

Studer, Matthias; Struffolino, Emanuela; Fasang, Anette Eva

Article — Published Version

Estimating the Relationship between Time-varying Covariates and Trajectories: The Sequence Analysis Multistate Model Procedure

Sociological Methodology

Provided in Cooperation with:

WZB Berlin Social Science Center

Suggested Citation: Studer, Matthias; Struffolino, Emanuela; Fasang, Anette Eva (2018) : Estimating the Relationship between Time-varying Covariates and Trajectories: The Sequence Analysis Multistate Model Procedure, Sociological Methodology, ISSN 1467-9531, Sage Publications, Thousand Oaks, CA, Vol. 48, Iss. 1, pp. 103-135, <https://doi.org/10.1177/0081175017747122>

This Version is available at:

<https://hdl.handle.net/10419/191543>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



ESTIMATING THE RELATIONSHIP BETWEEN TIME-VARYING COVARIATES AND TRAJECTORIES: THE SEQUENCE ANALYSIS MULTISTATE MODEL PROCEDURE

*Matthias Studer**
Emanuela Struffolino[†]
Anette E. Fasang[†]

Abstract

The relationship between processes and time-varying covariates is of central theoretical interest in addressing many social science research questions. On the one hand, event history analysis (EHA) has been the chosen method to study these kinds of relationships when the outcomes can be meaningfully specified as simple instantaneous events or transitions. On the other hand, sequence analysis (SA) has made increasing inroads into the social sciences to analyze trajectories as holistic “process outcomes.” We propose an original combination of these two approaches called the sequence analysis multi-state model (SAMM) procedure. The SAMM procedure allows the study of the relationship between time-varying covariates and trajectories of categorical states specified as process outcomes that unfold over time. The SAMM

*NCCR LIVES and University of Geneva, Geneva, Switzerland

[†]WZB Berlin Social Science Center and Humboldt University of Berlin, Berlin, Germany

Corresponding Author:

Matthias Studer, NCCR LIVES and IDESO, University of Geneva, 40, Boulevard du Pont-d’Arve, 1205 Geneva, Switzerland.

Email: matthias.studer@unige.ch

is a stepwise procedure: (1) SA-related methods are used to identify ideal-typical patterns of changes within trajectories obtained by considering the sequence of states over a predefined time span; (2) multistate event history models are estimated to study the probability of transitioning from a specific state to such ideal-typical patterns. The added value of the SAMM procedure is illustrated through an example from life-course sociology on how (1) time-varying family status is associated with women's employment trajectories in East and West Germany and (2) how German reunification affected these trajectories in the two subsocieties.

Keywords

event history analysis (EHA), sequence analysis, multistate model, German reunification, employment trajectories, life-course sociology

1. INTRODUCTION

Many theoretical questions in the social sciences address the relationship between longitudinal processes and time-varying covariates. Life-course and career researchers are interested in how changes in one life domain can influence trajectories in another (e.g., family and employment [Aisenbrey, Evertsson, and Grunow 2009]) and how changing economic conditions or family policies shape the transition to adulthood (Shanahan 2000). Social policy analysis is concerned with policy development processes that can be altered by specific events, such as wars or a change of government (Abbott 1995; Frank, Hironaka, and Schofer 2000). Related research questions are at the center of historical comparative sociology. Similarly, social movement scholars study how social movements unfold over time in response to trigger events (Minkoff 1995; Olzak 1989). Likewise, organizational ecology research examines how organizations develop over time: The relationship between processes and time-varying factors—for instance, the introduction of new technologies—is of core theoretical interest in this field as well (Carroll et al. 1993). These selected examples could be easily extended to other subfields of the social sciences. While the core units of analysis are located either on the micro, meso, or macro level, the similarity among all these examples is that they are not interested only in processes of metric outcomes, such as income or IQ. Instead, they focus on processes that consist of categorical states, including family trajectories, the implementation of specific policy programs, or the stages of group behavior prior to the outbreak of violent protest.

Two broad families of methodological strategies have been used to study the relationship between time-varying covariates and outcomes and trajectories of categorical states. The first strategy focuses on the occurrence of events or transitions (Allison 1984; Therneau and Grambsch 2000; Yamaguchi 1991) and relies on event history analysis (EHA) to estimate the effect of time-varying covariates on the risk of observing an event. However, the use of EHA is limited to modeling instantaneous changes, and it loses sight of the trajectory as a whole (Billari 2005). The second strategy emphasizes the holistic nature of trajectories or processes of categorical states by relying on sequence analysis (SA) (Abbott 1995; Studer and Ritschard 2016). While SA considers change and multiple transitions as lasting over longer time spans, in its traditional framework, studying the relationship between time-varying covariates and trajectories is impossible.

We propose an original combination of these two approaches, in what we call the *sequence analysis multistate model* (SAMM) procedure. The SAMM procedure is a stepwise application of (1) SA-related methods to identify ideal-typical trajectories understood as the sequence of states experienced by each individual during a given time span and (2) multistate event history models to study the probability of transitioning from a state to such ideal-typical trajectories. The SAMM procedure offers several advantages for studying processes. First, it allows for modeling the relationship between time-varying covariates and patterns of change within processes that unfold over long periods of time. This closely corresponds to the theoretical concept of trajectories as “process outcomes” (Abbott 2005). Second, studying trajectories holistically allows us to unveil potential interdependencies between states and transitions within trajectories. The social meaning of a given situation often depends on both previous and later events, which may be known in advance by the actors involved. For instance, an individual may start a new job although (or because) he or she knows that it will be only temporary. Finally, the SAMM procedure can handle censored observations, which is possible only to a very limited extent in the traditional SA framework, and this allows for inclusion in the analysis of trajectories that are only partially observed.

We demonstrate the added value of the SAMM procedure by using an original illustrative application in life-course sociology. Two theoretical principles in the life-course paradigm assume individual life-courses to be multidimensional (e.g., family and employment) and shaped by macrostructural and historical changes (Elder, Kirkpatrick

Johnson, and Crosnoe 2003). We employ the historically unique “social experiment” of German reunification to exemplify how the SAMM procedure can contribute to a better understanding of these two core life-course principles. Specifically, we assess how (1) time-varying statuses in the family domain are associated with women’s employment trajectories in East and West Germany (multidimensionality of life-courses) and (2) how the German reunification affected women’s employment trajectories in the two subsocieties (i.e., the impact of macrostructural change). Beyond previous research on life-courses during the German reunification (Bonin and Euwals 2002; Hauschild 2002; Klammer and Tillmann 2001; Mayer 2006; Trischler and Kistler 2010), our application uses data from more recent birth cohorts. This allows us to track differences and similarities in women’s employment trajectories in East and West Germany not only in the immediate transition period but also up to 20 years following reunification.

The remaining sections of this paper discuss the following topics. Section 2 briefly introduces background information on the German reunification. Section 3 reviews the complementary strengths and weaknesses of SA and EHA. Section 4 presents the SAMM procedure. In Section 5, we apply the procedure to our illustrative example and further provide robustness checks. We compare results from the SAMM procedure with standard MMs.

2. ILLUSTRATIVE EXAMPLE: THE GERMAN REUNIFICATION

Between 1955 and 1990, Germany was divided into the Federal Republic of Germany (FRG) in the West and the German Democratic Republic (GDR) in the East. The two subsocieties differed greatly in terms of both the ideational and institutional characteristics of their labor markets and welfare systems.

The GDR promoted a universal-breadwinner model within a communist egalitarian stratification system and aimed at achieving population growth through pronatalist family policies. The constitution guaranteed the “right and duty to work” (Kreyenfeld 2004). Women’s labor market participation rates reached 90 percent (Huinink et al. 1995). Almost-universal and day-long child care facilitated labor market participation among mothers (Huinink et al. 1995).

In the FRG, the male-breadwinner model was the core organizing principle of social policies (Brückner 2004). Tax-splitting among spouses reinforced incentives for marriage and a male-breadwinner/female-home-maker specialization (Fasang 2014). The infrastructure for public child care was limited, particularly for children under three years of age. In addition, the FRG was characterized by a normative climate in which mothers' employment was regarded as harmful to small children (Treas and Widmer 2000). During the decades before reunification, the female labor market participation rate in the FRG was only around 50 percent.

With reunification, the Eastern federal states adopted Western labor market institutions and social policies. The dramatic changes in the occupational structure and the destruction of about one-third of the jobs in the East led to persistent disadvantages in the economy of the former GDR (Goldstein and Kreyenfeld 2011; Kreyenfeld 2003). In the first years following reunification, the rapid privatization of the economy was counterbalanced by huge shares of subsidized jobs in the public sector, early retirement schemes, and generous social security transfers (Franz and Steiner 2000). However, after this short period, lower wages and higher unemployment rates continued to characterize the East compared with the West long after reunification (Goldstein and Kreyenfeld 2011). This challenged the initial expectations of progressive convergence toward a Western standard, following a critical adaptation period.

Most of the previous studies on the effects of reunification on labor market participation have focused on the West (e.g., Aisenbrey et al. 2009; Biemann, Fasang, and Grunow 2011; Brückner and Mayer 2005; Gundert and Mayer 2012; Manzoni, Härkönen, and Mayer 2014; Mayer, Grunow, and Nitsche 2010), examined differences before or after reunification in one of the two subsocieties (e.g., Diewald and Mach 2006; Diewald, Solga, and Goedicke 2006; Solga and Diewald 2001), or compared the East with the West only before or only after reunification (e.g., Diewald 2006; Diewald et al. 2006). A few studies made comparisons across contexts and periods simultaneously but relied on data for relatively old cohorts (i.e., up to 1970) (Mayer, Diewald, and Solga 1999; Rosenfeld, Trappe, and Gornick 2004; Simonson et al. 2011). Therefore, we know little about the labor market experiences among younger cohorts following reunification. This research gap originates both in a lack of appropriate methodology for studying the impact of macrostructural change on longitudinal life-course and the lack of data covering a sufficiently long time span.

In this context, we address two recurrent core research questions in life-course research: (1) how different life domains—namely, employment and family—are interrelated (Elder 1974; Elder et al. 2003) and (2) how macrostructural changes—such as the German reunification—shape individual life-course trajectories. In both cases, examining the relationships between time-varying covariates and trajectories is crucial.

Our analysis is based on the retrospective data of the Starting Cohort Six of the National Educational Panel Study (NEPS) (Blossfeld, Rossbach, and von Maurice 2011) for women born between 1944 and 1990 in East and West Germany ($N = 731$ and $N = 3,406$, respectively). We constructed individual employment trajectories from age 15 years to the maximum age of 40 years by coding each month according to one of three states: in education, employed, or out of employment (OE). For simplicity, in this illustrative application of the SAMM procedure, we did not distinguish between different types of education or different reasons for being OE, which, however, would be technically feasible.

3. EXISTING METHODOLOGICAL APPROACHES

We now present sequence analysis and event history analysis and highlight their complementary strengths and limitations from both methodological and conceptual viewpoints.

3.1. *Sequence Analysis*

Sequence analysis (SA) provides a holistic view of processes described as sequences—that is, successions of categorical states (Abbott 1995). From a more technical viewpoint, SA relies on a distance measure between sequences (or trajectories) of states, which allows for their comparison (Abbott and Forrest 1986; Elzinga 2005; Müller et al. 2008). Several distance measures are available, and choosing one should be based on their sensitivity in accounting for differences in timing, duration, or sequencing (for a review of distance measures, see Aisenbrey and Fasang 2010; Robette and Bry 2012; Studer and Ritschard 2016). The distances can be further analyzed by using discrepancy analysis (Struffolino, Studer, and Fasang 2016; Studer et al. 2011), multidimensional scaling (Piccarreta and Lior 2010), and cluster analysis to group similar trajectories (for a review of the available clustering algorithms in SA, see Studer 2013). The outcome of this last

procedure is a typology. The types are then interpreted as describing the main ideal-typical processes or trajectories. The remaining variation in the sequences within each type is usually ignored, assuming that a description of the social world requires a certain degree of simplification and that deviations reflect different realizations of the same underlying process (Studer 2013).

Short-term changes¹ and long-term dynamics are simultaneously considered. These features are in line with the life-course paradigm, which stresses the importance of studying the unfolding of trajectories understood as sequences of roles and social statuses (Elder et al. 2003). Therefore, SA is one of the most promising methods to study life-course research questions (e.g., Brzinsky-Fay 2007, 2010, 2014; Liefbroer and Toulemon 2010; Mayer 2009; Shanahan 2000).

Within the SA framework, change is operationalized as lasting over a period of time rather than as instantaneous. As Shanahan (2000) points out, important transitions can occur over several months or years and are usually less well defined than the study of a single event would suggest. He argues that important transitions may result from a succession of events. Abbott (2009) puts forward similar arguments when discussing the notion of “turning points” within processes. Finally, Brzinsky-Fay (2014:218) states that “measuring transitions means capturing a process with a specific time dimension.”

Brzinsky-Fay (2007) advocates for considering longer periods of time when studying employment trajectories as labor market integration or exclusion are processes that last over extended periods (see also Brzinsky-Fay 2010). Moreover, the labor market entry process might have become more complex in increasingly volatile labor markets with numerous internships, temporary jobs, and unemployment spells (Brzinsky-Fay 2007).

Sequence analysis additionally allows us to consider interdependencies between states in a trajectory in terms of the *duration*, *timing*, and *sequencing* of states (Abbott and Forrest 1986; Studer and Ritschard 2016). These dynamics illuminate the internal logic of trajectories by highlighting the necessary or avoidable steps, important turning points, and typical pattern of changes. The *sequencing* situates single events within longer trajectories. This is important because the social meaning and/or consequences of an event may depend on both previous and later events (Elder et al. 2003). For instance, Brzinsky-Fay (2010) points out that the meaning of part-time employment depends on both previous

and later states in employment trajectories. Part-time employment can signify either an employment entry or an exit process. Similarly, the social meaning of unemployment, for example, depends on its *duration* and *timing* with regard to other employment or nonemployment states (Brzinsky-Fay 2014). The same applies to other life domains—for instance, the *duration* between marriage and first childbirth.

However, SA also has several limitations. First, since trajectories are analyzed as a whole, when treating them as dependent variables, we can examine their relationships only with constant attributes or covariates measured before the starting point of the trajectory. Including covariates measured later—such as those in the middle or at the end of the trajectories—leads to conceptual issues related to the well-known problem of anticipatory analysis (Hoem and Kreyenfeld 2006). Indeed, disturbing the temporal order of events leads to accounting for the trajectories, or at least part of them, by a future measurement. The inclusion of censored observations is problematic even when using normalized distances or by adding a missing value state at the end of the sequences. Indeed, after clustering, the resulting typology is often based on the length of the sequences (i.e., observation time), and it rarely has substantive meaning (Elzinga and Studer forthcoming). Because the processes being compared by the distance measure need to be fully observed, most of the studies analyze only complete trajectories to preserve the holistic perspective. Therefore, the sample size is reduced and one usually excludes the observations for younger cohorts of individuals because their trajectories are only partially observed.

To analyze changes in the employment trajectories during German reunification, a conventional application of SA would select a set of fully observed trajectories and build a typology of the sequences. The link with the reunification would then be analyzed by looking at the frequency of each type of sequence in each birth cohort. However, this would offer a very crude estimation because most trajectories would include time points both before and after reunification. Moreover, we would exclude censored observations (i.e., from women born after 1970 whose employment trajectories are not observable from ages 15 to 40). Therefore, none of the trajectories would have started after reunification.

3.2. *Multistate Models*

Event history analysis (EHA) is another framework widely used to study processes and transitions between states. It includes a number of

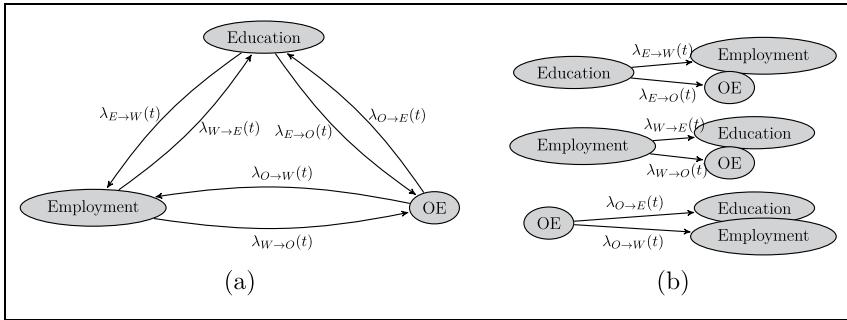


Figure 1. (a) The multistate model and (b) estimation strategy.

methods for estimating duration between two events—such as starting and ending an employment spell—or in a more or less similar way, the hazard of experiencing the second event after the first one.

Within this framework, multistate models (MMs) analyze state sequences by focusing on the hazard rate of observing transitions between states and, implicitly, the time spent in each state (Andersen and Keiding 2002; de Wreede, Fiocco, and Putter 2011; Putter, Fiocco, and Geskus 2007; Steele, Goldstein, and Browne 2004; Therneau and Grambsch 2000). Figure 1(a) visualizes the MM for our illustrative application and shows all possible transitions between the states used to define individual sequences: education, employment, and OE.²

MMs estimate the risk of experiencing a transition $\lambda_{a \rightarrow b}(t)$ over time t between two states a and b . The risks (and the effect of the explanatory factors) are estimated using the strategy displayed in Figure 1(b). We first consider a given state—say, “Education”—and then estimate the risks or chances of the transition to another state (i.e., “Employment” or “OE”). The two transitions, “Education \rightarrow Employment” and “Education \rightarrow OE,” can be seen as competing risks because once one of the two has occurred, individuals are no longer at risk of experiencing the other one. Then, another state is considered—say, “Employment”—and the risks associated with the transitions “Employment \rightarrow Education” and “Employment \rightarrow OE” are estimated. The procedure is repeated for all possible states.

MMs and the EHA framework have several advantages. First, MMs allow for the simultaneous analysis of transitions between several states and the time spent in each spell, which are crucial features of the dynamics of trajectories. Second, censored observations (i.e.,

individuals whose trajectories are not fully observed) can be included in the analysis. Finally, they allow us to measure the influence of possibly time-varying explanatory factors on the occurrence of a given event (Allison 1984; Blossfeld and Rohwer 2002; Courgeau and Lelièvre 1993; Hosmer and Lemeshow 1999; Yamaguchi 1991).

Some limitations of MMs have to be acknowledged. MMs conceive transitions as instantaneous—that is, they occur at one specific time point. As stated earlier, conceptualizing and analyzing changes and transitions over longer time spans is important. Moreover, by focusing on transitions rather than on the longitudinal sequencing of states, MMs fail to take a global view of the unfolding of trajectories and the interdependencies among states over time.

By using an MM for our illustrative application, we would be able to estimate how the transitions between two states is correlated with German reunification. Furthermore, the approach would allow us to include censored observations in the analysis. However, we would lose sight of the whole trajectory and not make a distinction, for instance, between working summer breaks and transitioning from education to stable employment, which is a crucial substantive difference.

4. SEQUENCE ANALYSIS MULTISTATE-MODEL PROCEDURE

So far, none of the available approaches fully addresses the methodological challenge of estimating the effect of time-varying covariates on trajectories. We propose the sequence analysis multistate-model (SAMM) procedure, which combines these two approaches in a step-wise analytical strategy. The SAMM procedure preserves the advantages of both approaches: It conceives change as lasting over a medium-term period while allowing us to study how time-varying explanatory covariates shape trajectories.

The SAMM procedure consists of two steps. First, we use an adapted form of SA to study the typical sequencing and the duration between the main events marking the trajectories over a medium-term period. Then, we use an MM to estimate the chances (or risks) of following each kind of typical sequence, as identified in the first step. We detail each of these steps on a general level, highlighting the necessary choices and available options. The illustrative application of the method follows in the next section.

A script distributed as a supplemental online appendix provides a step-by-step guide on how to implement the SAMM procedure in R.

4.1. Step 1: Identifying Typical Subsequences of Change

In the first step of the SAMM procedure, we identify subsequences that describe what follows over the specified period after each transition between two states. We then build a typology of these subsequences to identify ideal-typical sequences of changes along the trajectories.

4.1.1. Extracting Subsequences. To extract *subsequences of consecutive states*, we first set the time span, denoted by ℓ , over which the subsequences are to be analyzed. Then, for each transition between two states starting at time t , we extract a *subsequence of consecutive states* comprising the states from time t to $(t + \ell - 1)$ in the original sequence (i.e., we extract the subsequence that starts with a transition and lasts for ℓ time units). By doing so, our subsequences describe the transitions between two states as well as what follows over a period of ℓ time units. Since this is done for each observed transition in a sequence, there may be several subsequences for the same individual.³

We extract only the subsequences following a transition that occurred before $L - \ell$ time units, where L is the total length of the sequence. None of the subsequences that start after this time can be fully observed; this implies that our censoring time limit equals $L - \ell$, not L .

Figure 2 provides three examples of the extraction procedure. Sequence 1 represents a woman's employment trajectory that starts with an education spell. After 46 months, she experiences a transition from education to employment. In this example, we set $\ell = 60$ months. We therefore extracted the subsequence starting at position 46 (the last month spent in education) and lasting for 60 months (5 years) spent in employment. This subsequence is framed in a rectangle with solid lines in Figure 2. The employment spell that starts after education lasts for 193 months; this woman then experiences a transition from employment to OE. We thus extracted a second subsequence starting at position 238 and lasting for 60 months (i.e., the same duration as before). Finally, she experiences one last transition at time 268, but we cannot extract a subsequence of 60 months; we therefore discard this subsequence. Indeed, transitions occurring after 240 time units (i.e., the length of the sequence $L = 300$, minus $\ell = 60$) are not included in the analysis.



Figure 2. Three examples of the subsequence-extraction procedure. Extracted subsequences are marked with solid or dashed rectangular borders. The censored time is represented by the vertical black bar.

In some applications, as for our illustrative example, many women experienced several transitions between states. In this case, the extracted subsequences potentially overlap, as shown for sequence 2 in Figure 2. Here, the first extracted subsequence embeds the next four transitions. Finally, the length of sequence 3 equals $L=164$ time units because it refers to an individual who has not reached the age of 40 years yet but whose subsequences (starting with transitions occurring before $164 - 60 = 104$ time units) will nevertheless be included in the sample.

If $\ell=2$, then our subsequences would be of length 2 and they would therefore coincide with the instantaneous transitions. As ℓ increases, we consider longer subsequences, thus describing potentially more complex interdependencies between a first transition and the states that follow. We could therefore analyze medium- or long-term effects and interdependencies. In an extreme case, when the lengths of the subsequences approaches those of the full sequences, we would almost be in the usual SA framework, but our subsequences would still be aligned at the first transition. If ℓ is too high, the dynamics of many trajectories may be disregarded because only a few subsequences that follow a transition will likely be fully observed. In contrast, as ℓ decreases, fewer observations are excluded, but only shorter-term dynamics are analyzed. The choice of ℓ should be based on several trials with different settings and should ultimately be based on substantive arguments, the research question, and data availability.

There are several substantive reasons for studying subsequences. First, the subsequences following a transition may be known (at least to some extent) by the actors. For instance, students may plan a short-term job during the summer break before restarting their studies in autumn. The same applies if a woman accepts a new job even if (or because) she knows that it is only short term. These subsequences may therefore convey meaningful information even from the viewpoint of the actors.

Second, the subsequence that follows a prior transition may be a consequence thereof. That is, the first transition may be the stepping-stone for a more profound transition lasting over a certain period. As we already argued in Section 3.1, transitions are not necessarily instantaneous and may last for a period of time (Abbott 2009).⁴ In breaking down a change into single transitions, we may fail to describe the underlying dynamics of the trajectory (Shanahan 2000). By using subsequences, we implicitly consider the dynamics of change over a medium-term period—something that is more substantively meaningful in many applications.

4.1.2. Subsequence Clustering. Once all subsequences have been extracted, we cluster them by using SA to build a typology. The obtained ideal-typical subsequences summarized in clusters can be interpreted as typical changes along the trajectories that follow a transition between states. This step requires choosing a distance measure for comparing the sequences and a clustering algorithm.

According to Studer and Ritschard (2016), the choice of a distance measure is a substantive one, which should be based on the performance of a distance in accounting for three dimensions: timing, duration (or spacing), and sequencing. Note that analyzing time since first transition is different from analyzing time understood as age, for instance. So if there is a substantive reason to focus more on timing or sequencing, the distance measure should reflect this interest.⁵

The standard sequence clustering procedure (Studer 2013) needs to be adjusted to generate meaningful distinctions for trajectories following a given spell. In the following MM analysis, we analyze the risk of following each type of subsequences, *departing from a given state*. A separate cluster analysis is conducted for each pool of subsequences that start with the same state (i.e., one of education, employment, or OE for our example application). In all studies that use a SA typology for subsequent analysis (either as an independent or dependent variable), the

cluster quality should be sufficiently high to guarantee high within-cluster similarity (Studer 2013). Furthermore, we should ensure that within cluster heterogeneity is not linked to key factors such as age at the start of the subsequence. In our specific case, we can expect a higher cluster quality since the subsequences are shorter than the whole sequence and probably show lower variation. For this reason, they are probably less complex to cluster.

4.2. Step 2: Estimating the Relationships between Trajectories and Time-varying Covariates

After identifying the typical subsequence of changes, we use MMs to estimate the relationships between covariates and the clusters of typical subsequences. Using this approach, we assess how the covariates are associated with the hazard rate of following each type of subsequence cluster while departing from a given state. Individuals with censored observations and those who did not experience any subsequences (i.e., those in a stable sequence without any transition during the entire observation period) are included in the analysis in the at-risk population.

The competing risks are not transitions between states but rather transitions from one state to a sequence of states over a medium-term period (here five-year subsequences). To this purpose, MMs can be adapted to study competing risks, when the different states can be interpreted as different starting conditions for the competing events (Steele et al. 2004).

4.2.1. Estimation Method. Any estimation method suitable for competing risks (see e.g., Andersen and Keiding 2002) can be used for MMs (for more information, see the excellent tutorial by Putter et al. 2007). We propose to consider two factors when choosing among these methods: the underlying measurement of time and the possible recurrence of spells for the same individual trajectory. First, if the *true* durations to be estimated are continuous, we then recommend using Cox models, which are relatively simple and widely available. When the time dimension is measured on a discrete time scale, the “Efron” method should be used to appropriately handle ties in durations (for a review of ties handling methods, see Hertz-Picciotto and Rockhill 1997). On the other side, when the durations to be estimated are discrete or when the durations are only measured with a very crude

approximation such as years, then logistic or multinomial discrete-time models should be preferred (Allison 1982).

Second, if several spells in the same state can be observed in the same sequence, then a model including a frailty term or a random intercept should be preferred. These models relax the assumption of independence of spells belonging to the same individual (Blossfeld and Hamerle 1990; Galler and Poetter 1990; Mayer and Tuma 1990; Wu 2003). Models with frailty terms also provide more accurate estimates of individual-level covariates (Bijwaard 2014) and consider unobserved heterogeneity.⁶ One can use a Cox proportional hazard model with frailties or random intercept (Lunn and McNeil 1995; Putter et al. 2007; Therneau and Grambsch 2000) or a multinomial or logistic discrete-time models with random intercept (Steele et al. 2004). These models take into account multiple observations (in this case subsequences) nested within individuals. On the other side, if each spell can occur only once in each sequence, then the usual Cox model or logistic model can be used.

When using a Cox model or a logistic discrete-time model (with or without a random intercept), we usually estimate a separate model for each competing risk (i.e., typical subsequences of changes here). In this setting, we estimate the chances to experience a specific typical subsequence of changes instead of (1) any other or (2) remaining in the spell. This is achieved by considering a new data set where the competing risks are recoded as censored observations. This strategy assumes that the effect of covariates is transition-dependent and cause-specific hazard rates are nonproportional (with a separate risk curve being used for each transition).

5. ILLUSTRATIVE APPLICATION OF THE SAMM PROCEDURE

We now present an application of the SAMM procedure to assess (1) how time-varying statuses in the family domain are associated with women's employment trajectories in East and West Germany and (2) the association between the reunification and women's employment trajectories.

5.1. Step 1: Subsequence Extraction and Clustering

For the first step of the SAMM procedure, we extracted subsequences from the trajectories and clustered them. Here, we chose a subsequence length $\ell = 60$ months: A five-year period provides a view of medium-

term dynamics within the trajectories, and it is a common time span in the analysis of labor market entry trajectories (Brzinsky-Fay 2014). In online Appendix B, we provide robustness checks testing for different ℓ values. We extracted 14,622 subsequences from 4,137 employment trajectories. This gives an average of 3.5 subsequences per trajectory.

We then clustered the extracted subsequences by using SA. We chose the optimal matching distance measure with constant costs,⁷ which is sensitive to duration—a key aspect of employment trajectories—while still being sensitive to timing and sequencing (Studer and Ritschard 2016). We used partitioning around the medoid clustering method as it has the advantage of minimizing a global criterion measuring residual variation (Kaufman and Rousseeuw 1990). The number of clusters was chosen to maximize the average silhouette width (ASW). We obtained good cluster quality according to the thresholds given in Kaufman and Rousseeuw (1990), with three groups for each pool of subsequences starting with education (ASW = .65), OE (ASW = .64), or employment (ASW = .62). The relative frequency plots presented in Figure A1 of online Appendix A further confirm this high within-cluster homogeneity. We also made sure that the within-cluster heterogeneity is not linked to age using index plots as presented in Figure A3 of online Appendix A. Even if we keep three clusters after each ending spell, the choice of the same number of groups for each set of subsequences is not compulsory. For example, we could very well have chosen only one type of transition for leaving an education spell (regardless of the destination) and three for the others depending on the ASW values.

Figure 3 shows state distribution plots of the subsequence clusters for each starting state. The clusters are labeled according to the medoid (i.e., the subsequence with the lowest average distance to all others in the cluster). The percentages in Figures 3(a), 3(b), and 3(c) refer to the share of extracted subsequences starting in education, out of employment, and employment, respectively, and not to individual sequences. For instance, the percentages in Figure 3(a) are calculated as the number of subsequences in each cluster divided by the number of subsequences that start with education.

Figure 3(a) presents the clusters related to education. First, for 10 percent of the subsequences, education is followed by being OE either immediately or within the next five years (left panel). Second, almost 25 percent of the transitions out of education are only temporary, with a quick return usually after two months (middle panel). The brief

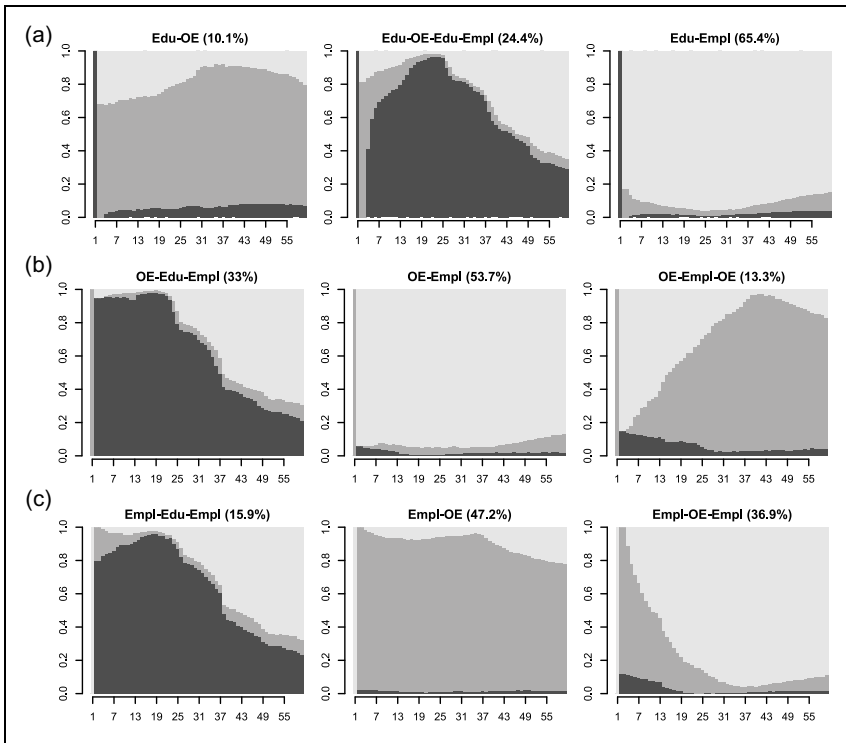


Figure 3. Clusters of subsequences: distribution of the states (y-axis) at each month following a transition (x-axis). Percentages for each plot represent the proportion of subsequences in each cluster departing from a specific state.
Source: NEPS data.

interruptions in continuing education consist of phases both in and out of employment. Finally, almost two-thirds of subsequences that start with education are swiftly followed by employment that lasts for the next five years (right panel). Some women in this group experience a short OE spell before starting employment.

Figure 3(b) presents the clusters following a move away from OE. First, in one-third of the subsequences, women return to education and mostly continue into employment in the next five years (left panel). This pattern likely follows the two-month break of education identified previously. Second, the majority of OE spells end with employment uptake (54 percent, middle panel). Finally, 13 percent of the subsequences show high employment volatility with only brief temporary employment spells before returning to OE.

The clusters after leaving employment are shown in Figure 3(c). First, some women return to education (around 16 percent of the subsequences, left panel). Additional visualization with sequence index plots (available on request from the authors) showed that recurrent transitions between employment and education are common in this group. Second, almost half of the subsequences (47 percent) are shown to be OE over the next five years (middle panel). Finally, 37 percent return to employment after a short period of OE, often within less than one year.

The subsequence clusters identified trajectories of back-and-forth dynamics, such as “Edu–OE–Edu,” “OE–Empl–OE,” or “Empl–OE–Empl.” These kinds of dynamics enable us to distinguish between short summer jobs and more stable transitions into employment and highlight the volatility of employment trajectories with recurrent moves of “OE–Empl–OE” or “Empl–OE–Empl.” In addition, the clusters summarize not only instantaneous transitions but also “medium-term transitions” between states within a five-year period. In contrast to direct instantaneous transitions (i.e., between two states, without an intermediate state in between), we define *medium-term transitions* as subsequences in which the transition to another state occurs after a longer period of time. The clusters of “Edu–Empl,” “Edu–OE,” “OE–Empl,” and “Empl–OE” exemplify medium-term transitions. Table 1 highlights the difference between instantaneous and medium-term transitions in a contingency table. More than 50 percent of the direct transitions “Edu–OE” in fact are part of a subsequence of “Edu–OE–Edu–Empl.” Therefore, direct “Edu–OE” transitions are difficult to interpret out of the context of the longer-term trajectory. The SAMM procedure allows us to clearly identify these medium-term transitions and distinguish them from faster back-and-forth movements, which is not possible in the conventional MM.

5.2. Step 2: Multistate Models

In the second step of the SAMM procedure, we estimate the effect of explanatory factors on the chances to follow each type of subsequences using an MM. Figure 4 presents the MM for our example: Each arrow is a hazard rate to be estimated. We therefore need to estimate nine hazard functions (i.e., one for each pair of ending spell and typical subsequence cluster).

The underlying durations to be estimated here are continuous, and we might observe several spells in the same state for each individual. We

Table 1. Percentage of Direct Transitions Classified in Each Subsequence Cluster

Clustering	From Education to		From OE to		From Employment to	
	OE	Empl	Edu	Empl	Edu	OE
Edu–OE	18.02	5.23				
Edu–OE–Edu–Empl	52.07	7.21				
Edu–Empl	29.92	87.56				
OE–Edu–Empl			86.14	2.78		
OE–Empl			8.49	79.47		
OE–Empl–OE			5.37	17.75		
Empl–Edu–Empl					69.45	3.93
Empl–OE					6.07	56.35
Empl–OE–Empl					24.48	39.72
Total	100.00	100.00	100.00	100.00	100.00	100.00

Note: EDU = education; EMPL = employed; OE = out of employment.
Source: NEPS data.

therefore used a Cox model with random effects to estimate the MM. The following covariates were included in the MM.

The relationship between women’s employment trajectories and family status is estimated by including two covariates: being in a union and having at least one child.⁸ Family status is expected to be related to transitions in and out of employment. A covariate for East and West and two separate dummies for reunification in the East and West were included as proxies of macrolevel change.

We additionally controlled for the period, measured in months and coded as a continuous variable. In fact, several developments besides reunification—such as cultural change and educational expansion—might have affected employment trajectories within our observational window. By omitting the period covariate, we would have attributed all changes in employment trajectories to the reunification of the two German subsocieties.⁹

We also included several control variables. First, age was added as a three-degree polynomial transformation to ensure that the other effects are not related to age.¹⁰ Second, we included parental education, measured as the highest number of years of education between the parents. Finally, we introduced dummies for calendar months because education-related transitions (starting or ending) typically occur between June and September and to control for yearly economic cycles (e.g., those related

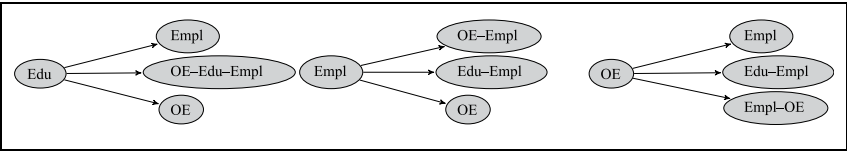


Figure 4. Estimation strategy for the multistate model combined with sequence analysis.

to unemployment). Moreover, given the retrospective nature of our data, recall bias could result in a higher number of transitions during some key months, such as trimesters, semesters, or the beginning of the year. To facilitate interpretation, we standardized all continuous covariates; their coefficients are therefore unit-free.

5.2.1. Interpretation. Table 2 shows the SAMM procedure’s estimates for women’s employment trajectories in East and West Germany. The estimates refer to the association between (time-varying) covariates and the likelihood of ending a spell in one of the ideal-typical subsequence clusters. Overall findings both on the association between family life-courses and the German reunification on women’s employment trajectories are largely in line with expectations.

Having a child reduces a woman’s likelihood to enter any of the ideal-typical subsequences following OE. Women therefore tend to remain out of employment longer after having a child. In addition, they have a higher likelihood to exit from either education or employment into long-term OE after childbirth. Being in a cohabiting or married union is associated with shorter education spells for women that either lead to employment or withdrawal from the labor market. Whether employment uptake or withdrawal occurs after entering a union likely depends on the partner’s employment status and the couples’ gender-specific division of labor.

Concerning our first research question on the impact of family life-courses, the SAMM procedure uniquely allowed us to distinguish the differential impact of having a partner on women’s periods out of employment: Although women in a union are more likely to experience long-term periods out of employment, short-term transitory periods out of employment within overall more dynamic trajectories are more frequent for women who are not in a union.

Concerning reunification, we find stronger effects for East German women’s employment than for their West German counterparts. This is

Table 2. Sequence Analysis Multistate Model (SAMM) Procedure for Women's Employment Trajectories in East and West Germany

	From Education to			From OE to			From Employment to		
	OE	OE-Edu-Empl	Empl	Edu-Empl	Empl	Empl-OE	Edu-Empl	OE	OE-Empl
East	-.54** (.21)	-.22* (.10)	.16* (.08)	.41*** (.12)	1.18*** (.12)	-.09 (.26)	-.15 (.17)	-2.02*** (.17)	-.11 (.11)
West: Reunif	.13 (.19)	-.07 (.11)	.38*** (.08)	-.13 (.12)	.11 (.09)	-.16 (.18)	.12 (.16)	.01 (.11)	-.14 (.11)
East: Reunif	.77* (.33)	-.12 (.19)	.36* (.15)	-.49* (.20)	-.50*** (.15)	.89** (.30)	.34 (.26)	1.72*** (.20)	.35* (.15)
Union	1.33*** (.14)	-.71*** (.16)	.30*** (.06)	-.1.14*** (.12)	-.38*** (.08)	.14 (.15)	-.63*** (.12)	1.3*** (.09)	.04 (.08)
Child	1.16*** (.19)	.04 (.30)	-.21† (.11)	-.2.18*** (.17)	-.1.57*** (.08)	-.1.09*** (.14)	-.35* (.17)	.78*** (.08)	.15† (.08)
Period	-.25** (.09)	.16*** (.04)	-.48*** (.04)	.27*** (.05)	.01 (.05)	.11 (.09)	.32*** (.08)	-.12† (.06)	.09 (.06)
Age	-.03 (.20)	-.97*** (.20)	.28** (.09)	-.1.36*** (.14)	.30*** (.07)	-.19 (.13)	-.1.48*** (.17)	-.32*** (.09)	-.10 (.09)
Age ²	-.20* (.09)	.40** (.15)	-.56*** (.04)	.37*** (.07)	-.37*** (.04)	-.07 (.07)	.01 (.09)	-.30*** (.07)	-.22*** (.07)
Age ³	.20* (.10)	.14 (.11)	.43*** (.04)	.10 (.07)	.16*** (.04)	.05 (.07)	.08 (.09)	-.07 (.06)	.01 (.06)
Omitted output									
Maximum level of education of the parents									
Month in the year dummies									
LogLik (NULL)	-4,213.03	-11,275.72	-28,453.23	-11,227.59	-16,470.47	-3,878.55	-4,912.44	-12,989.74	-10,565.26
LogLik	-3,699.03	-10,185.50	-26,688.88	-9,104.69	-15,888.96	-3,718.34	-4,532.83	-12,217.23	-10,389.36
AIC	984.00	2,136.45	3,484.70	4,201.79	1,119.02	276.42	715.22	1,501.01	307.80
Number of events	532	1,673	3,777	1,570	2,141	469	705	1,571	1,259
Number of observations	263,292	263,292	263,292	141,764	141,764	141,764	458,124	458,124	458,124
Frailty (S. D.)	1.66	.77	1.16	.73	1.05	.85	1.11	1.45	.77

Note: EDU = education; EMPL = employed; OE = out of employment; AIC = Akaike Information Criterion.

Source: NEPS data.

† $p < .10$. * $p < .05$. ** $p < .01$. *** $p < .001$.

in line with the profound shift from the communist GDR to the capitalist FRG model in the East while the institutional context in the West remained similar. Findings support both convergence and divergence of women's employment trajectories after the reunification.¹¹

On the one hand, women's employment trajectories in both subsocieties converged on a higher likelihood of extended periods out of the labor force after the reunification, albeit the coefficients are mostly insignificant in the West. Note that periods out of employment cover both unemployment and family-related leaves in our analysis with arguably more unemployment in the East and more family-related interruptions in the West.

On the other hand, we also see persistent differences after reunification with a higher likelihood of extended periods out employment for West German women compared with East German women. This is expressed both in East German women's higher likelihood of entering the straightforward route to employment in the "OE–Empl" (column 6) group and their lower likelihood of exiting employment via the "Empl–OE" (column 9) group compared with the West. Importantly, East German women's employment trajectories have diverged from their West German counterparts with higher employment volatility. For East German women, we find an elevated likelihood of transitioning into frequent back-and-forth movements between employment and nonemployment after reunification, which is not the case in the West. This is visible in the positive reunification effects for the East (East: Reunif) on entering the "OE–Empl–OE" (column 7) and the "Empl–OE–Empl" (column 10) clusters that are insignificant with a negative sign for the West (West: Reunif). Higher employment volatility for East German women after reunification is further substantiated with their lower likelihood to enter more stable employment trajectories, such as the "OE–Empl" (column 6) pattern. Also with regard to our second research question on the German reunification, the SAMM procedure enabled us to identify East German women's higher employment volatility with various back-and-forth movements as one of the most distinctive diverging trends in East and West German women's employment after reunification.

In addition to these main findings, the period and age covariates can also be of interest. The period captures general trends in employment trajectories over time. It shows the most pronounced effects on subsequences related to education. The findings support that both the duration of education and the likelihood of reentering education after employment or

periods out of employment has increased over time. This is visible in the negative period effect for the “Edu–OE” (column 2) and “Edu–Empl” (column 4) clusters as well as the positive period effects for patterns of reentry to education: the “Edu–OE–Edu–Empl” (column 3), “OE–Edu–Empl” (column 5), and “Empl–Edu–Empl” (column 8) clusters. Age is certainly important in many applications. Some ideal-typical patterns might be more likely to occur at younger or older ages. As expected, in our application, the hazard of entering subsequences that mark a return to education decreases with age: “Edu–OE–Edu–Empl” (column 3), “OE–Edu–Empl” (column 5), or “Empl–Edu–Empl” (column 8).

5.2.2. Robustness Check: Varying Subsequence Length. We ran the SAMM procedure with different subsequence lengths ℓ as a robustness check. We set the subsequence length at 12, 24, 36, 48, and 72 months. The results of the subsequences clustering are very stable. We kept three groups after each ending state to have more comparable results and found in all cases very similar patterns of change. However, the frequencies of each type are slightly different. Therefore, when ℓ is small (i.e., 12 or 24), the number of subsequences classified in back-and-forth patterns such as “OE–Empl–OE” is naturally much lower. We then checked whether each of the statements made in Section 5.2.1 would still be supported when changing the length of the subsequences. The full comparison is available in online Appendix B. The results are remarkably stable. We observe the greatest difference in the case showing a subsequence length of 12, which is also the most different on a substantive level. All statements made here can also be made when $\ell = 72$. Only two statements cannot be made when using $\ell = 48$. The period effect on the chance to follow the “Edu–OE” pattern is no longer significant. The same applies to the effect of the reunification in the East on the chance to follow the “OE–Empl–OE” (but it is still significantly more common in EA than in WA).

5.2.3. Comparison with Standard MMs. To highlight the advantages of the proposed methodology, we compare results from the SAMM procedure with those from conventional MMs, in which competing risks are direct transitions between states (Table 3). In this case, MM does not identify an increase in medium-term transitions to OE after education in the East. In the subsequence-based model, we can distinguish between short education interruptions (which do in fact constitute a continuation

of education) and medium-term transitions to OE: In the conventional MM, the two subsequences are aggregated.

Importantly, by using the conventional MM, we are not able to observe any increase in back-and-forth movement between OE and employment in the East after reunification—because, for instance, transitions to short-term or medium-term employment cannot be distinguished. Finally, we can see that the statistical significance of the “East: Reunif” coefficients are somewhat smaller in the conventional model, indicating that the distinction between subsequence types are relevant to detect changes in the degree of volatility of women’s employment trajectories.

6. EXTENSIONS OF THE SAMM PROCEDURE

We discuss here some possible extensions of our proposed framework. The SAMM procedure relies on two kinds of information: a sequence of spells S that defines the spells to be analyzed with the MM and the subsequences or subtrajectory T that follow each of these spells. In our analysis of employment trajectories, we use the same alphabet—namely, the set of possible states, Σ_S and Σ_T —to describe S and T , respectively. However, we could have chosen two different alphabets. Conceptually, Σ_S describes the different starting conditions in the risk of experiencing the competing risks (i.e., the subsequences T that follow). As noted by Steele et al. (2004), the typically high number of hazard functions that need to be estimated is one of the main limitations of MMs; they therefore recommend considering only very broad differences and limit the size of Σ_S . More subtle differences in the starting conditions can be considered by including additional covariates in the spell-specific models.¹² For instance, the employment rate can be added to the model to distinguish between full-time and part-time employment.

The subsequences T define the studied dynamics of the trajectories. In some cases, we may benefit from a more precise description of these dynamics. This can be achieved by considering a more detailed alphabet for T .¹³ For instance, in distinguishing the various reasons for being OE—such as unemployment or parental leaves—we may be able to better describe the dynamics of casual employment. As another example, distinguishing between full-time and part-time employment might better describe how women restart an employment spell after employment interruptions.

Finally, in some applications, the influence of previous paths on subsequence of change is of central interest. In this case, we suggest adding

Table 3. Transition-based Multistate Model for Women's Employment States in East and West Germany

	From education to		From OE to		From employment to	
	OE	Empl	Edu	Empl	Edu	OE
East	-.41*** (.09)	.22** (.07)	.39*** (.11)	.94*** (.10)	-.08 (.15)	-.97*** (.09)
West: Reunif	-.18* (.09)	.43*** (.08)	-.18 (.11)	.12 (.08)	.15 (.15)	-.10 (.08)
East: Reunif	.26† (.16)	.22 (.14)	-.38* (.18)	-.26* (.13)	.35 (.24)	.95*** (.12)
Union	.38*** (.09)	.24*** (.06)	-1.07*** (.11)	-.29*** (.07)	-.82*** (.11)	.76*** (.06)
Child	1.13*** (.13)	-.29* (.11)	-1.95*** (.14)	-1.39*** (.07)	-.33* (.15)	.46*** (.06)
Period	.16*** (.04)	-.47*** (.03)	.25*** (.05)	.04 (.04)	.20** (.07)	.05 (.04)
Omitted output						
	Maximum level of education of the parents					
	Age (Third-degree polynomial)					
	Month in the year dummies					
LogLik (NULL)	-16,778.27	-27,161.32	-12,036.80	-19,525.61	-5,519.54	-22,943.53
LogLik (integrated)	-15,460.10	-25,573.75	-9,864.16	-19,014.75	-5,143.16	-22,384.47
AIC (integrated)	2,592.35	3,131.15	4,301.29	977.72	708.76	1,074.11
Number of events	2,346	3,636	1,641	2,539	759	2,776
Number of observations	263,292	263,292	141,764	141,764	458,124	458,124
Frailty (S.D.)	.90	.89	.56	.74	1.05	.89

Note: EDU = education; EMPL = employed; OE = out of employment; AIC = Akaike Information Criterion.

Source: NEPS data.

† $p < .10$. * $p < .05$. ** $p < .01$. *** $p < .001$.

indicators of previous paths to the MMs. For instance, we can add the time already spent in each state or dummies for the states experienced to the models.

7. CONCLUSIONS

The relationship between time-varying covariates and processes is at the base of a number of theoretical questions in the social sciences. As a combination of SA and an MM, the SAMM procedure allows us to study patterns of change—namely, subsequences following a transition—along processes. Compared with the conventional methodological approaches, the SAMM procedure offers several advantages to the analysis of the relationship between trajectories and time-varying covariates for different units of analysis. Our illustrative study of women's employment trajectories in Germany highlights the SAMM procedure's usefulness by identifying an increase in volatility of employment trajectories in the East after reunification, signified by more frequent medium-term transitions from education to OE, more recurrent back-and-forth dynamics between OE and employment, and more usual long-term OE spells after education while being in a union.

More generally, within the SAMM procedure, transitions, turning points, and changes are conceived as lasting over a certain period; they are not instantaneous events, as is usually assumed in EHA. Furthermore, studying patterns of change makes it possible to uncover potential interdependencies among states and transitions along the trajectories. The SAMM procedure also considers the duration of a spell—an important dimension of life trajectories and processes. More importantly, within this framework, we can estimate the effects of time-varying covariates on the identified patterns of change. This is crucial for our illustrative example so that we can not only study how individual trajectories are linked to macrostructural changes but also address the issue of linked lives or how various life domains are entwined (Elder 1974). Finally, unlike the typical SA, the SAMM procedure can handle censored observations: In our application, this characteristic makes it possible to include younger cohorts whose trajectories are only partially observed.

In this article, we presented an illustrative application of the SAMM procedure within the field of life-course research, but the general framework can be extended to a wide range of fields and disciplines where there is a theoretical interest in studying the complex relationship

between time-varying factors and processes that are coded as a sequence of categorical states.

Acknowledgments

We warmly thank the anonymous reviewers for very helpful comments. We also greatly appreciate the suggestions made by participants in the USP Writing Workshop at WZB Berlin on an earlier version of the manuscript.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This publication benefited from the financial support of the Swiss National Centre of Competence in Research LIVES—Overcoming Vulnerability: Life Course Perspectives (NCCR LIVES), which is financed by the Swiss National Science Foundation (Grant Number 51NF40-160590).

Notes

1. Changes can be transitions between these states as well as the events that mark the trajectories because transitions can be formalized as the simultaneous occurrence of a set of events (Studer et al. 2010).
2. We usually distinguish between transient states, if individuals can leave a state, and absorbing (sometimes called terminal) states, if individuals are not followed after having reached this state. Being deceased is an example of an absorbing state.
3. Our definition of “subsequence” therefore differs from the one proposed by Elzinga (2005), where a subsequence x of a sequence X is defined as a sequence obtained by deleting any number of states in X . In his definition, the states in x are therefore not necessarily consecutive in X .
4. Abbott (2009) rightly notes that turning points are defined as such *ex-post*. They may be defined as sequences of changes that lead to a profound change in the trajectory.
5. Hence, timing would refer to the time that elapses from this first transition, for example.
6. That is, individual-level characteristics—such as employment motivation—which are not included in the model.
7. Optimal matching analysis calculates the distance between two sequences as the cost of turning one sequence into another based on a set of transformation operations (MacIndoe and Abbott 2004).
8. Most women with children are either married or cohabiting, with both effects tending to cumulate when having children. Indeed, these two family dimensions are strongly associated in our data (Cramer’s $v = 0.57$). For this illustrative application, we do not distinguish between marriage and cohabitation.
9. A more complex transformation of the period covariate is not needed here, as the reunification dummies can only estimate a somewhat linear relationship. Moreover, this complex transformation could partially reflect the effect of reunification itself.

10. The effect of age on the hazard rate to experience each subsequence is likely to be nonlinear. We could think, for instance, that the hazard rate of the transition from employment to midterm out of employment shows a peak at some point. If one wants to exclude that the other covariates capture the nonlinear relationship of age (e.g., because the child covariate is strongly linked to age), adding a polynomial transformation of age (or using a piecewise model) is needed. For our illustrative example, we use a three-degree polynomial, which is significant for some patterns.
11. Based on the results provided, after reunification East and West can be compared by contrasting the “West: Reunif” coefficients with the sum of the “East” and “East: Reunif” coefficients. Additional analyses to assess the statistical significance of differences between East and West after the reunification—in which the reference categories are changed—are available on request from the authors.
12. Spell-specific covariates can be included in an MM (Putter, Fiocco, and Geskus 2007).
13. One may even consider a spell-specific definition for \sum_T since we use a different typology for each spell.

References

- Abbott, Andrew. 1995. “Sequence Analysis: New Methods for Old Ideas.” *Annual Review of Sociology* 21:93–113.
- Abbott, Andrew. 2005. “The Historicity of Individuals.” *Social Science History* 29: 1–13.
- Abbott, Andrew. 2009. “A propos du concept de Turning Point.” Pp. 187–211 in *Bifurcations: les sciences sociales face aux ruptures et à l'événement*, edited by M. Bessin, C. Bidart, and M. Grossetti. Paris: La Découverte.
- Abbott, Andrew, and John Forrest. 1986. “Optimal Matching Methods for Historical Sequences.” *Journal of Interdisciplinary History* 16:471–94.
- Aisenbrey, Silke, Marie Evertsson, and Daniela Grunow. 2009. “Is There a Career Penalty for Mothers’ Time Out? A Comparison of Germany, Sweden, and the United States.” *Social Forces* 88:573–606.
- Aisenbrey, S., and A. E. Fasang. 2010. “New Life for Old Ideas: The ‘Second Wave’ of Sequence Analysis Bringing the ‘Course’ Back into the Life Course.” *Sociological Methods and Research* 38:420–62.
- Allison, Paul D. 1982. “Discrete-time Methods for the Analysis of Event Histories.” Pp. 61–98 in *Sociological Methodology*. Vol. 12, edited by S. Lienhardt. San Francisco: Jossey-Bass Publishers.
- Allison, Paul D. 1984. *Event History Analysis: Regression for Longitudinal Event Data*. Newbury Park, CA: Sage.
- Andersen, Per Kragh, and Niels Keiding. 2002. “Multi-state Models for Event History Analysis.” *Statistical Methods in Medical Research* 11:91–115.
- Biemann, T., A. E. Fasang, and D. Grunow. 2011. “Do Economic Globalization and Industry Growth Destabilize Careers? An Analysis of Career Complexity and Career Patterns over Time.” *Organization Studies* 32:1639–63.
- Bijwaard, Govert E. 2014. “Multistate Event History Analysis with Frailty.” *Demographic Research* 30:1591–620.

- Billari, Francesco C. 2005. "Life Course Analysis: Two (Complementary) Cultures? Some Reflections with Examples from the Analysis of the Transition to Adulthood." *Advances in Life Course Research* 10:261–81.
- Blossfeld, Hans-Peter, and Alfred Hamerle. 1990. "Unobserved Heterogeneity in Hazard Rate Models: A Test and an Illustration from a Study of Career Mobility." Pp. 241–52 in *Event History Analysis in Life Course Research*, edited by K. U. Mayer and N. B. Tuma. Madison, WI: University of Wisconsin Press.
- Blossfeld, Hans-Peter, and Götz Rohwer. 2002. *Techniques of Event History Modeling: New Approaches to Causal Analysis*. 2nd ed. Mahwah, NJ: Lawrence Erlbaum.
- Blossfeld, Hans-Peter, Hans-Günther Rossbach, and Jutta von Maurice, eds. 2011. *Education as a Lifelong Process—The German National Educational Panel Study (NEPS)*. Wiesbaden, Germany: Springer.
- Bonin, Holger, and Rob Euwals. 2002. "Participation Behavior of East German Women after German Unification." SSRN Scholarly Paper ID 297087, Social Science Research Network, Rochester, NY.
- Brückner, H. 2004. *Gender Inequality in the Life Course: Social Change and Stability in West Germany 1975–1995*. Somerset, NJ: Transaction Publishers.
- Brückner, Hannah, and Karl Ulrich Mayer. 2005. "De-standardization of the Life Course: What It Might Mean? And If It Means Anything, Whether It Actually Took Place?" *Advances in Life Course Research* 9:27–53.
- Brzinsky-Fay, Christian. 2007. "Lost in Transition? Labour Market Entry Sequences of School Leavers in Europe." *European Sociological Review* 23:409–22.
- Brzinsky-Fay, Christian. 2010. "The Concept of Transitional Labour Markets: A Theoretical and Methodological Inventory." WZB Discussion Paper No. SP I 2010-507, WZB Social Science Center, Berlin.
- Brzinsky-Fay, Christian. 2014. "The Measurement of School-to-work Transitions as Processes." *European Societies* 16:213–32.
- Carroll, Glenn R., Peter Preisendoerfer, Anand Swaminathan, and Gabriele Wiedenmayer. 1993. "Brewery and Brauerei: The Organizational Ecology of Brewing." *Organization Studies* 14:155–88.
- Courgeau, Daniel, and Éva Lelièvre. 1993. *Event History Analysis in Demography*. Oxford, UK: Clarendon Press.
- de Wreede, Liesbeth C., Marta Fiocco, and Hein Putter. 2011. "mstate: An R Package for the Analysis of Competing Risks and Multi-state Models." *Journal of Statistical Software* 38:1–30.
- Diewald, Martin. 2006. "The Quest for a Double Transformation: Trends of Flexibilization in the Labor Market of East and West Germany." Pp. 214–36 in *After the Fall of the Wall. Life Courses in the Transformation of East Germany*, edited by M. Diewald, A. Goedicke, and K. U. Mayer. Stanford, CA: Stanford University Press.
- Diewald, Martin, and Bodgan W. Mach. 2006. "Comparing Paths of Transition: Employment Opportunities and Earnings in East Germany and Poland during the First Ten Years of the Transformation Process." Pp. 237–68 in *After the Fall of the Wall. Life Courses in the Transformation of East Germany*, edited by M. Diewald, A. Goedicke, and K. U. Mayer. Stanford, CA: Stanford University Press.

- Diewald, Martin, Heike Solga, and Anne Goedicke. 2006. "Old Assets, New Liabilities? How Did Individual Characteristics Contribute to Labor Market Success or Failure after 1998?" Pp. 65–88 in *After the Fall of the Wall. Life Courses in the Transformation of East Germany*, edited by M. Diewald, A. Goedicke, and K. U. Mayer. Stanford, CA: Stanford University Press.
- Elder, Glen H. 1974. *Children of the Great Depression: Social Change in Life Experience*. Chicago, IL: University of Chicago Press.
- Elder, Glen H., Monica Kirkpatrick Johnson, and Robert Crosnoe. 2003. "The Emergence and Development of Life Course Theory." Pp. 3–19 in *Handbook of the Life Course*, edited by J. T. Mortimer and M. J. Shanahan. New York: Springer.
- Elzinga, C. H. 2005. "Combinatorial Representations of Token Sequences." *Journal of Classification* 22:87–118.
- Elzinga, Cees, and Matthias Studer. Forthcoming. "Normalization of Distance and Similarity in Sequence Analysis." *Sociological Methods and Research*.
- Fasang, Anette Eva. 2014. "New Perspectives on Family Formation: What Can We Learn from Sequence Analysis?" Pp. 107–28 in *Advances in Sequence Analysis: Theory, Method, Applications*, edited by P. Blanchard, F. Bühlmann, and J.-A. Gauthier. New York: Springer International Publishing.
- Fasang, Anette Eva, and Tim Futing Liao. 2014. "Visualizing Sequences in the Social Sciences." *Sociological Methods & Research* 43:643–76.
- Frank, David John, Ann Hironaka, and Evan Schofer. 2000. "The Nation-state and the Natural Environment over the Twentieth Century." *American Sociological Review* 65:96–116.
- Franz, Wolfgang, and Viktor Steiner. 2000. "Wages in the East German Transition Process: Facts and Explanations." *German Economic Review* 1:241–69.
- Galler, Heinz P., and Ulrich Poetter. 1990. "Unobserved Heterogeneity in Models of Unemployment." Pp. 226–40 in *Event History Analysis in Life Course Research*, edited by K. U. Mayer, and N. B. Tuma. Madison, WI: University of Wisconsin Press.
- Goldstein, Joshua R., and Michaela Kreyenfeld. 2011. "Has East Germany Overtaken West Germany? Recent Trends in Order-specific Fertility." *Population and Development Review* 37:453–72.
- Gundert, Stefanie, and Karl Ulrich Mayer. 2012. "Gender Segregation in Training and Social Mobility of Women in West Germany." *European Sociological Review* 28: 59–81.
- Hauschild, Christine. 2002. "Die empirische Typisierung von Versichertenbiografien." *Deutsche Rentenversicherung* 57:539–89.
- Hertz-Picciotto, Irva, and Beverly Rockhill. 1997. "Validity and Efficiency of Approximation Methods for Tied Survival Times in Cox Regression." *Biometrics* 53:1151–56.
- Hoem, J. M., and M. Kreyenfeld. 2006. "Anticipatory Analysis and Its Alternatives in Life-course Research. Part 1: The Role of Education in the Study of First Childbearing." *Demographic Research* 15:461–84.
- Hosmer, David W., and Stanley Lemeshow. 1999. *Applied Survival Analysis: Regression Modeling of Time to Event Data*. New York: Wiley.

- Huinink, J., K. U. Mayer, M. Diewald, H. Solga, A. Sørensen, and H. Trappe. 1995. *Kollektiv und Eigensinn. Lebensverläufe in der DDR und danach*. Berlin, Germany: Akademie Verlag.
- Kaufman, L., and P. J. Rousseeuw. 1990. *Finding Groups in Data: An Introduction to Cluster Analysis*. New York: Wiley.
- Klammer, Ute, and Katja Tillmann. 2001. "Erwerbsbiografien als Mosaik-Längsschnittergebnisse zur Zusammensetzung von Erwerbsbiografien und ihren Veränderungen, zum Einkommenserwerb im Lebensverlauf und den Folgen für die Absicherung im Alter." Pp. 141–223 in *Flexicurity: Soziale Sicherung und Flexibilisierung der Arbeits-und Lebensverhältnisse*, Düsseldorf, Germany: WSI und Hans Böckler Stiftung.
- Kreyenfeld, Michaela. 2003. "Crisis or Adaptation—Reconsidered: A Comparison of East and West German Fertility Patterns in the First Six Years after the 'Wende.'" *European Journal of Population/Revue Européenne de Démographie* 19:303–29.
- Kreyenfeld, Michaela. 2004. "Fertility Decisions in the FRG and GDR: An Analysis with Data from the German Fertility and Family Survey." *Demographic Research* 3: 276–318.
- Liefbroer, Aart C., and Laurent Toulemon. 2010. "Demographic Perspectives on the Transition to Adulthood: An Introduction." *Advances in Life Course Research* 15: 53–58.
- Lunn, Mary, and Don McNeil. 1995. "Applying Cox Regression to Competing Risks." *Biometrics* 51:524–32.
- MacIndoe, Heather, and Andrew Abbott. 2004. "Sequence Analysis and Optimal Matching Techniques for Social Science Data." Pp. 387–406 in *Handbook of Data Analysis*, edited by M. Hardy and A. Bryman. London: SAGE Publications.
- Manzoni, Anna, Juho Härkönen, and Karl Ulrich Mayer. 2014. "Moving on? A Growth-curve Analysis of Occupational Attainment and Career Progression Patterns in West Germany." *Social Forces* 92:1285–312.
- Mayer, Karl Ulrich. 2006. "After the Fall of the Wall: Living Through the Post-socialist Transformation in East Germany." Pp. 1–28 in *After the Fall of the Wall: Life Courses in the Transformation of East Germany*, edited by M. Diewald, A. Goedicke, and K. U. Mayer. Stanford, CA: Stanford University Press.
- Mayer, Karl Ulrich. 2009. "New Directions in Life Course Research." *Annual Review of Sociology* 35:413–33.
- Mayer, Karl Ulrich, Martin Diewald, and Heike Solga. 1999. "Transitions to Postcommunism in East Germany: Worklife Mobility of Women and Men Between 1989 and 1993." *Acta Sociologica* 42:35–53.
- Mayer, Karl Ulrich, Daniela Grunow, and Natalie Nitsche. 2010. "Mythos Flexibilisierung?" *KZfSS Kölner Zeitschrift für Soziologie und Sozialpsychologie* 62: 369–402.
- Mayer, Karl Ulrich, and Nancy Brandon Tuma. 1990. "Life Course Research and Event History Analysis: An Overview." Pp. 3–22 in *Event History Analysis in Life Course Research*, edited by K. U. Mayer and N. B. Tuma. Madison, WI: University of Wisconsin Press.

- Minkoff, Debra C. 1995. *Organizing for Equality: The Evolution of Women's and Racial-ethnic Organizations in America, 1955–1985*. Rutgers, NJ: Rutgers University Press.
- Müller, Nicolas S., Alexis Gabadinho, Gilbert Ritschard, and Matthias Studer. 2008. "Extracting Knowledge from Life Courses: Clustering and Visualization." In *Data Warehousing and Knowledge Discovery, 10th International Conference, DAWAK 2008, Turin, Italy, September 2-5*, edited by I.-Y. Song, J. Eder, and T. M. Nguyen. Vol. LNCS 5182, pp. 176–85. Berlin: Springer.
- Olzak, Susan. 1989. "Analysis of Events in the Study of Collective Action." *Annual Review of Sociology* 15:119–41.
- Piccarreta, Raffaella, and Orna Lior. 2010. "Exploring Sequences: A Graphical Tool Based on Multi-dimensional Scaling." *Journal of the Royal Statistical Society, Series A (Statistics in Society)* 173:165–84.
- Putter, Hein, M. Fiocco, and R. B. Geskus. 2007. "Tutorial in Biostatistics: Competing Risks and Multi-state Models." *Statistics in Medicine* 26:2389–430.
- Robette, Nicolas, and Xavier Bry. 2012. "Harpoon or Bait? A Comparison of Various Metrics in Fishing for Sequence Patterns." *Bulletin of Sociological Methodology/Bulletin de Méthodologie Sociologique* 116:5–24.
- Rosenfeld, Rachel A., Heike Trappe, and Janet C. Gornick. 2004. "Gender and Work in Germany: Before and after reunification." *Annual Review of Sociology* 30:103–24.
- Shanahan, Michael J. 2000. "Pathways to Adulthood in Changing Societies: Variability and Mechanisms in Life Course Perspective." *Annual Review of Sociology* 26: 667–92.
- Simonson, J. L., Romeo Gordo, and N. Kelle. 2011. "The Double German Transformation: Changing Male Employment Patterns in East and West Germany." Working Paper 391, German Socio-Economic Panel (SOEP).
- Solga, Heike, and Martin Diewald. 2001. "The East German Labour Market after German Unification: A Study of Structural Change and Occupational Matching." *Work, Employment, and Society* 15:95–126.
- Steele, Fiona, Harvey Goldstein, and William Browne. 2004. "A General Multilevel Multistate Competing Risks Model for Event History Data, with an Application to a Study of Contraceptive Use Dynamics." *Statistical Modelling* 4:145–59.
- Struffolino, Emanuela, Matthias Studer, and Anette Eva Fasang. 2016. "Gender, Education, and Family Life Courses in East and West Germany: Insights from New Sequence Analysis Techniques." *Advances in Life Course Research* 29:66–79.
- Studer, Matthias. 2013. "WeightedCluster Library Manual: A Practical Guide to Creating Typologies of Trajectories in the Social Sciences with R." LIVES Working Papers 24, NCCR LIVES, Switzerland.
- Studer, Matthias, Nicolas S. Müller, Gilbert Ritschard, and Alexis Gabadinho. 2010. "Classer, discriminer et visualiser des séquences d'événements." *Revue des nouvelles technologies de l'information RNTI* E-19:37–48.
- Studer, Matthias, and Gilbert Ritschard. 2016. "What Matters in Differences between Life Trajectories: A Comparative Review of Sequence Dissimilarity Measures." *Journal of the Royal Statistical Society, Series A (Statistics in Society)* 79:481–511.

- Studer, Matthias, Gilbert Ritschard, Alexis Gabadinho, and Nicolas S. Müller. 2011. "Discrepancy Analysis of State Sequences." *Sociological Methods and Research* 40: 471–510.
- Therneau, Terry M., and Patricia M. Grambsch. 2000. *Modeling Survival Data: Extending the Cox Model*. New York: Springer.
- Treas, Judith, and Eric D. Widmer. 2000. "Married Women's Employment over the Life Course: Attitudes in Cross-national Perspective." *Social Forces* 78:1409–36.
- Trischler, F., and E. Kistler. 2010. "Gute Erwerbsbiographien—Arbeitspapier 1: Erwerbsverläufe im Wandel." *Bericht an die Hans-Böckler-Stiftung* (https://www.boeckler.de/pdf_fof/95619.pdf).
- Wu, Lawrence L. 2003. "Event History Models for Life Course Analysis." Pp. 477–502 in *Handbook of the Life Course*, edited by J. Mortimer and M. Shanahan. New York: Springer.
- Yamaguchi, Kazuo. 1991. *Event History Analysis. Applied Social Research Methods Series 28*. Newbury Park, CA: Sage.

Author Biographies

Matthias Studer, PhD in socioeconomics, is a senior researcher at the Swiss National Centre of Competence in Research (NCCR) program "LIVES Overcoming Vulnerability: Life Course Perspectives" and a lecturer at the Faculty of Social Sciences of the University of Geneva. His research interests include quantitative methods for longitudinal data analysis, sequence analysis, gendered career inequalities, and labor market and social policy evaluation. He is one of the TraMineR developers, and he recently published on discrepancy analysis in sociological methods and research and a comparison of sequence analysis distance measures in the *Journal of the Royal Statistical Society (Series A)*.

Emanuela Struffolino is senior researcher at the "Demography and Inequality" research group at the WZB Berlin Social Science Center and associate researcher at Humboldt University of Berlin. She obtained her doctorate in sociology from the University of Milano-Bicocca and has worked as a research fellow at the Swiss NCCR program LIVES Overcoming Vulnerability: Life Course Perspectives. Her research interests include life-course sociology, gender inequalities in the labor market, social stratification, and methods for longitudinal data.

Anette E. Fasang is a professor of microsociology at Humboldt University of Berlin and is head of the "Demography and Inequality" research group at the WZB Berlin Social Science Center. She obtained her doctorate from Jacobs University Bremen and completed postdoctoral research at Yale University and Columbia University. Her research interests include social demography, stratification, life-course sociology, family demography, and methods for longitudinal data analysis.