and events. On the other hand, we also know that life courses come in a few, dominant types or classes (e.g. Ritschard and Oris 2005). A GLCM should be able to reproduce the variation observed and the most frequent patterns therein.
- (e) We know that the probability of the most important transitions taking place during the life course, is age-dependent. For example, elderly have different rates of partnering than young adults (e.g. Sassler 2010). Such age-dependencies should be recognized by a GLCM.

- (f) We know that various aspects of the life course are correlated (e.g. Liefbroer and Corijn 1999; Rindfuss et al. 1996) and such correlations between the variables v_i appearing in eq. (7.2) should be reproduced by a GLCM.
- (g) We also know that life courses of different people are linked, within families (e.g. King and Elder Jr. 1995; Liefbroer and Elzinga 2012) and in broader social structures. A GLCM should be able to accommodate such linkages.

Taken together, these simple requirements imply GLCM's of tremendous complexity, logically and computationally, even if we confine to modelling just one aspect of the life course and observe it through just one kind of measurement. Even such restricted models will have to satisfy the first five of the above requirements a-g.

Outline

In the next section, we begin our exposition by concisely discussing well-known methods of analyzing life course data against the background of the requirements put up above. Then we informally introduce the concept of an HMM and some issues related to parameter estimation and interpretation of its structural components. In the third section we discuss an application of HMMs to family formation data (Han et al. 2020) and illustrate some of the issues discussed in Sect. 7.2. In Sect. 7.4, we discuss using the potential of HMMs as building blocks in a GLCM: HMMs can be used to model correlation between trajectories in different life-course domains like family formation and labor market careers and to model the interaction between life courses within different social strata, i.e. to model linked lives.

7.2 Methods for Life-Course Analysis

Over the past decades, life course analysis has been dominated by two paradigms: Event History Analysis (EHA) and Sequence Analysis (SA). An interesting application of both EHA and SA to the same research question can be found in Eerola and Helske (2016).

EHA (e.g. Blossfeld et al. 2007) amounts to predicting the (non-)occurrence and timing of life course events or very short sequences of such events through logistic regression models and such models can be adapted to accommodate for autocorrelation (Feijten and Mulder 2002). However, EH multistate models are not suitable to predict complete, extended life courses of the form of (7.1)–(7.2) and do not have a component that represents a cognitive process.

SA (e.g. Cornwell 2015) is not a model at all; it is a toolbox to describe life courses and identify frequent patterns among these. Distances used to construct such frequent patterns or clusters can be used to test hypotheses about the effect of covariates on cluster-membership (Studer et al. 2011) but there is no mechanism that generates or accounts for the sequences or the variability within clusters.

Recently, we have seen other models employed in life course research. Barban and Billari (2012) proposed Latent Class Analysis (LCA) and e.g. Pakpahan et al. (2017) proposed to use Structural Equation Models (SEM's). LCA is the

probabilistic variant of SA's distance-based clustering (Han et al. 2017) and generates variation of life courses within classes but it does not have a component that could represent a decision-making process and it has no memory. Even the sophisticated models proposed by Pakpahan and his collaborators do not satisfy all of the requirements stated in a.-e. above.

Therefore, we here discuss a broad class of models that satisfies all of the requirements as stated in the Introduction: the class of Hidden Markov Models (HMM's) (Fig. 7.1).

HMM's consist of two main components: a Markov chain over a set of unobservable, latent states and a set of probability distributions over a set of multivariate observables (e.g. marital status, residence, labor market status, health condition, income, etc.), one such distribution for each latent state. When the system is in a latent state, it will generate or emit precisely one observable according to the associated probability distribution. The latent states are thought to represent the unobservable, individual decision-making processes, pertaining to the decisions (agency) to be taken and the observables (multivariate or univariate) result from these decisions. Covariates (structure) may affect the decisions and the state-switching. The state-switching process, i.e. the Markov chain, has a memory: the probability of a switch to a particular state depends on the state-history of the process – even if only the last state occupied is relevant (a first-order chain), such dependency may create long-term auto-correlation in the state-switching. If necessary, generalizations of Markov-models (e.g. Pegram 1980; Berchtold and Raftery 2002) can be used to explicitly model longer time-lags (Fig. 7.2).

Because of the probabilistic nature of the process, it will generate a great variety of sequences of observables, even for one and the same path along the latent states. If the Markov chain is well parameterized, it will generate a few paths that are more frequent than others, thus generating distinct classes of observable sequences that are, within classes, quite similar but not identical. In the simplest version of an HMM, the state transition probabilities are not age-dependent. However, as long as particular states, i.e. life course decisions are taken roughly at the same age for subjects that belong to the same cohort, using a single, time-homogeneous latent Markov chain to model the life courses of that cohort will not suffer from this feature. If necessary, time-inhomogeneous models can be formulated or even models wherein the waiting time distributions for state transitions can be made

Fig. 7.1 A state-switching circuit, representing a first-order, 2-state Markov chain and its transition probability matrix. The states are labeled as “0” and “1” and the arrows represent the transition probabilities

Fig. 7.2 A graph showing the time-window $(t-I, t+I)$ of a Hidden Markov Model. At each time t , the system is at some latent state $S_t = q_j$ and emits an observable O_t . Note that the hidden state S_t is not necessarily different from S_{t+1} . The observable is a random sample from the set of observables, according to a probability distribution that is specific for each state $q_i, i = 1, \dots, k$. Covariates v_i and v_j may affect both the state switching and/or the sampling of observables

age-dependent (e.g. Dewar et al. 2012). Multi-channel life course data can be modelled by correlated latent chains and linked life courses of related people could be modelled by linked HMM's like the ones proposed in Elzinga et al. (2007).

The statistical theory of estimating HMM's originates from the work of Leonard Baum and his collaborators (e.g. Baum et al. 1970) and the classical paper by Lawrence Rabiner (1989) popularized and stimulated the use and further development of the statistical theory. Today, there is an abundance of literature on the theory and estimation of even very complex models and variants. Good introductory texts are Bartolucci et al. (2012) and Zucchini et al. (2016).

7.3 The Nomological Net as a Testing Environment

Estimating an HMM is problematic because of the irregular shape of the surface of the likelihood function in the parameter space. Judging the fit of the model is problematic because of the large number of parameters. Therefore, as we will argue in this section, we have to study the adequacy of a particular HMM in a nomological net (e.g. Preckel and Brunner 2017) to decide on the credibility of this particular model as a valid explanation or rendering of life course genesis. First, we discuss some problems in judging model-fit and then we discuss the nomological net and the role of HMM's therein.

Problems in Fitting HMMs

Although today, the theory and estimation technique of HMM's is well developed and fast computers are generally available, even in the social sciences, estimating HMM's is still quite a challenge.

First, the number of free parameters of an HMM is big: with k latent states and n observables, this number amounts to $k^2 + kn - k - 1$ and often these parameters will be quite different in size. Thus, the surface of the likelihood function involved in the associated maximization problem will be quite irregular and so, we may expect that this optimization will often be trapped on a local maximum. Therefore, it is advisable to repeat the calculations for many different sets of initial guesses and hope that one finds a configuration that is close to the maximum sought for.

Second, because of this big number of parameters, there is, even for small, restricted models, always a parameter configuration that generates predicted distributions that closely approximate the observed distributions (e.g. Eggar 2002).

Finally, an important structural parameter of an HMM is the number k of latent states: all estimates of the initial-, transition- and emission-probabilities are performed, given the fixed, user-defined number k .

Therefore, the absolute fit of an HMM cannot be tested statistically and so one has to rely on comparisons of models with different numbers of latent states and/or differently restricted probabilities and these comparisons have to use some likelihood-related criterion like BIC or AIC (e.g. Burnham and Anderson 2002). However, such likelihood-based model-selection is possible only when the plot of such criteria against the number of model parameters shows a clear knee. Unfortunately, such pronounced knees are often hardly distinguishable.

Hence, the problem arises of how to deal with HMM's as life course models. Not using them because of the above-mentioned problems, seems a waste of descriptive and explanatory power: there are, as far as we know, no modelling tools that satisfy all of the requirements to be imposed upon GLCM's. On the other hand, we have to face the fact that often HMM's cannot be used to statistically test causal structures.

Nomological Nets and Theory Testing

A nomological net (Fig. 7.3) is a representation of at least two concepts, operationalisations in the form of data, and the theories and methods that formulate and unveil the relations between all these entities (e.g. Cronbach and Meehl 1955; Preckel and Brunner 2017). If generally accepted, we call such a network “a body of facts”. Here the relevant network is comprised of concepts like “life course”, “linked lives”, “agency”, “adulthood”, etc., data like produced within the Generations and Gender Program or as available from the National Longitudinal Survey of Youth

Fig. 7.3 Rendering of a general nomological net with theories (concepts and their relations), data (measured operationalizations of concepts) and methods (analytical procedures to test theories) & tools (data-rendering techniques). For explanation, see text

(NLSY97; Moore et al. 2000) or narratives from homeless in a big city, methods like SA (implicitly relying on theories of similarity between separate life course events and the similarity between time series) and EHA (using a variants of local independence), demographic theory (Second Demographic Transition theory), sociological theories (Blau & Duncan's theory on status attainment, Bourdieu's field theory, models and methods of educational attainment, etc.). All these concepts, data, theories and methods are linked in various ways.

Of course, the relations between data, theories and methods must be consistent: different methods applied to the same data should not lead to incompatible results, different theories should not lead to predictions that are inconsistent, etc. Attempts to construct GLCM's probably cannot rely on statistical testing or model selection techniques but instead will be validated through a confrontation with the "body of known facts" as emerges from the social-demographic nomological net.

So, if we know from several datasets and survival analysis that fertility histories are affected by gender and education, then estimated HMMs with gender and education as covariates, should produce the same effects. If this would not be the case, these HMMs would be incompatible with the nomological net of social demography and thus cannot be valid renderings of family formation. Therefore, we can use the social-demographic nomological net to look for covariates and test the HMMs to see whether or not these models reproduce known effects. This exactly what we will demonstrate in the next section with an application of HMMs to family formation trajectories. We will use modern Second Demographic Transition theory to decide between competing models and re-estimate the model with education and gender as covariates.

7.4 Modelling Family Formation and HMM: An Application

For this application (for details, see Han et al. 2020) we used data from the Generations and Gender Programme (Fokkema et al. 2016), a longitudinal survey among 18–79 year olds in nineteen countries. We selected 1900 French subjects (56% female, 19% higher educated), born between 1956 and 1965 and constructed fertility histories (0, 1, 2 or more children), partnership histories (living single, in cohabitation or married) and a binary trajectory for having or not having left the parental home.

Second Demographic Transition Theory (SDT, see e.g. Lesthaeghe 1995) predicts that, after or upon leaving the parental home, people consider family formation trajectories that either do or do not involve cohabiting. Hence, family formation would involve at least three and at most five decisions, as tabulated in Table 7.2; modelling this decision process by an HMM would thus require 5 latent states. Therefore, we estimated (using LMest, see Bartolucci et al. 2017) a 5-state HMM. In Fig. 7.4, we show the state switching circuit (the probabilities of "self-transitions" omitted) of the estimated model, in Fig. 7.6 we show the state occupancy plot and

Fig. 7.4 Circuit switching diagram of the estimated 5-state HMM. The thickness of the arrows reflects the probabilities (shown in small print) of transition to another state (based on Han et al. 2020)

Table 7.1 Estimated emission probability distributions of a 5-state HMM (based on Han et al. 2020)

Nr. of children/State	1	2	3	4	5
0	1	1	.40	0	0
1	0	0	.60	.53	0
2	0	0	0	.32	.67
3+	0	0	0	.15	.33
Partnership/state	1	2	3	4	5
Single	.98	.67	0	.42	0
Cohabiting	.02	.33	0	.58	0
Married	0	0	1	0	1
Left home/state	1	2	3	4	5
Yes	1	0	.02	.01	.01
No	0	1	.98	.97	.99

in Table 7.1, we show the estimated probabilities of emitting observables, conditional upon state occupancy.

From Table 7.1, we observe that in latent state 1 (LS1), none of the subjects has left the parental home while in LS2 and LS3, almost all have. Furthermore, we observe that in LS2, none have married while in LS3, all have. From Fig. 7.4, we see that the only transitions from LS1 are to either LS2 or to LS3. So, we conclude that the decision taken in LS1 is the decision on how and when to leave the parental home. Similar reasoning, using the emission probabilities and the state-switching circuit, leads to the interpretations of the decision processes of the other latent states, shown in Table 7.2. Apparently, a 5-state HMM fits in well with modern

Table 7.2 Interpretation of latent states in a 5-state HMM in terms of observed demographic events and mental decision processes (adapted from Han et al. 2020)

State	Observation	Decision process
LS1	Living in parental home	When/how to leave parental home
LS2	Childless, not married, Residential-independent	Mode of partnership
LS3	Married with 0 or 1 child	Family extension
LS4	Non-marital family	Marriage
LS5	Multi-child family	Staying together

Fig. 7.5 State occupancy plot of the estimated 5-state HMM (based on Han et al. 2020)

demographic notions about the decline of traditional family values. In Fig. 7.5, we show the state-occupancy plot of the estimated 5-state HMM.

The validity of the model is further corroborated by the results of including gender, education and the interaction of these effects as covariates. We know that men experience marriage and parenthood later than women (Aassve et al. 2002; Andersson and Philipov 2002). We also know that the higher educated generally delay marriage and parenthood (Kravdal and Rindfuss 2008; Liefbroer and Corijn 1999) but also that in many countries, the higher educated are very reluctant to enter

Table 7.3 Estimated odds of transitions in a 5-state HMM with gender, education and their interaction in the form of weights of a logistic regression equation. The odds that are significantly ($p < 0.1$) different from 1 are shown in bold (adapted from Han et al. 2020)

Transition	Male	High education	Interaction
LS1→LS2	1.06	1.41	0.82
LS1→LS3	0.40	0.40	1.25
LS2→LS3	0.78	0.74	1.56
LS2→LS4	0.72	0.50	1.16
LS3→LS5	0.88	1.26	0.91
LS4→LS3	1.46	0.53	2.00
LS5→LS4	0.64	0.73	1.50

unmarried cohabitation (Perelli-Harris et al. 2010). We estimated the effects of these covariates on the state transition rates through logistic regression and obtained results as summarized in Table 7.3.

Clearly, the covariates affect the transition probabilities in a way that is in accordance with what we know about their effects from other studies, therewith validating the application and interpretation of the 5-state HMM.

7.5 Linked Trajectories – Linked HMMs

Often, life course trajectories are linked in the sense that the trajectories affect each other. Such linkages can be intra-personal as for example in the case of labor market careers and family formation trajectories where choices in the one area affect the choices in the other area and vice versa. Life course linkages can also be inter-personal as for example in the case of “linked-lives” where e.g. parent’s family formation patterns seem to affect children’s family formation patterns (e.g. Liefbroer and Elzinga 2012). At the same time, such linkages may be mutual or non-mutual. From these two dichotomies – inter- vs. intra-personal and mutual vs. nonmutual – arises a simple classification of types of linkages and, as will appear in the sequel, these types require different modelling. Intra-personal linkages will probably always be mutual: it is hard to imagine that, within the same person, the course of trajectory A will affect trajectory B while the course of trajectory B will not affect trajectory A. However, if the trajectories belong to different persons, the trajectories may mutually affect each other as for example in behavioral sequences of therapists and patients or of negotiating parties. On the other hand, parents’ patterns of family formation may affect children’s patterns while it is hard to see how children’s family formation patterns would influence the parent’s trajectories. We summarize these considerations in Table 7.4 and discuss some consequences for modeling with HMMs below.

Table 7.4 Classification scheme for linked life course trajectories

	Intra-personal	Inter-personal
Mutual	Type I	Type II
	e.g. family formation and labor market careers	e.g. family formation in peer groups
Non-mutual	Type IV	Type III
	?	e.g. family formation in parents and children

Fig. 7.6 Intra-personal linkage of HMMs A and B. Circles represent latent states, squares represent observables. Hatched arrows mean “is affected by”

Intra-Personal Linkage

If life course genesis can be represented by HMMs of which the latent states represent unobservable decision processes, it is only natural that decisions in one domain affect decisions in the other domain and that it is the interacting state switching processes that produce the correlation between the trajectories observed. Perhaps sometimes it could be argued that there is interaction between the emission processes too but this is not a very parsimonious assumption, neither conceptually, nor from the perspective estimating all the parameters involved and the precision of that estimation given the number of observed sequences. Therefore, the most promising and parsimonious hypothesis is that intra-personal linkage is caused only by interacting state switching, i.e. a model structured as depicted in Fig. 7.6.

Inter-Personal Linkage

Possible HMM-based models for interpersonal linkage must be structurally different from models for intra-personal linkage; different persons have no access to each other’s mental processes – different persons only have access to the observables emitted from the other’s decision processes.

Thus, the only way that the one HMM may affect the other HMM is via it’s emitted observables. But these observables may affect both the other HMM’s state switching and/or the other HMM’s emittance process. We summarize this structure

Fig. 7.7 Mutual inter-personal linkage (Type II) of HMMs A and B. Circles represent latent states, squares represent observables. Hatched Arrows mean “is affected by”

in Fig. 7.7. The reader notes that in Fig. 7.7, depending on the substantive process modeled, HMMs A and B need not be symmetric and need not to contain both of the outgoing hatched arrows.

A schematic representation, analogous to Fig. 7.7, of non-mutual inter-personal linkage arises by erasing either of the two bundles of hatched arrows.

Discussing the logistic-regression equations to estimate the models implied by Figs. 7.6 and 7.7, is beyond the scope of this chapter; the interested reader is referred to Zucchini et al. (2016).

7.6 Summary and Conclusions

We set out to model the production, the genesis of life course sequences. This is different from modelling the variance of life course related events. This genesis modelling is different since its primary purpose is to test or to validate a mechanism that produces sequences that are close to sequences as observed, that are realistically affected by covariates and are correlated between life course domains. If such modelling is successful, one of its results is a good approximation of the variance of life courses themselves and their constituting events, timings and orderings, i.e. of the results of regression-based methods that produce decompositions of observed variances. In that sense, the modelling of the generative mechanism is more encompassing than designing an (additive) decomposition of variances and covariances. On the other hand, regression-based models will continue to be relevant to unequivocally establish the (relative) contribution of the components, the independents, a property that is indispensable to explore, to inventory a research area or to motivate policy-decisions.

Thus, we started listing the requirements that models of the mechanism of life course generation should satisfy and arguing that none of the methods presently used in life course research can satisfy all of these requirements. We then informally introduced HMM's as a class of models that could be used to fulfill these requirements, i.e. generate realistic, data-like sequences.

To demonstrate the power of these models, we applied them to study family formation in relation to SDT-theory and validated the model using gender and education as covariates. We feel that this demonstration is convincing, despite the fact that we hardly discussed problems of estimation and model selection and did not explore more sophisticated models that would allow for, e.g. age-graded transition rates or variable time-lags. We also tried to guide the reader in modeling correlation of trajectories by presenting a typology of linkages and the consequences thereof for modeling such linkages

We hope to compensate these shortcomings in future research and that others will help exploring the power and potential of HMM's in related areas of life course research.

Acknowledgements The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013)/ERC Grant Agreement n. 324178.

References

- Aassve, A., F.C. Billari, S. Mazucco, and F. Ongaro. 2002. Leaving home: A comparative analysis of ECHP data. *Journal of European Social Policy* 12 (4): 259–275.
- Allison, P.D. 1982. Discrete-time methods for the analysis of event histories. *Sociological Methodology* 13: 61–98.
- Andersson, G., and D. Philipov. 2002. Life-table representations of family dynamics in Sweden, Hungary and 14 other FFS countries: A project of descriptions of demographic behavior. *Demographic Research* 7: 67–144.
- Arulampalam, W. 2001. Is unemployment really scarring? Effects of unemployment experiences on wages. *The Economic Journal* 111 (475): 585–606.
- Barban, N., and F.C. Billari. 2012. Classifying life course trajectories: A comparison of latent class and sequence analysis. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 61 (5): 765–784.
- Bartolucci, F., A. Farcomeni, and F. Pennoni. 2012. *Latent Markov models for longitudinal data. Statistics in the social and behavioral sciences*. Boca Raton, FL: CRC Press.
- Bartolucci, F., S. Pandolfi, and F. Pennoni. 2017. LMest: An R package for latent Markov models for categorical longitudinal data. *Journal of Statistical Software* 81: 2–38.
- Baum, L.E., T. Petrie, G. Soules, and N. Weiss. 1970. A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. *Annals of Mathematical Statistics* 41 (1): 164–171.
- Berchtold, A., and A.E. Raftery. 2002. The mixture transition distribution model for high-order Markov chains and non-Gaussian time series. *Statistical Science* 17 (3): 328–356.
- Blossfeld, H.-P., K. Golsch, and G. Rohwer. 2007. *Event history analysis with STATA*. Mahwah: Lawrence Erlbaum.
- Bourdieu, P. 1977. *Outline of a theory of practice*. Cambridge: Cambridge University Press.

- Brzinsky-Fay, C. 2007. Lost in transition? Labour market entry sequences of school leavers in Europe. *European Sociological Review* 23 (4): 409–422.
- Burnham, K.P., and D.R. Anderson. 2002. *Model selection and multi-model inference. A practical information-theoretic approach*. 2nd ed. Heidelberg: Springer.
- Cornwell, B. 2015. *Social sequence analysis: Methods and applications*. Cambridge: Cambridge University Press.
- Cronbach, L.J., and P.E. Meehl. 1955. Construct validity in psychological tests. *Psychological Bulletin* 52 (4): 281–302.
- Dewar, M., C. Wiggins, and F. Wood. 2012. Inference in hidden Markov models with explicit state duration distributions. *IEEE Signal Processing Letters* 19 (4): 235–238.
- Dronkers, J., and J. Härkönen. 2008. The intergenerational transmission of divorce in cross-national perspective: Results from the fertility and family surveys. *Population Studies* 62 (3): 273–288.
- Eerola, M., and S. Helske. 2016. Statistical analysis of life history calendar data. *Statistical Methods in Medical Research* 25 (2): 571–597.
- Eggart, M.H. 2002. Validity of fitting a first-order Markov chain model to data. *Journal of the Royal Statistical Society, Series D* 51 (2): 259–265.
- Elder, G.H. 1974. *Children of the great depression: Social change in life experience*. Chicago: University of Chicago Press.
- Elzinga, C.H., A.W. Hoogendoorn, and W. Dijkstra. 2007. Linked Markov sources: Modeling outcome-dependent social processes. *Sociological Methods & Research* 36 (1): 26–47.
- Feijten, P., and C.H. Mulder. 2002. The timing of household events and housing events in The Netherlands: A longitudinal perspective. *Housing Studies* 17 (5): 773–792.
- Fokkema, T., A. Kveder, N. Hiekel, T. Emery, and A.C. Liefbroer. 2016. Generations and Gender Programme wave 1 data collection: An overview and assessment of sampling and fieldwork methods, weighting procedures, and cross-sectional representativeness. *Demographic Research* 34: 499–524.
- Giddens, A. 1984. *The constitution of society: Outline of the theory of structuration*. Berkeley: University of California Press.
- Han, S.Y., A.C. Liefbroer, and C.H. Elzinga. 2017. Comparing methods of classifying life courses: Sequence analysis and latent class analysis. *Longitudinal and Life Course Studies* 8 (4): 319–341.
- . 2020. Mechanisms of family formation: An application of Hidden Markov Models to a life course process. *Advances in Life Course Research* 43: 100265.
- Hewson, M. 2010. Agency. In *Encyclopedia of case study research*, ed. A. Mills, G. Durepos, and E. Wiebe, 13–17. Thousand Oaks: Sage.
- Kalbfleisch, J.D., and R.L. Prentice. 1980. *The statistical analysis of failure time data*. Hoboken: Wiley.
- King, V., and G.H. Elder Jr. 1995. American children view their grandparents: Linked lives across three rural generations. *Journal of Marriage and the Family* 57 (1): 165–178.
- Kravdal, Ø., and R.R. Rindfuss. 2008. Changing relationships between education and fertility: A study of women and men born 1940 to 1964. *American Sociological Review* 73 (5): 854–873.
- Lesthaeghe, R. 1995. The second demographic transition in Western countries: An interpretation. In *Gender and family change in industrial countries*, ed. K. Oppenheim Mason and A.-M. Jensen, 17–62. Wotton-under-Edge: Clarendon Press.
- Liefbroer, A.C., and M. Corijn. 1999. Who, what, where, and when? Specifying the impact of educational attainment and labour force participation on family formation. *European Journal of Population* 15 (1): 45–75.
- Liefbroer, A.C., and C.H. Elzinga. 2012. Intergenerational transmission of behavioral patterns: How similar are parents' and children's demographic trajectories? *Advances in Life Course Research* 17 (1): 1–10.

- McVicar, D., and M. Anyadike-Danes. 2002. Predicting successful and unsuccessful transitions from school to work by using sequence methods. *Journal of the Royal Statistical Society, Series A* 165 (2): 317–334.
- Moore, W., S. Pedlov, P. Krishnamurti, and K. Wolter. 2000. *National Longitudinal Survey of Youth (NLSY97)*, Technical sampling report. Chicago: National Opinion Research Center.
- Pakpahan, E., R. Hoffmann, and H. Kröger. 2017. Statistical methods for causal analysis in life course research: An illustration of a cross-lagged structural equation model, a latent growth model, and an autoregressive latent trajectories model. *International Journal of Social Research Methodology* 20 (1): 1–19.
- Pegram, G.G.S. 1980. An autoregressive model for multi-lag Markov chains. *Journal of Applied Probability* 17 (2): 350–362.
- Perelli-Harris, B., W. Sigle-Rushton, M. Krayenfeld, T. Lâppegard, R. Keizer, and C. Berghammer. 2010. The educational gradient of childbearing with cohabitation in Europe. *Population and Development Review* 36 (4): 775–801.
- Preckel, F., and M. Brunner. 2017. Nomological nets. In *Encyclopedia of personality and individual differences*, ed. V. Zeigler-Hill and T. Shackelford. Cham: Springer.
- Rabiner, L.R. 1989. A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE* 77 (2): 257–286.
- Rindfuss, R.R., S.P. Morgan, and K. Offutt. 1996. Education and the changing age pattern of American fertility: 1963–1989. *Demography* 33 (3): 277–290.
- Ritschard, G., and M. Oris. 2005. Life course data in demography and social sciences: Statistical and data-mining approaches. *Advances in Life Course Research* 10: 283–314.
- Sassler, S. 2010. Partnering across the life course: Sex, relationships, and mate selection. *Journal of Marriage and the Family* 72 (3): 557–575.
- Studer, M., G. Ritschard, A. Gabadinho, and N.S. Müller. 2011. Discrepancy analysis of state sequences. *Sociological Methods & Research* 40 (3): 471–510.
- US Department of Labor, Bureau of Labor Statistics. The NLSY97. <http://www.bls.gov/nls/nlsy97.htm>.
- Vermunt, J.K. 1997. *Log-linear models for event histories*. Thousand Oaks: Sage.
- Walker, E.A., A. Gelfand, Wayne J. Katon, Mary P. Koss, M. Von Korff, D. Bernstein, and J. Russo. 1999. Adult health status of women with histories of childhood abuse and neglect. *The American Journal of Medicine* 107 (4): 332–339.
- Zucchini, W., I. MacDonald, and R. Langrock. 2016. *Hidden Markov models for time series: An introduction using R*. 2nd ed. Boca Raton: CRC Press.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.