Sequence Analysis as a Tool for Family Demography

Nicola Barban
Institute of Economic and Social Research, University of Essex
nicola.barban@essex.ac.uk

Maria Sironi
Department of Quantitative Social Science, UCL
m.sironi@ucl.ac.uk

March 2018

1 Family Change Over Time

If we think about a definition for the word *family*, we will immediately realize how difficult this task can prove to be. However, this was not the case six or seven decades ago. As a matter of fact, family structure has changed considerably over time, and family trajectories are much more heterogeneous now than they used to be in the past. Most people in Western societies followed a similar path in the 1950s and 1960s, which was moving out of the parental home to get married and have children (Allan et al., 2001; Furstenberg, 2014). Moreover, these events would happen relatively early on in people's life. For example, in the United States and in many European countries, most people were married and with children in their mid-twenties (Bianchi, 2011; Cherlin, 2010). Since then, many things have changed. The age at leaving home and the age at first marriage have increased in many, if not all, western societies; cohabitation has become a very standard stage in young adults' family formation; divorce rates have rose considerably since the 1970s (even though their growth slowed down and recently even halted in some countries); childbearing has been postponed, with a concomitant decrease in fertility rates; and stepfamilies have become more and more common (Amato et al., 2008; Baizan et al., 2003; Billari, 2004, 2001a; Iacovou, 2010; Schoen et al., 2007; Sobotka, 2008; Stevenson and Wolfers, 2007). All of these trends imply not only changes in *timing* of events, but also changes in the *sequencing* and the *duration* of events, lending to an increase in the complexity of life course trajectories.

Moreover, as suggested in the recent demographic and sociological literature, family changes have been different in different contexts, with varying degrees of postponement and de-standardization in life course trajectories. The variability

in the transition to adulthood and family formation has been defined as 'convergence towards diversity' or 'convergence to divergence' (Fokkema and Liefbroer, 2008; Billari and Liefbroer, 2010; Billari and Wilson, 2001). Why have family trajectories changed so much - towards more heterogeneous and less standardized behaviors - and why did the changes differ across countries? Four potential explanations have previously been proposed for the 'convergence to divergence,' each one contributing differently to the speed and the degree of change.

The first one refers to the Second Demographic Transition - SDT hereafter - as theorized by Lesthaeghe and Van de Kaa (Lesthaeghe, 2010; Van De Kaa, 1987), which describes the process of individualization in demographic behaviors across Europe since the 1960's. The second explanation is related to country-specific institutions, which affect life trajectories through welfare regimes (Esping-Andersen, 2013; Mayer, 2001). Connected to the institutional explanation, the third explanation refers to changes in the socioeconomic context, labor market structure and women's revolution (Esping-Andersen and Billari, 2015; McDonald, 2000; Schoen, 2011). The fourth explanation is connected to long-term cultural differences among countries, such as the strength of intergenerational ties (Reher, 1998).

According to the Second Demographic Transition theory, individualization and secularization are at the basis of new demographic behaviors. Individualization implies greater flexibility in life trajectories and it allows for less normative behaviors, aiming at adult self-realization and self-fulfilment (Lesthaeghe, 2010). Hence, longer periods are spent in states, such as single living, unmarried and childless cohabitation (Billari, 2001a). The preference shift towards more non-material needs, such as freedom of expression, emancipation, and autonomy, was followed or complemented by a change in social norms and values. It was also supported by technological innovations, such as the pill and other forms of efficient contraception. The expression of these new preferences and values resulted in the main features of the SDT: a rise in non-marital cohabitation, which became more widespread and more acceptable as an alternative to marriage; delayed marriage and lower rates of 'ever marrying'; an increase in divorce rates, which led to an increase in re-marriages; delayed and lower fertility, but also a rise in non-marital childbearing; and higher levels of union instability, which has been at the basis of a rise in single-parent families, multipartnered fertility, and step-families. Modernization and secularization processes were obviously instrumental to the ideanational changes described above. Notably, the SDT did not start at the same time everywhere, and also the speed of its diffusion has been different across countries. New demographic behaviors started spreading first in Northern Europe, and then to the rest of Europe and other developed societies, such as the United States (Lesthaeghe, 1995; Lesthaeghe and Neidert, 2006). For this reason, family change has been so heterogeneous across different contexts.

The lack of clear evidence of convergence in family trajectories in industrialized societies may also be related to diverse institutional structures. The most obvious way in which this can affect demographic behaviors is through welfare regimes. For example, generous social benefits in Social Democratic countries

2

allow for an early transition to adulthood and family formation, giving young adults more certainty on their future economic conditions. In contrast, Southern European countries show late exit from the parental home and late marriage and childbearing, given the lower level of benefits and the consequent stronger dependence on the family of origin (Esping-Andersen, 2013; Ferrera, 1996; Mayer, 2001; Trifiletti, 1999).

Moreover, in order to understand the recent changes in family formation, we need to consider the expansion of education that has taken place in the last decades of the 20th century - especially among women -, and the significant changes in the socioeconomic context due to global competitive pressure. An increase in the years of schooling has caused a delayed entry in the labor market and a delay in union formation and childbearing (Blossfeld and Huinink, 1991; Liefbroer and Corijn, 1999). Since the 1970s, earnings inequality and employment instability have grown and job tenure has declined with growing turnover (Sironi and Furstenberg, 2012), which also contributes to a delay and destandardization in union formation (Danziger and Ratner, 2010). Female education and labor force participation rates have increased substantially and, together with the gender equality movement, have led to gender competition among partners. The competition has undermined traditional family behaviors, such as marriage and childbearing, and has driven changes in family behavior. For instance, cohabitation was the opportunity for economically independent women to have a union of equals, and for men to eliminate the constraints and obligations of traditional marriage (Schoen, 2011). These structural shifts have been accompanied by changes in family and gender equality policies, which have had different effects in various settings. In fact, according to McDonald's theory (McDonald, 2000; Mcdonald, 2006), the efficacy of these policies depends on whether institutions have adapted to the new role of women in the society; this implies a further source of heterogeneity in family change across countries.

The fourth class of explanations of family change and its geographical heterogeneity is based on long-term cultural differences across societies. Reher's work shows that in southern Europe there is an emphasis on a vertical relationship between generations and 'strong' family ties. On the contrary, in the north, Germanic tradition and the reformation contributed to the development of a 'weak' family (Reher, 1998). This also implies that external shocks affect 'strong' and 'weak' family systems differently. For example, the sexual revolution (i.e. the advent of female contraception for unmarried women and the legalization of abortion in the 1980s, which postponed family formation) had a negligible impact in Northern Europe, where family ties were weak and children lived by choice out of their parents' home. The same shock, in contrast, had a major impact in Southern Europe, where family ties were strong and children by choice postponed the exit from the parental home (Giuliano, 2007). The cultural inheritance model predicts that convergence will not take place at a national level: factors, which lead to homogenization of behaviors throughout the world, like those implied by modernization and economic globalization, will not have the same impact on societies with different cultural and historical origins, thus leading to diverse and possibly divergent coping behaviors and timing

of transitions.

The above description of family change over time, and the attempt to identify explanations responsible for the change, clearly show how complex family trajectories have become in the past decades. This complexity and the de-standardization of life course pathways call for innovative tools and methodologies to analyze family formation. Many aspects of family trajectories and family change have been analyzed individually, such as the postponement of marriage, the increase in the mean age at first birth, and the diffusion of cohabitation. However, only a few studies have taken into account the interrelation among different events, and the possible influence they have on each other. It is necessary to look at the process of union formation and the subsequent family pathways from a holistic point of view. Therefore, we think that *sequence analysis* is the appropriate tool to analyze family histories, taking into account timing, sequencing, duration of events, and therefore, addressing the complexity that characterizes these demographic processes.

In this chapter, we will review the sequence analysis technique, and we will discuss the way in which sequence analysis has been used so far in family demography. Moreover, we will illustrate the most relevant developments and innovative procedures relative to this technique (e.g., 'multichannel sequence analysis'). Finally, given the heterogeneity of family change in different contexts, in the second part of the chapter, we will use data from the European Social Survey to illustrate an empirical application of sequence analysis, and we will describe family trajectories across 23 European countries.

2 A Formal Representation for Life Course Trajectories

A life course trajectory can be described as the representation, over the course of an individual's age (or any other alternative time reference), of an ordered sequence of life course events. The concept of trajectory derives from the representation of the life course paradigm proposed by Elder (1985), in which trajectories are based on the occurrence of events in multiple life domains. For example, one may want to describe the evolution of residential independence, marital status, and childbearing over an individual's life course.

Trajectories can be envisioned as sequences of transitions that are enacted over time. A life course transition is a discrete life change or event within a trajectory (e.g., from single to married) often accompanied by socially shared ceremonies and rituals, such as a graduation or a wedding ceremony. On the other hand a trajectory is a long-term pathway, with age-graded patterns of development in major social institutions such as the family. A primary objective of life course analysis is to study the entire development of life trajectories for different group of individuals. In particular, researchers are often interested in the *timing* (at what age different life transition happen), the *quantum* (what and how many transitions happen), and the *sequencing* (which transition come

first and which after) of life course events.

A convenient representation of life course trajectories is to depict them as sequences of linked states within a conceptually defined range of behaviors or experiences. In statistical terms, this can be translated into a categorical time series. For each individual i we can then associate a variable s_{it} that indicates her/his life course status at time t. As one can assume that s_{it} takes a finite number of values, trajectories are then represented as categorical time series.

In other terms, trajectories can be represented as strings or sequences of characters, with each character denoting one particular state. The state-space, (i.e the set of possible states from which sequences are constructed) has a finite number of elements and represents all the possible combinations of events that an individual can take in each time period.

For instance, the marital status trajectory of a woman who is single for 4 years since the start of our observation (e.g., age 18), then starts a cohabitation lasting 3 years and then marries and remains married for 7 years can be described as follows:

$$\text{SSSSCCCMMMMMMM}$$

In this case, the state-space has 3 values (S=single; M=married; C=cohabiting) and the trajectory illustrates the marital status from age 18 to age 29.

More formally, we can define the life course trajectory of length T of individual i as the set of realizations of a discrete-time stochastic process $S_t : t \in T$ with state-space $\Sigma = \{\sigma_1, \ldots \sigma_K\}$. The trajectory of the individual i can be then described by the sequence $s_i = \{s_{i1} \ldots s_{iT}\}$.

For practical reasons, a more compact representation of sequences, which we shall use later on, involves counting the repetitions of a state, which in the former example becomes as follows:

$$\text{(S,4)-(C,3)-(M,9)}$$

At the same time, individuals experience events in multiple life domains. A constrain of classic methods for the analysis of life course trajectories is the difficulty to analyze simultaneously different life course domains. Representing life course trajectories as sequences of states can accommodate the occurrence of transitions in multiple domains.

For instance, we can describe childbearing using two states (childless; parent) or residential independence (living with parents; living single or with a partner). Assuming that the person above started living outside the parental home at age 20 and had a child at age 24, we can describe her life course trajectories as follows:

   i. Marital Status   (S,4)-(C,3)-(M,9)
   ii. Independence   (Living with parents,2)-(Living independently,14)
   iii. Childbearing   (Childless,6)-(Parent,10)

Life course domains can be treated separately or jointly, depending on the required analysis. Researchers in family demography are often interested in the interdependence of different life transitions. A common practice is to adapt the state-space to the combination of all possible events. In this example, the combination of the three domains has 3*2*2=12 possible states. However, we can assume that some states are not possible together or are joined with others, e.g. marriage and cohabitation assume that the individual no longer lives with her parents (if this is not the case, it will be treated as it was). Table 1 shows the combination of possible states in a hypothetical family trajectory.

TABLE 1 HERE

The resulting life course trajectory can be then rewritten as follows:

(S,2)-(SL,2)-(C,1)-(CP,2)-(MP,9).

Life course sequences can be described with graphical representation (i.e. chronograms, see Figure 1) or by studying the occurrence, timing and order of particular events (Barban, 2013). Other possibilities for describing trajectories' characteristics include the study of heterogeneity and dynamics in life course states. For instance, Elzinga et al. (2007) proposed to take into account, besides the number of transitions, the duration in different states. They introduced a new index, called turbulence, that is a composite measure of two aspects: variability in the time spent in different states and the number of distinct subsequences that can be extracted from the main sequence. Turbolence can be used to give an overall measure of the grade of disorder of a life trajectory (Elzinga, 2010; Elzinga and Liefbroer, 2007; Widmer and Ritschard, 2009).

3 Measuring Dissimilarity in Life Course Trajectories

Most often sequence analysis is used to quantify distances between categorical time series, i.e. life course trajectories.

Optimal Matching algorithm (OM) is the most used algorithm that has been applied within social sciences. Basically, OM expresses distances between sequences in terms of the minimal amount of effort, measured in terms of edit operations, that is required to change two sequences such that they become identical. The OM dissimilarity measures are derived from the measure originally proposed in the field of information theory and computer science by Vladimir Levenshtein (Levenshtein, 1966) and later adapted to the social sciences (Abbott and Hrycak, 1990; Abbott, 1995).

Three basic operations on sequences are used: $\Omega = \{\iota, \delta, \sigma\}$, where ι denotes *insertion* (one state is inserted into the sequence), δ denotes *deletion* (one state is deleted from the sequence) and σ denotes *substitution* (one state is replaced by another state into the sequence). To each of these elementary operations $\omega_k \in \Omega$, a specific cost can be assigned using a cost function $c(\omega) : \Omega \rightarrow \mathcal{R}^+$.

If K operations must be performed to transform one observed sequence $\mathbf{s}_1$ into another $\mathbf{s}_2$ such that

$$\mathbf{s}_2 = \omega_1 \circ \omega_2 \circ \cdots \circ \omega_K(\mathbf{s}_1) = \omega_{.}(\mathbf{s}_1),$$

then the transformation cost is defined as $\sum_{j=1}^{K} c(\omega_j)$. The distance between two sequences can thus be defined as the minimum cost, independently from the order of the operators, of transforming one sequence into the other one:

$$\mathcal{D}_s(\mathbf{s}_1, \mathbf{s}_2) = \min_{\omega_{.}}\{\sum_{j=1}^{K} c(\omega_j) \text{ s.t. } \mathbf{s}_2 = \omega_{.}(\mathbf{s}_1)\}.$$

The choice of the operations' costs determines the matching procedure and influences the results obtained.

Several other dissimilarity measures have been proposed to compare life course trajectory. An exhaustive review can be found in (Studer and Ritschard, 2016).

Alternative measures and techniques have been developed as a response to sequence analysis' critiques (Wu, 2000; Aisenbrey and Fasang, 2010). A first critic to the use of OM is the role of the sequence operations. Differently from biological sequences from which sequence analysis derives, social sequences are time referenced. Therefore, the operations imply modifications in the time scale. In particular, insertion and deletion operations modify the time scale in order to match identically coded states but occurring at different moments in their respective sequences.

A second limitation, is the choice of costs in the use of OM for social sciences because their arbitrariness and the absence of link to social theory. Critics argue that the resulting distances are meaningless from a sociological point of view (Levine, 2000). In the case in which there is not a clear ranking between the different states, the definition of cost is necessarily arbitrary. A common practice is to set constant costs independent on the states that are substituted or, most commonly, to adopt a data-driven approach,that uses substitution costs inversely proportional to transition frequencies (Piccarreta and Billari, 2007). This cost specification takes into account the occurrence of events, weighting more those transitions that are less frequent.

A third critic to OM is the treatment of missing data and censoring among sequences. Unequal sequence length due to censoring may affect the calculation of distances between sequences. To overcome this limitation, Elzinga (2003) proposed different measures, such as the *Longest Common Subsequence (LCS)* for categorical time series that are valid for sequences of different length and do not require any cost specification.

Sequence analysis and OM are often used in conjunction with cluster analysis or other data reduction techniques to identify patterns in the data and highlight typical life course trajectories (Barban and Billari, 2012).

Cluster analysis is used to reduce the enormous number of possible trajectories to a meaningful and treatable number of typical life course trajectories.

Doing so, each observation (i.e. individual) is assigned to a "typology" based on its distance to all the other sequences. Starting from the dissimilarity matrix, i.e. the set of all pairwise distances calculated with a sequence analysis algorithm, cluster analysis is used to agglomerate iteratively sequences in order to reduce the number of groups. Different cluster algorithms can be used (e.g. single linkage, complete linkage, Ward's method, k-means clustering)[1].

This approach has several advantages. Complexity is reduced and differences in timing, tempo and ordering of events is maintained between clusters. The cluster solution can be then used as a "new" categorical variable, often incorporated in some multivariate regression model. This allows researcher to study the determinants of different life course patterns using clusters as a dependent variable (McVicar and Anyadike-Danes, 2002). Alternatively, this group classification can be used as an independent variable if the researcher is interested in understanding the effect of different life course trajectories (or rather group of trajectories) on successive outcomes (e.g. the effect of different trajectory of family formation on health outcomes (Barban, 2013)).

Clustering sequences has some limitations. First, the variability within clusters is not taken into account, only between cluster differences are analyzed. Second, the number of cluster solutions also implies a certain degree of arbitrariness, although several measures of goodness of fit have been recently proposed (Studer, 2013). Third, time varying covariates cannot be included in a regression model, although a recent development of sequence analysis called "Sequence Analysis Multistate Model Procedure" addresses this problem by combining sequence analysis and an event history model (Studer et al., 2018). Last, regression models based on sequence clustering usually do not address causality, but provide a multivariate description of life course dynamics. To overcome this limitation, a recent article by Barban et al. (2018) proposed a new matching approach based on sequence similarity to investigate the causal effect of the timing of life-course events on subsequent outcomes.

4 Sequence Analysis for Family Demography

Sequence analysis has been vastly adopted in family demography to study complex phenomena and multiple demographic transitions simultaneously (see e.g., Billari (2001a,b); Fasang (2014)).

By focusing on the analysis of entire trajectories rather than the occurrence of single events, sequence analysis overcomes the limitations of event history analysis in studying multiple interrelated events. Different life domains such as family formation, childbearing, and employment are merged together in complex life sequences and analyzed in a holistic manner. Rather than investigating the age at different life course events (e.g. age at marriage and children progression), sequence analysis can be employed to describe the occurrence and

---

[1]Cluster techniques differ mainly by the linkage algorithm used, i.e. the way different observation are merged together. For an exhaustive description of clustering techniques, see Kaufman and Rousseeuw (2009)

the timing of events together with their sequencing (i.e. the order). This allows researchers to identify patterns and create life course typologies. Sequence analysis has been applied to a vast range of life domains, for instance work-family trajectories in the transition to adulthood (Aassve et al., 2007); school to work transitions (Brzinsky-Fay, 2014); family formation among immigrants and second generations (Kleinepier et al., 2015; Ferrari and Pailhé, 2017); childlessness (Jalovaara and Fasang, 2017); gender differences (Jalovaara and Fasang, 2015).

Combining multiple life domains has the drawback of increasing complexity because of the higher number of state-space dimensions. Moreover, not all the combinations of states are always relevant for the analysis. For this reason, a different approach that analyzes separately different domains (e.g. employment history and partnership history) and combines them in a single dissimilarity measure is sometimes used. This approach is called "multichannel sequence analysis" and has often been applied to the study of family trajectories (Gauthier et al., 2010; Pollock, 2007). The advantage of multichannel sequence analysis is that different sets of costs can be specified for each domain, while a single measure of sequence dissimilarity is computed from the observed data.

Sequence analysis has been used to investigate several empirical aspects in the sociology of the family. A common debate in family demography is whether the overall structure of the life course has changed and became "de-standardized," "de-institutionalized," and increasingly "individualized" (Macmillan and Copher, 2005; Shanahan, 2000). It is not clear, however, what are the consequences of a de-standardization of family life course. Sequence analysis, by measuring life course dissimilarities, has been used to investigate the degree of heterogeneity and complexity over different populations and birth cohorts (Bras et al., 2010; Elzinga and Liefbroer, 2007).

A noteworthy application of sequence analysis is the study of life course interactions. A pillar of the life course paradigm is that human lives are embedded in social relationships with kins and friends across the life span (Elder, 1985). How do individuals affect each others' life courses? Several studies have attempted to investigate cross-influence in family events through an event history framework (e.g. Balbo and Barban, 2014; Lyngstad and Prskawetz, 2010). Sequence analysis can be used to understand similarities of specific trajectories, for instance siblings or parents and children trajectories. By using the entire information of the dissimilarity matrix, it is possible to analyze siblings' similarities in family formation (Raab et al., 2014). The same authors implemented sequence analysis, and in particular multichannel sequence analysis, to study the intergenerational transmission of family formation (Fasang and Raab, 2014).

Sequence analysis is thus a powerful tool to investigate various classical topics in family demography. Most of the studies focus on a single country, but recently, sequence analysis has been used for comparative research. Sironi et al. (2015) used multichannel sequence analysis to investigate the role of parental social class in the transition to adulthood in Italy and the United States. Also, sequence analysis has been used to test the second demographic transition theory in post-socialist countries (Romania and the Russian Federation) compared to France (Potârcă et al., 2013). A wider cross-country comparison using sequence

analysis to explore family trajectories has been applied to the Generations and Gender Survey (GGS) and to SHARELIFE, the life course module of SHARE (Schwanitz, 2017; Van Winkle, 2017).

5 An Application of Sequence Analysis to 23 Countries

5.1 Data

To illustrate the use of comparative sequence analysis for family demography, we use the European Social Survey (ESS). The ESS is a multi-country survey that started in 2002, and that counts eight rounds, one every two years (from 2002 to 2016). Its main aim is to outline the attitudes of the different regions towards religion, politics, and moral issues, while also depicting their social habits and how they are changing over time. Since this study was conceived to draw a long-term picture about social and cultural changes, many issues have been addressed through the questionnaire: people's value and ideological orientations, people's cultural/national orientations, and socio-demographic characteristics. These issues are part of the so-called Core Module, which stays unchanged over time. The questionnaire includes also a second part that is dedicated to specific themes, the so-called Rotating Module.

We focus on Round 3 of the ESS collected in 2006, which includes the rotating module 'timing of life'. This module collects detailed information on timing of various life course events, attitudes related to the ideal age, and oldest/youngest age for specific events, and planning for retirement. The third round includes 23 countries (Austria, Belgium, Bulgaria, Switzerland, Cyprus, Germany, Denmark, Estonia, Spain, Finland, France, Great Britain, Hungary, Ireland, Netherlands, Norway, Poland, Portugal, Russia, Sweden, Slovenia, Slovakia, Ukraine) and 43,000 individuals born between 1905 and 1992. Since we want to look at family history between age 15 and age 35, we select only individuals who are 35 years old and older (born before 1973) and who provide information on age at leaving home, age at first non-marital cohabitation, age at first (successive) marriage, and age at first (successive) child. Our sample includes 30,597 individuals, born between 1905 and 1972, and 55.5% of which are women. The number of observations per country ranges from 701 in Cyprus to 2,161 in Denmark.

5.2 Identifying Typical Trajectories of Family Formation

We apply sequence analysis to the respondents of Round 3 of the European Social Survey. We begin by constructing family trajectories based on transitions to partnership and to parenthood. We use the same state-space described in section 3 composed by 7 possible states. Figure 1 shows the chronogram from age 15 to age 35.

FIGURE 1 HERE

10

We then calculated the sequence dissimilarity among all the individuals in the sample. We created a matrix of dimension $30,597 \times 30,597$ containing all the pairwise distance measures based on the Optimal Matching algorithm. Since we do not have *a priori* knowledge of ranking of family states, we adopted inverse transition costs, that assign higher costs to less frequent transitions. The analysis has been performed using the package `TraMineR` in `R` (Gabadinho et al., 2011).

As we are analyzing life course trajectories from age 15 to age 35, the number of possible combinations of sequences in family formation is enormous. It follows that a convenient empirical strategy aims to reduce all the possible trajectories to a more manageable number. We used a cluster analysis to specify five groups of trajectories, representative of the entire set of sequences (men and women together). We did not run gender-specific cluster analysis in order to have a gender-neutral classification that can be used to understand the distribution of men and women in different life course groups in different countries. Specifically we adopted hierarchical clustering and we based our cluster solution on the comparison of different measures of cluster performance provided by the R package `weightedCluster` package described in Studer (2013) (Table 2). Cluster performance is measured by looking at the goodness of classification under different cluster solutions. Silhouette, for instance, is a measure of how sequences are similar to their own group compared to other groups. The silhouette index ranges from -1 to $+1$, where high values indicate that sequences are well matched to those of their own group. We computed measures of cluster performances for different number of groups (from 2 to 10 cluster solutions) and reported the "optimal" number of groups for different indexes. Four out of six indexes indicated that the optimal solution is composed by five groups.

TABLE 2 HERE

A description of the sequences in each group can be found in table 3 and figure 2 . Clusters can also be described using their *medoid* sequences (Aassve et al., 2007). A medoid is the observation with the minimum distance from other individuals in a cluster. The medoid of the first group, for example, is *(S,19)-(M,2)*, indicating that the most central sequence corresponds to an individual who stays single for 19 years (from age 15 to 33), marries at age 34 and stays married for two years. The advantage of using medoid sequences is to define the cluster using a real sequence that best represents the group.

TABLE 3 HERE

The five groups can be described as follows:

1. **Late transition** *(S,19)-(M,2)* This group represents individuals who started family transition very late. They represent 13.21% of the sample. They stay single for the majority of the sequence and eventually experience a transition to union formation. About 43% of them is married and 30% experiences parenthood by age 35.

2. **Marriage and childbearing** *(S,6)-(M,1)-(MP,12)* This is the largest group in the sample, representing approximately 65% of the total sample. It is composed by men and women that follow a more gradual and traditional pattern i.e. Single-Married-Married Parents. The majority of transitions are traditional and cohabitation is generally absent or short.

3. **Cohabitation** *(S,5)-(SL,3)-(C,7)-(CP,6)* This group is compose by men and women who experience long-term cohabitations. Individuals in this group are generally younger, and they spend most of their time (up to age 35) in cohabitation. Childbearing is common (68% of them experience parenthood before age 35). Cohabitation is sometimes transformed into marriage (23% of men and women in this group marry by age 35). Also, cohabitation is often preceded by independent living. Overall, this is the most "modern" group, as described by the SDT theory.

4. **Transition to independence** *(S,4)-(SL,14)-(M,1)-(MP,2)* This group is characterized by individuals who spend most of their time not in partnership but living outside the parental home. They experience an early transition to independence living, but late family formation. About half of this group marries by age 35, but they marry late with a median age at marriage of 34.

5. **Marriage, no children** *(S,6)-(SL,2)-(M,13)* Men and women in this group experience marriage but not childbearing. This group is characterized by early family formation and long periods spent in a marriage without children. About one quarter of individuals in this group experience childbearing towards the end of the observation period. This is the smallest cluster, representing about 5% of the sample.

FIGURE 2 HERE

5.3 Differences Across Countries and Birth Cohorts

The typologies of family trajectories that we constructed can be used to describe family differences across countries and changes over time. As reported in table 4, there are significant differences in family dynamics across European countries. For instance, the proportion of individuals who can be classified in the "late transition" group varies from 6% in Denmark to over 26% in Spain. Bulgaria is the country with the higher proportion of individuals in the "traditional", cluster characterized by marriage and childbearing, while Sweden is the country with the highest proportion of people in the "cohabitation" group (26.6%). Ireland and Switzerland are the countries with the highest proportion (17%) of people in the fourth cluster, "transition to independence". The fifth group, "marriage with no children", is the smaller cluster in size with a proportion of people that varies from 2% in Slovenia to 8% in United Kingdom.

TABLE 4 HERE

How did family formation change over time? To answer this question, we analyzed how the distribution of family trajectories varies across birth cohort. By visually inspecting the distribution in each cluster over time (Figure 3), several trends emerge. First, it is possible to observe an inverse U-shaped trend in "traditional" family formation. This group reached a peak among cohorts born in the 1940s, and then declined substantially among younger cohorts. At the same time, we observe a dramatic increase in cohabitation, with almost no gender differences. A group that shows substantial gender differences is the cluster named "late transition". While for women the proportion declined and remained low, men experienced an increase starting from the cohort born in the 1940s. We do not observe particular trends and variations for the last two groups, at the aggregate level.

FIGURE 3 HERE

To explore this further and attempt to understand whether there is a convergence of family formation across different countries, it is possible to use sequence analysis and study country-specific variation over time. We give an example of such analysis, showing how countries differ in the trends of family trajectories over time.

We estimated a multinomial regression model to calculate the probability of cluster membership, using a very simple set of covariates: gender, country fixed effects and birth year. In this way, we prevent that sample differences across countries affect the interpretation of the results. For sake of simplicity, we report the predictive probabilities calculated from this model for the two clusters that exhibit grater changes over time: "marriage and childbearing" and "cohabitation." Figures 4 and 5 show country differences in family formation over time. Interestingly, it is possible to notice that the decline of traditional family formation has not been homogeneous in European countries. While we observe a sharp decline in Western Europe, this trend has not been paired in Eastern European countries. On the contrary, the proportion of women in this typology of family formation has increased in the post-Soviet societies.

FIGURE 4 AND 5 HERE

The second group we study is the cluster characterized by cohabitation. Figure 5, shows that cohabitation has become more and more common in European countries. Most of the countries experienced a rise in cohabitation. Nordic countries, together with France, have the highest proportion of individuals born after 1950 in this cluster. Eastern European countries, together with Portugal and Spain, lag behind, with slower change in the proportion of people experiencing cohabitation.

6   Discussion and Conclusion

The past few decades have witnessed substantial changes in family structure. There has been a postponement in the beginning of family formation, new types

of unions have emerged (e.g. cohabitation), the complexity of families has increased, with divorces, remarriages, multi-partnered fertility. Several factors may account for the dramatic change in marriages rates, age at first birth, diffusion of cohabitation and divorce experienced by most countries in European. Individualization and secularization, women empowerment and gender equality, institutional structures, and long-term cultural differences across societies have been proposed as possible causes of heterogeneity in life course trajectories. Any attempt to identify explanations for family change has to deal with the diversity and de-standardization that family trajectories have gone through in the past decades.

The complexity in family change poses several methodological challenges, and standard event history techniques do not take into account the interrelation among different events, and the possible influence they have on each other. Hence, we reviewed the use of *sequence analysis* as a holistic approach for the study of life course in family demography. Sequence analysis is a series of techniques that take into account timing, sequencing, duration of events, to address the complexity that characterizes family demographic processes.

After formally describing how sequence analysis can be used for family demography, we showed how it is possible to reduce the number of life course trajectories into typologies using cluster analysis. We also reviewed the recent literature in the field that has implemented sequence analysis to investigate family change and the interaction of family events with other life domains. Finally, we illustrated an example of how to use sequence analysis to describe family trajectories by using data for 23 European countries. This empirical application shows how different family trajectories can be classified in few typologies and how the distribution of different family formation pathways changes in different societies over time. Our analysis shows major changes in family formation such as the increase in cohabitation and the gradual decline of traditional family formation. At the same time, we noted how late transition to adulthood has become more common, in particular among men.

Sequence analysis has been rarely used for cross-country comparative analysis. Our empirical analysis shows the value of using this tool for comparing family change in different countries. For instance, traditional family formation has been declining mostly in Western Europe but not in Eastern European countries, at least for the birth cohorts considered in this analysis. The analysis proposed here is only illustrative of the potential use of sequence analysis for comparative studies in family demography and does not claim to be exhaustive. In particular, we do not provide any explanation of why different trends in family formation have been observed in the last decades in different countries. Nevertheless, we think that sequence analysis provides a useful and powerful set of techniques that can deal with the complexity of the life course and that it represents a valid tool to investigate the pressing and emerging issues in family demography.

References

Aassve, Arnstein, Francesco C Billari, and Raffaella Piccarreta. 2007. "Strings of adulthood: A sequence analysis of young British women's work-family trajectories." *European Journal of Population/Revue européenne de Démographie* 23:369–388.

Abbott, Andrew. 1995. "Sequence analysis: new methods for old ideas." *Annual Reviews in Sociology* 21:93–113.

Abbott, A and A Hrycak. 1990. "Measuring Resemblance in Sequence Data: An Optimal Matching Analysis of Musicians' Careers." *American Journal of Sociology* 96:144–185.

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 & Research* 38:420–462.

Allan, Graham, Sheila Hawker, and Graham Crow. 2001. "Family diversity and change in Britain and Western Europe." *Journal of Family Issues* 22:819–837.

Amato, Paul R., Nancy S. Landale, Tara C. Havasevich-Brooks, Alan Booth, David J. Eggebeen, Robert Schoen, and Susan M. McHale. 2008. "Precursors of young women's family formation pathways." *Journal of Marriage and Family* 70:1271–1286.

Baizan, P, A Aassve, and F C Billari. 2003. "Cohabitation, marriage, and first birth: The interrelationship of family formation events in Spain." *European Journal of Population/Revue européenne de Démographie* 19:147–169.

Balbo, Nicoletta and Nicola Barban. 2014. "Does Fertility Behavior Spread among Friends?" *American Sociological Review* 79:412–431.

Barban, Nicola. 2013. "Family Trajectories and Health: A Life Course Perspective." *European Journal of Population / Revue européenne de Démographie* 29:357–385.

Barban, Nicola and Francesco 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)* 61:765–784.

Barban, N., X. de Luna, E. Lundholm, I. Svensson, and F. C. Billari. 2018. "Causal Effects of the Timing of Life-course Events: Age at Retirement and Subsequent Health." *Sociological Methods & Research* 0:0049124117729697.

Bianchi, S. M. 2011. "Family Change and Time Allocation in American Families." *The ANNALS of the American Academy of Political and Social Science* 638:21–44.

Billari, Fc and Chris Wilson. 2001. "Convergence towards diversity? Cohort dynamics in the transition to adulthood in contemporary Western Europe." *Max Planck Institute for Demographic . . .* 49:0–29.

Billari, Francesco C. 2001a. "Sequence Analysis in Demographic Research." *Canadian Studies in Population* 28:439–458.

Billari, Francesco C. 2001b. "The analysis of early life courses: Complex descriptions of the transition to adulthood." *Journal of Population Research* 18:119–142.

Billari, Francesco C. 2004. "Becoming an Adult in Europe : A Macro (/ Micro) -Demographic Perspective." *Demographic Research* Special Co:15–43.

Billari, Francesco C. and Aart C. Liefbroer. 2010. "Towards a new pattern of transition to adulthood?" *Advances in Life Course Research* 15:59–75.

Blossfeld, Hans-Peter and Johannes Huinink. 1991. "Human Capital Investments or Norms of Role Transition? How Women's Schooling and Career Affect the Process of Family Formation." *American Journal of Sociology* 97:143–168.

Bras, Hilde, Aart C Liefbroer, and Cees H Elzinga. 2010. "Standardization of pathways to adulthood? An analysis of Dutch cohorts born between 1850 and 1900." *Demography* 47:1013–1034.

Brzinsky-Fay, Christian. 2014. "The measurement of school-to-work transitions as processes: about events and sequences." *European Societies* 16:213–232.

Cherlin, Andrew J. 2010. *The marriage-go-round: The state of marriage and the family in America today*. Vintage.

Danziger, Sheldon and David Ratner. 2010. "Labor Market Outcomes and the Transition to Adulthood." *The Future of Children* 20:133–158.

Elder, Glen H. Jr.;. 1985. "Perspectives on the life course." *Life course dynamics: Trajectories and transitions* 1:23–49.

Elzinga, Cees, A W Hoogendoorn, and W Dijkstra. 2007. "Linked Markov Sources: Modeling Outcome-Dependent Social Processes." *Sociological methods and research* 36:26.

Elzinga, Cees H. 2003. "Sequence Similarity: A Nonaligning Technique." *Sociological Methods & Research* 32:3–29.

Elzinga, Cees H. 2010. "Complexity of Categorical Time Series." *Sociological Methods & Research* 38:463–481.

Elzinga, Cees H and Aart C Liefbroer. 2007. "De-standardization of family-life trajectories of young adults: A cross-national comparison using sequence analysis." *European Journal of Population/Revue européenne de Démographie* 23:225–250.

Esping-Andersen, Gosta. 2013. *The three worlds of welfare capitalism*. John Wiley & Sons.

Esping-Andersen, Gøsta and Francesco C. Billari. 2015. "Re-theorizing Family Demographics." *Population and Development Review* 41:1–31.

Fasang, Anette Eva. 2014. "New Perspectives on Family Formation: What Can We Learn from Sequence Analysis?" In *Advances in sequence analysis: Theory, method, applications*, pp. 107–128. Springer.

Fasang, Anette Eva and Marcel Raab. 2014. "Beyond Transmission: Intergenerational Patterns of Family Formation Among Middle-Class American Families." *Demography* 51:1703–1728.

Ferrari, Giulia and Ariane Pailhé. 2017. "Transition to adulthood in France: Do children of immigrants differ from natives?" *Advances in Life Course Research* 31:34–56.

Ferrera, Maurizio. 1996. "The 'southern model' of welfare in social Europe." *Journal of European Social Policy* 6:17–37.

Fokkema, Tineke and Aart C. Liefbroer. 2008. "Trends in living arrangements in Europe: Convergence or divergence?" *Demographic Research* 19:1351–1418.

Furstenberg, Frank F. 2014. "Fifty Years of Family Change: From Consensus to Complexity." *Annals of the American Academy of Political and Social Science* 654:12–30.

Gabadinho, Alexis, Gilbert Ritschard, Nicolas Séverin Mueller, and Matthias Studer. 2011. "Analyzing and visualizing state sequences in R with TraMineR." *Journal of Statistical Software* 40:1–37.

Gauthier, Jacques-Antoine, Eric D Widmer, Philipp Bucher, and Cédric Notredame. 2010. "Multichannel sequence analysis applied to social science data." *Sociological methodology* 40:1–38.

Giuliano, Paola. 2007. "Living Arrangements in Western Europe: Does Cultural Origin Matter?" *Journal of the European Economic Association* 5:927–952.

Iacovou, Maria. 2010. "Leaving home: Independence, togetherness and income." *Advances in Life Course Research* 15:147–160.

Jalovaara, Marika and Anette Eva Fasang. 2015. "Are there gender differences in family trajectories by education in Finland?" *Demographic Research* 33:1241–1256.

Jalovaara, Marika and Anette Eva Fasang. 2017. "From never partnered to serial cohabitors: Union trajectories to childlessness." *Demographic Research* 36:1703–1720.

Kaufman, Leonard and Peter J Rousseeuw. 2009. *Finding groups in data: an introduction to cluster analysis*, volume 344. John Wiley & Sons.

Kleinepier, Tom, Helga AG de Valk, and Ruben van Gaalen. 2015. "Life paths of migrants: a sequence analysis of Polish migrants' family life trajectories." *European Journal of Population* 31:155–179.

Lesthaeghe, Ron. 1995. "The Second Demographic Transition in Western Countries: An Interpretation." *Gender and Family Change in Industrialized Countries* pp. 17–62.

Lesthaeghe, Ron. 2010. "The unfolding story of the second demographic transition." *Population and Development Review* 36:211–251.

Lesthaeghe, Ron J. and Lisa Neidert. 2006. "The second demographic transition in the United States: Exception or textbook example?" *Population and Development Review* 32:669–698.

Levenshtein, VI. 1966. "Binary Codes Capable of Correcting Deletions, Insertions and Reversals." *Soviet Physics Doklady* 10:707.

Levine, Joel H. 2000. "But what have you done for us lately? Commentary on Abbott and Tsay." *Sociological methods & research* 29:34–40.

Liefbroer, A 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* 15:45–75.

Lyngstad, Torkild Hovde and Alexia Prskawetz. 2010. "Do siblings' fertility decisions influence each other?" *Demography* 47:923–934.

Macmillan, Ross and R Copher. 2005. "Families in the life course: Interdependency of roles, role configurations, and pathways." *Journal of Marriage and the Family* 67:858–879.

Mayer, Karl Ulrich. 2001. "The Paradox of Global Social Change and National Path Dependencies." In *Inclusions & Exclusions in European Societies*, pp. 80–89.

McDonald, P. 2000. "Gender equity, social institutions and the future of fertility." *Journal of population Research* 17:1–16.

Mcdonald, Peter. 2006. "Low fertility and the state: The efficacy of policy." *Population and Development Review* 32:485–510.

McVicar, Duncan and Michael Anyadike-Danes. 2002. "Predicting successful and unsuccessful transitions from school to work by using sequence methods." *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 165:317–334.

Piccarreta, Raffaella and Francesco C Billari. 2007. "Clustering work and family trajectories by using a divisive algorithm." *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 170:1061–1078.

Pollock, Gary. 2007. "Holistic trajectories: a study of combined employment, housing and family careers by using multiple-sequence analysis." *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 170:167–183.

Potârcă, Gina, Melinda Mills, and Laurent Lesnard. 2013. "Family Formation Trajectories in Romania, the Russian Federation and France: Towards the Second Demographic Transition?" *European Journal of Population / Revue européenne de Démographie* 29:69–101.

Raab, Marcel, Anette Eva Fasang, Aleksi Karhula, and Jani Erola. 2014. "Sibling Similarity in Family Formation." *Demography* 51:2127–2154.

Reher, D S. 1998. "Family Ties in Western Europe: Persistent Contrasts." *Population and Development Review* 24.

Schoen, Robert. 2011. "Gender competition and family change." *Genus* 66.

Schoen, Robert, Nancy Landale, and K Daniels. 2007. "Family transitions in young adulthood." *Demography* 44:807–820.

Schwanitz, Katrin. 2017. "The transition to adulthood and pathways out of the parental home: A cross-national analysis." *Advances in Life Course Research* 32:21–34.

Shanahan, Michael. 2000. "Pathways to Adulthood in Changing Societies: Variability and Mechanisms in Life Course Perspective." *Annual Reviews in Sociology* 26:667–692.

Sironi, Maria, Nicola Barban, and Roberto Impicciatore. 2015. "Parental social class and the transition to adulthood in Italy and the United States." *Advances in Life Course Research* 26:89–104.

Sironi, Maria and Frank F. Furstenberg. 2012. "Trends in the Economic Independence of Young Adults in the United States: 1973-2007." *Population and Development Review* 38:609–630.

Sobotka, Tomáš. 2008. "Overview chapter 6: The diverse faces of the second demographic transition in Europe."

Stevenson, B and J Wolfers. 2007. "Marriage and divorce: Changes and their driving forces." *Journal of Economic Perspectives* 21:27–52.

Studer, Matthias. 2013. "WeightedCluster Library Manual: A practical guide to creating typologies of trajectories in the social sciences with R."

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* 179:481–511.

Studer, Matthias, Emanuela Struffolino, and Anette E. Fasang. 2018. "Estimating the Relationship between Time-varying Covariates and Trajectories: The Sequence Analysis Multistate Model Procedure." *Sociological Methodology* 0:0081175017747122.

Trifiletti, Rossana. 1999. "Southern European welfare regimes and the worsening position of women." *Journal of European Social Policy* 9:49–64.

Van De Kaa, D J. 1987. "Europe's second demographic transition." *Population bulletin* 42:1–59.

Van Winkle, Zachary. 2017. "Family Trajectories Across Time and Space: Increasing Complexity in Family Life Courses in Europe?" *Demography* .

Widmer, Eric D and Gilbert Ritschard. 2009. "The de-standardization of the life course: Are men and women equal?" *Advances in Life Course Research* 14:28–39.

Wu, L L. 2000. "Some comments on "Sequence Analysis and Optimal Matching Methods in Sociology: Review and Prospect:." *Annual Review of Sociology* 29:41–64.

Tables

Short name	Marital status	Childbearing	Living outside parental home
S (Single)	Single	No	No
SL (Single Left)	Single	No	Yes
SP (Single Parent)	Single	Yes	Yes
C (Cohabiting)	Cohabiting	No	Yes
CP (Cohabiting Parent)	Cohabiting	Yes	Yes
M (Married)	Married	No	Yes
MP (Married Parent)	Married	Yes	Yes

Table 1: State-space of family events

Index	Abbrv.	Best number of groups	Stat
Point Biserial Correlation	PBC	5	0.75
Hubert's Gamma	HG	5	0.88
Average Silouette Width	ASW	5	0.41
Calinksi- Harabasz index	CH	2	7438.17
Pseudo R^2	R2	8	0.48
Hubert's C	HC	5	0.06

Table 2: Best number of cluster solutions according to several measures of goodness of fit available in the `weightedCluster` package (Studer, 2013).

	Late Transition	Marriage and Childbearing	Cohabitation Cohabitation	Transition to Independence	Marriage, no Children	All sample
% Ever cohabited	21.87	19.80	100.00	16.48	27.48	26.75
% Ever married	42.90	100.00	23.36	55.49	100.00	82.60
% Ever had child	30.36	99.95	68.36	44.22	13.06	79.30
% Women	44.70	60.20	50.68	40.09	55.42	55.51
Median year of birth	1953	1950	1962	1950	1951	1949
Average age at first transition	27.35	20.67	19.60	19.23	20.34	21.11
Median age at marriage	-	23	-	34	23	23
Median age at parenthood	-	24	30	-	-	25
% Sample	13.21	65.52	8.17	8.07	5.03	100.0
N	4,042	20,047	2,500	2,469	1,539	30,597

Table 3: Descriptive statistics. Typologies of life course trajectories.

Country	Late transition	Marriage and childbearing	Cohabitation	Transition to independence	Marriage no children	N
AT	15.50	61.14	8.78	6.90	7.69	1,652
BE	15.02	69.88	4.66	2.85	7.59	1,265
BG	11.04	80.75	1.42	3.68	3.11	1,060
CH	12.30	54.69	9.67	16.86	6.47	1,406
CY	15.26	75.04	0.57	7.13	2.00	701
DE	15.55	62.05	8.10	7.59	6.71	2,161
DK	6.01	57.51	20.77	9.70	6.01	1,165
EE	8.75	69.14	8.28	9.60	4.23	1,063
ES	26.27	60.79	3.17	6.75	3.02	1,260
FI	10.30	61.94	12.27	11.10	4.38	1,369
FR	10.41	63.65	14.46	7.84	3.65	1,480
GB	12.26	58.70	9.40	11.25	8.39	1,787
HU	11.77	77.24	2.70	3.05	5.23	1,147
IE	23.36	47.26	7.52	17.52	4.34	1,130
NL	16.19	58.23	9.38	10.49	5.70	1,439
NO	6.00	62.99	16.45	10.94	3.62	1,216
PL	11.64	76.63	3.21	5.41	3.12	1,091
PT	20.31	66.61	2.19	4.32	6.57	1,689
RU	10.52	77.12	2.77	4.86	4.74	1,626
SE	8.37	52.96	25.56	10.30	2.81	1,350
SI	13.99	69.75	7.00	6.90	2.36	1,015
SK	11.49	79.70	1.76	4.08	2.97	1,079
UA	8.99	78.08	2.97	5.46	4.50	1,446

Table 4: Distribution of group membership by country.
AT=Austria; BE=Belgium; BG=Bulgaria; CH=Switzerland; CY=Cyprus; DE=Germany; DK=Denmark; EE=Estonia; ES=Spain; FI=Finland; FR=France; GB=United Kingdom; HU=Hungary; IE=Ireland; NL=Netherlands; NO=Norway; PL=Poland; PT=Portugal; RU=Russia; SE=Sweden; SI=Slovenia; SK=Slovakia; UA= Ukraine
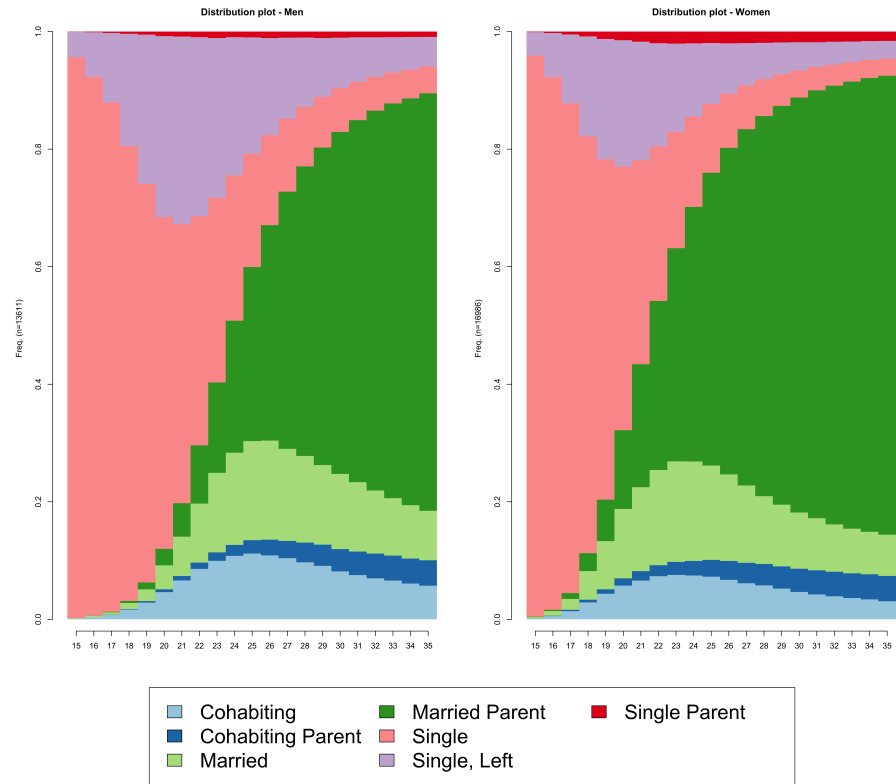
Figures

Figure 1: Distribution plot of life course states for men and women from age 15 to age 35. Data are based on 23 countries from Wave 3 of the European Social Survey
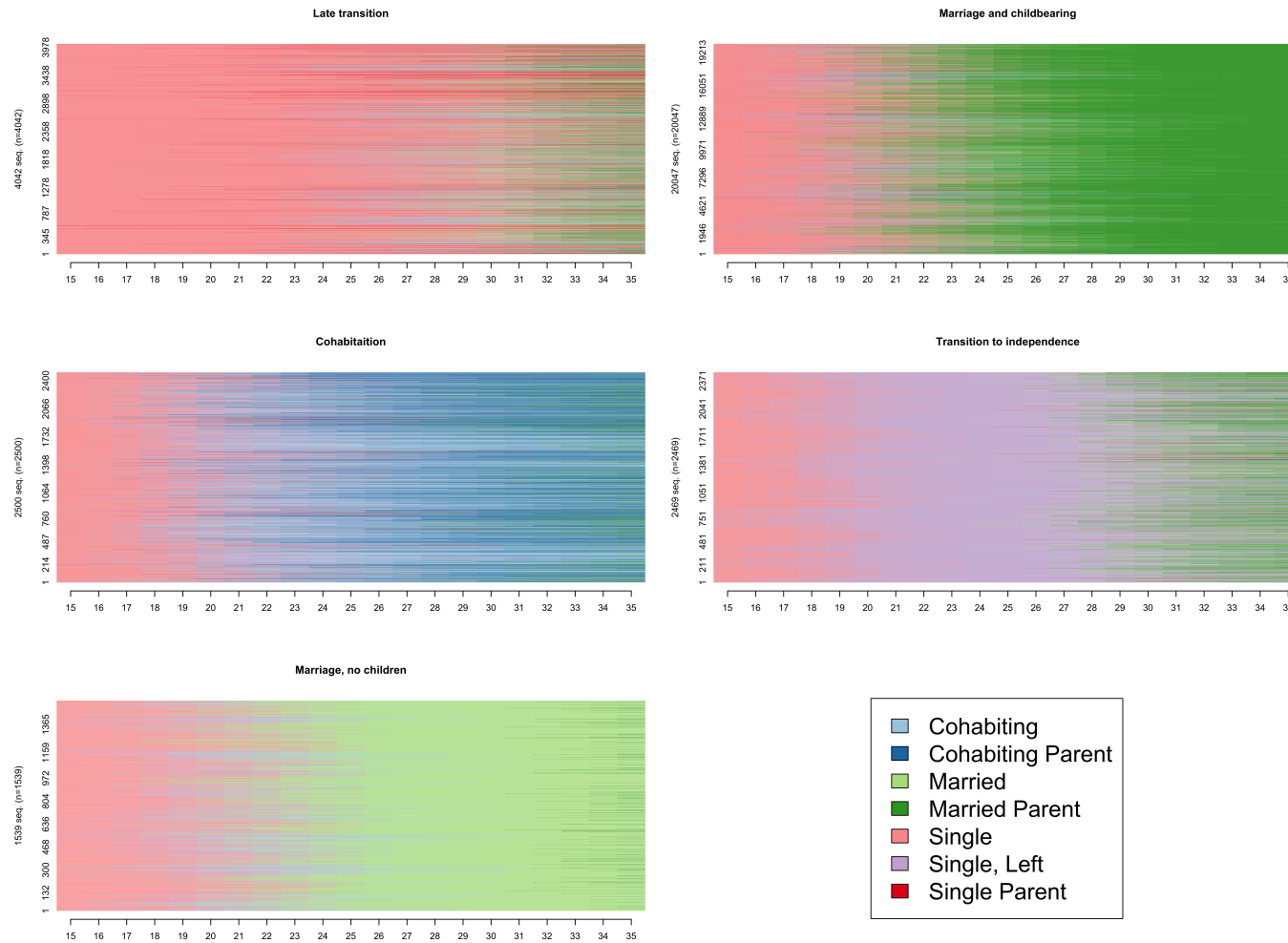
Figure 2: Graphical representation of groups of life course trajectories obtained after applying cluster analysis. Each line represents a life trajectory from age 15 to 35. Wave 3 of European Social Survey. Men and women.
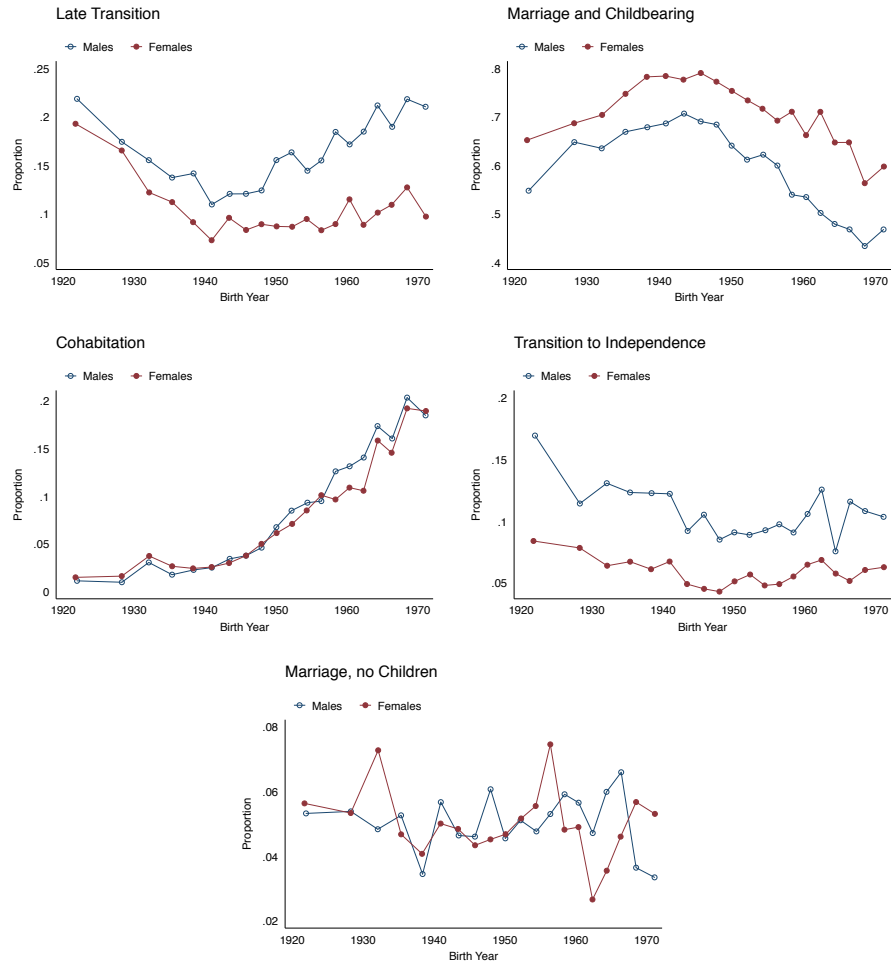
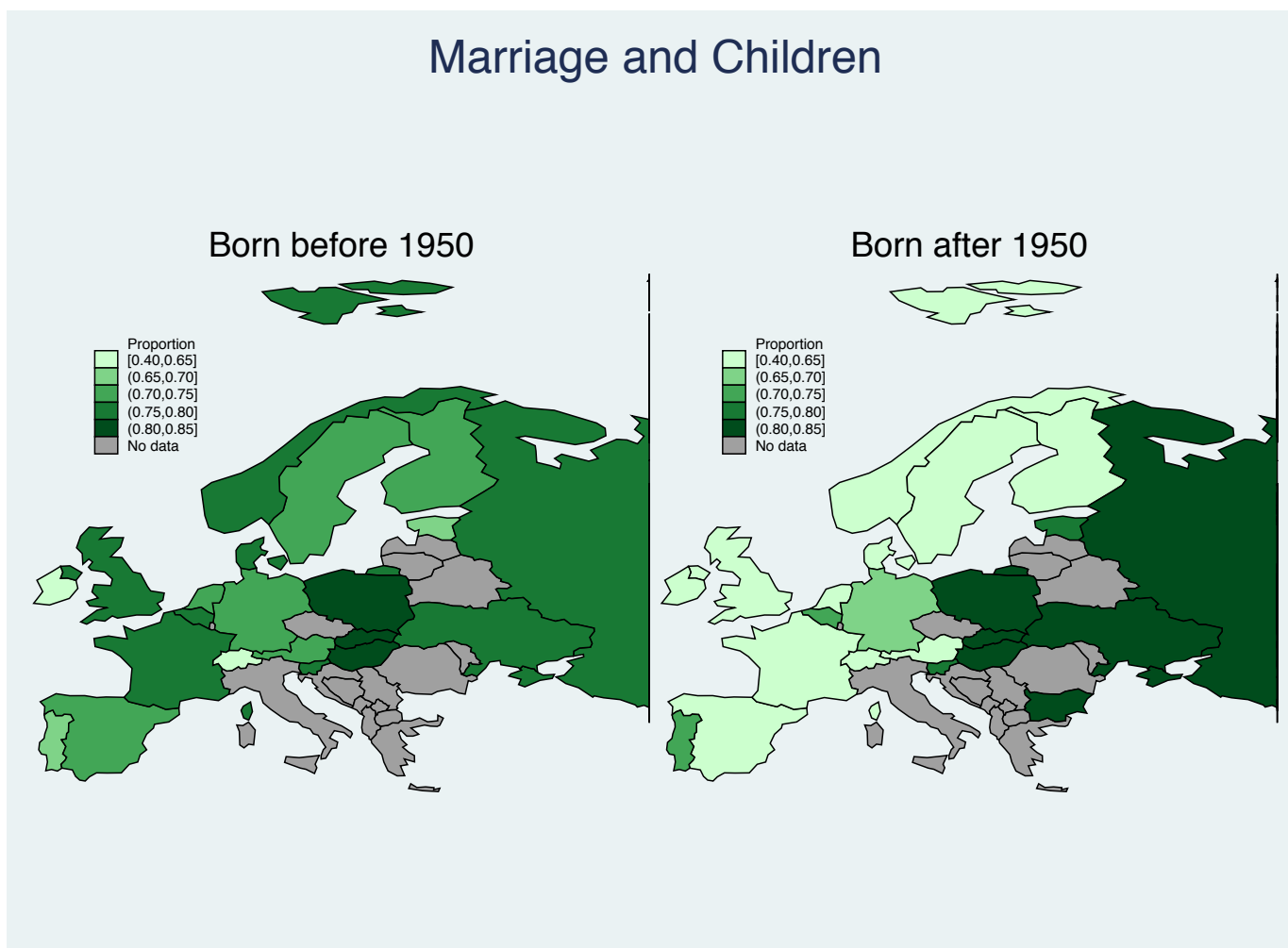Figure 3: Group membership by sex and birth cohort. Binned scatter plots controlling for country fixed effects.

Figure 4: Predicted probability of cluster membership. Cluster 2: "Marriage and Childbearing". Women only.
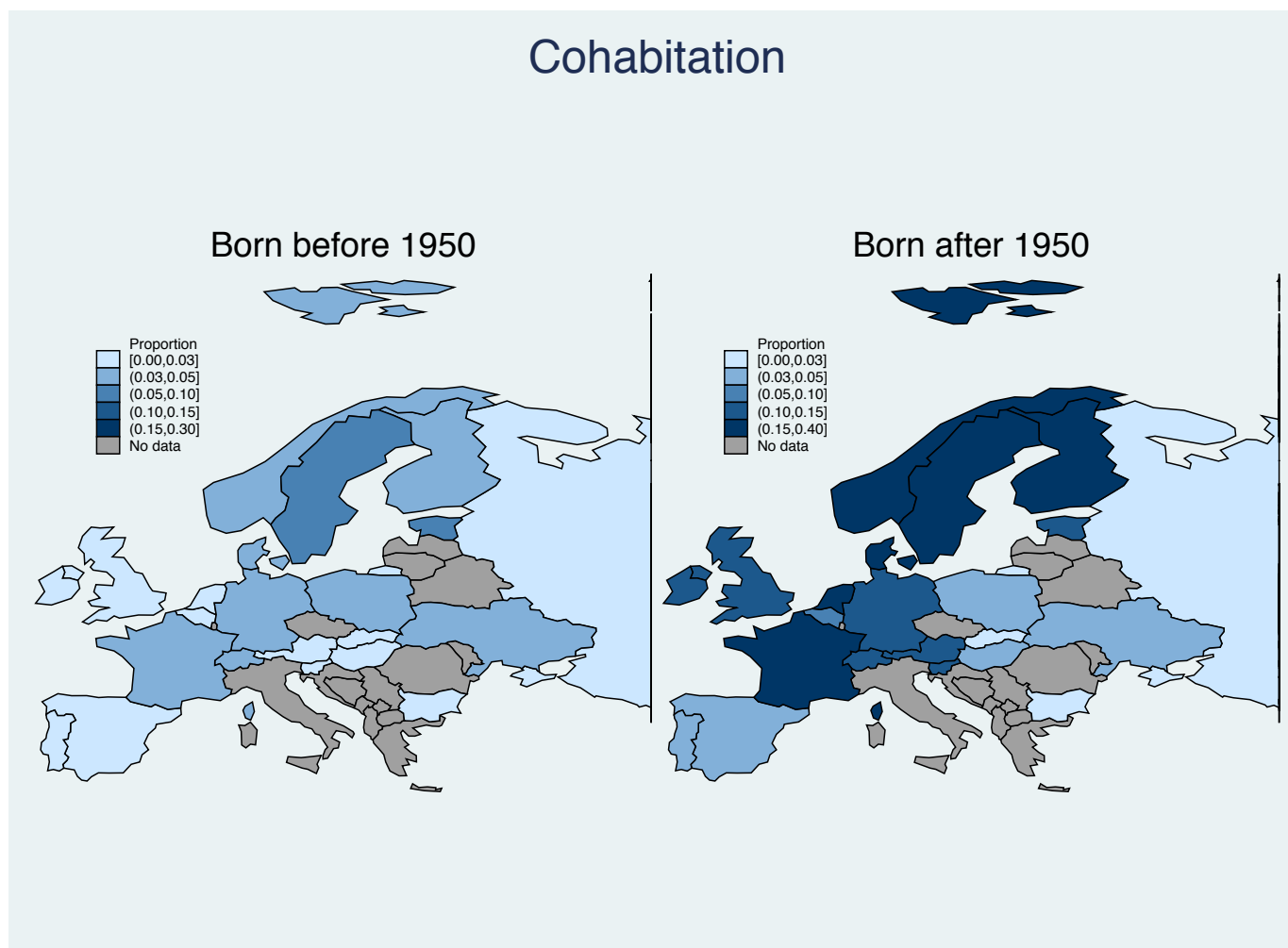
Figure 5: Predicted probability of cluster membership. Cluster 3: "Cohabitation". Women only