MAPLES: a general method for the estimation of age profiles from standard demographic surveys (with an application to family and fertility)

Roberto Impicciatore
"Dondena" Centre for Research on Social Dynamics & DEAS, University of Milan

Francesco C. Billari
"Dondena" Centre for Research on Social Dynamics & Department of Decision Sciences, Università Bocconi

Abstract

In this paper we propose MAPLES (Method for Age Profiles Longitudinal EStimation), a general method for the estimation of age profiles which uses standard micro-level retrospective demographic survey data. After the specification of data requirements, the method is implemented through a data processing routine and the estimation of specific regression models. For the more relevant transitions in the field of living arrangements and fertility, MAPLES estimates smoothed age profiles and relative risks for time-fixed and time-varying covariates. An example of application is carried on data from Italy. The major advantage of this method is that it can be applied to every setting where micro-level data on transitions are available from a large-scale representative survey (e.g., Fertility and Family Survey; Generations and Gender) and for different kind of transitions. MAPLES is implemented through the R software package and it can be applied to any suitable dataset through the execution of R functions.

Acknowledgments: this paper has been prepared within the MicMac project. The data for this analysis have been provided by ISTAT (Italian National Institute of Statistics). The author would like to thank Eva Beaujouan for her many helpful comments and discussions regarding this work.

1. Introduction

The estimation of smooth age profiles for demographic events is a problem of general interest, which has triggered a substantial amount of research over the last centuries. Starting from mortality, e.g. with Gompertz model, scholars have also attacked fertility and marriage. This type of estimation has then been embedded within the event history framework, making room for the role of time-constant and time-varying covariates. More recently, models that minimize the strength of the assumption on the shape of the underlying profiles have been proposed and used. An example is the fertility model by Schmertmann (2003) based on splines. The interpolation of age profiles through splines had been proposed earlier on by McNeil et al. (1977).

When the interest is not on a single transition, but on a multistate set of transitions, the need for a general method for the estimation of age profiles is even more evident. This becomes crucial when developing multistate population forecasts, i.e. when forecasts are developed in order to account for the complexity of life course trajectories. A typical example is the need to forecast individuals according to statuses such as living arrangement and family status. This is the approach, for instance, of the "MicMac" population forecasting framework, which aims at explicitly taking into account the life course trajectories of individuals in population forecasts (see, e.g., Willekens, 2005; van der Gaag *et al*., 2006). The life course is viewed as a sequence of states and events; each event marks a transition from one state to another (see also the statistical approach developed in Andersen *et al*., 1993). The study of a single transition is based on the estimation of its transition rates (from the original state to the destination state, in a defined state space). From the literature on living arrangements and fertility, we know these behaviours are strongly related to age. Indeed, such variation with age has traditionally been exploited in demographic forecasting.

In this paper, our first aim is to describe a general method for the estimation of age profiles for the main transitions experienced by individuals, as far as living arrangement and fertility behaviours are concerned, which we have been developed within the MicMac project. It is called MAPLES (Method for Age Profiles Longitudinal EStimation). The idea of MAPLES to be able to start from standard retrospective demographic surveys, such as Fertility and Family Surveys or Generation and Gender Surveys, in order to be able to estimate age profiles for various transitions. We present the model also by developing a specific application to Italian data.

This paper is structured as follows. In Section 2 we present, through a flow-chart, the basic components of MAPLES. In Section 3 we discuss the data preparation step. In Section 4 we present the construction of the transition-specific data matrix. In Section 5 the regression and smoothing procedure is presented. Section 6 presents an application to Italian data. The Appendix contains some of the routines and program examples as boxes.

2. The components of MAPLES

Let us start by giving a stepwise description of the components MAPLES, before going into the details of each step.

1. *Data preparation.* In the first step, data are prepared for subsequent computations. Information from the original dataset has to be adapted. The starting dataset has to include the dates of the events to be studied, separately for men and women. Throughout the following text, the word *date* refers to the time point of a certain event measured in months and years. After the specification of the state space and of the possible transitions, an input data file is prepared for subsequent analyses.

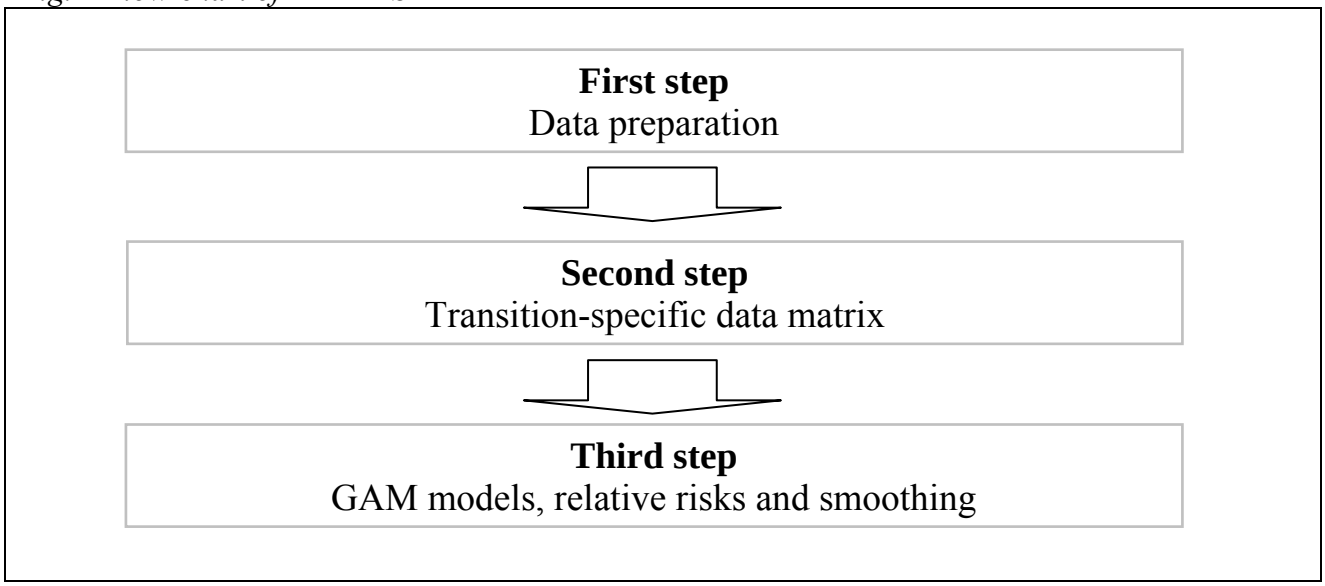
When the initial dataset is ready, some data consistency checks are executed, such as the insertion of missing months, the specification of status variables and the computation of decimal dates and ages at various events. This step is common to every transition.

2. *Preparation of the transition-specific data matrix.* In the second step, MAPLES computes episode-data for each specific transition, taking into account both time-fixed and time-varying covariates. It transforms the "micro-data" structure (one row = one individual) into what we can define a multistate "macro-data" structure (one row = one combination of age and levels of covariates). The resulting data matrix contains events and time of exposure for a specific period of time up to the interview (*window of observation*). This is the *transition-specific data matrix*, as its computation depends on the definition of episodes that are specific to each transition.

3. *Model estimation.* In the third step, we estimate age profiles by modelling the observed set of events and exposure times, which are stored in the transition-specific data matrix, using a smoother function. In particular, MAPLES uses GAMs (Generalized Additive Models) (Hastie & Tibshirani, 1990) in a way that permits to jointly estimate the baseline age profile and the effect of covariates as multiplicative changes from the baseline. For each row of the *transition-specific data matrix*, we model the logarithm of the transition rate (events divided by time of exposure) by adding a (smooth) function of age and a set of fixed covariates. In order to remove the proportionality assumption which would be implied by the summation of log-rates, MAPLES considers the multiplicative change given by a covariate in a piecewise way, i.e. separately for each sub-interval of age. A final smoothing procedure ensures that the estimated final profiles are continuous over age.

MAPLES has been developed with the aim to construct a flexible method that may be applied in different contexts, when survey micro-data are available. As a result, this method can be applied to every setting where relatively standard family and fertility micro-level data are available from a large-scale representative survey. Moreover, it takes into account the most important interactions between different trajectories through the specification of different covariates, both time-constant and time-varying over the life course. Being based on regression models, the method also permits to evaluate confidence intervals and to test hypotheses. The whole method is implemented in R software and it could be easily recalled as an R function in order to be applied to real data.

Fig. 1 Flow-chart of MAPLES



3. Data preparation

In the first step, the main focus is the preparation of the initial dataset and the computation of basic quantities at the individual level such as dates in a decimal format, ages at various events and status variables. In order to specify the characteristics of the input data file, we need to define unambiguously the state space and the set of transitions that we want to analyze. Moreover, we have to cope with inconsistent and missing data.

3.1 State space and transitions

We explain the functioning of MAPLES by focusing on the specific example we developed. As in de Beer et al. (2006), we consider transitions in three life course domains: 1) marital status; 2) fertility and 3) living arrangement. The choice of these transitions is strongly influenced by micro data availability. More specifically, we define the state space on the basis of the standard information that is generally available in Gender and Generation Surveys (Vikat *et al*., 2007). Moreover, states are chosen in such a way that transition from state A to state B is caused by a non-repeatable event. We call the generic transition as TR*X* where *X* is an identification code.

Let us define the state space for each of the domains. As far as *marital status* is concerned, we distinguish between first marriage and second (or following) marriage. We do not consider further transitions after the entry in the second marriage. In table 1 we can see the (qualitative) shape of the transition matrix.

Table 1. Marital status. State space and transitions

From \ to	Never married	First marriage	Second marriage	Divorced	Widowed
Never married	-	TR1			
First marriage		-		TR2	TR3
Second marriage			-		
Divorced			TR4	-	
Widowed			TR5		-

There are many possible transitions in the field of *living arrangements*, but it is rare to have detailed information. This forces us to limit the number of possible states. Given the usual information contained in the GGS, we focus our attention to the following states: at parental home, alone/with others (never in union), first union, separated (after 1st union disruption), second union. We do not consider further transitions after the entry in the second union (table 2)

As far as *fertility* is concerned, the possible states are the following: childless, 1 child, 2 children, 3 children, 4 or more children. Transition such as 0 → 2, 1→3, etc. caused by multiple births are not taken into account. A childless woman who has a twin birth simply experiences the transition 0→1 and 1→2 at the same date. The transition matrix is given in table 3.

Table 2. Living arrangement. State space and transitions

From \ to	at parental home	Alone/with others (never in union)	First union	Separated (after 1st union disruption)	Second union
at parental home (never in union)	-	TR7	TR6		
Alone/with others		-	TR8		
First union			-	TR9	
Separated (after 1st union disruption)				-	TR10
Second union					-

Table 3. Fertility (own children ever born). State space and transitions

From \ to	childless	1 child	2 children	3 children	4+ children
Childless	-	TR11			
1 child		-	TR12		
2 children			-	TR13	
3 children				-	TR14
4+ children					-

3.2 The input data file

Transitions as presented earlier are available from retrospective questionnaires. In order to feed the subsequent steps, an input data file has to be specified. In this case we discuss a more practical instance. Table 4 reports the record structure that is needed in input datafile (as required by the implementation of MAPLES using the R software). Variables highlighted in gray (id, weight, date of birth, date of interview, sex, and education) are compulsory: missing values are not allowed. Further information is optional and we may simply do not include it, or part of it, in the dataset. Weights must be normalized (average weight must be 1). If data have no weights it is sufficient to specify a unit weight for all individuals in the sample. We assume that individuals in the dataset are aged 18 and more at the interview. All dates are expressed in calendar month (format MM: 1 to 12) and year (format YYYY).
We now consider the presence of typical time-constant covariates: sex and education (the latter taken here as time-constant only for simplicity). Sex is coded as follows:

1. Men
2. Women

Edu is the level of education reported by respondents (with at least 18 years old) at the interview. We consider this variable as *time-fixed* in the sense that values remain constant throughout the biography. The variable is coded as follows:

1. Primary (ISCED0 *pre-primary education* and ISCED1 *first stage of basic education*)
2. Lower secondary (ISCED2 *second stage of basic education*)

3. Upper secondary (ISCED3 *upper secondary education* and ISCED4 *post secondary non-tertiary education*)
4. Tertiary (ISCED5 *first stage of tertiary education* and ISCED6 *second stage of tertiary education*)

A transition is well-defined when we know which event causes it, at which point in time it occurs and when the individual starts to be *at risk* of living this event. Moreover, at a certain point in time the individual may experience an event that does not permit to follow his/her life course further on, i.e. the observation is censored (Blossfeld and Rowher, 2002). With longitudinal retrospective data this usually happens at the interview or, for example, at the death of spouse when we are studying the transition to divorce for married people.

Table 4. Initial dataset record structure.

variable name	Description	Format
id	Identification number (individual level)	8
weight	Normalized Weight	10
ybirth	Year of birth	4
mbirth	Month of birth	2
yint	Year of interview	4
mint	Month of interview	2
yexit	Year of exit from parental home	4
mexit	Month of exit from parental home	2
ymarr	Year of first marriage	4
mmarr	Month of first marriage	2
ydiv	Year of divorce (first marriage)	4
mdiv	Month of divorce (first marriage)	2
yved	Year of death of spouse (first marriage)	4
mved	Month of death of spouse (first marriage)	2
ypartn	Year of first union (cohabitation or marriage)	4
mpartn	Month of first union (cohabitation or marriage)	2
ydiss	Year of first union (cohabitation or marriage) disruption	4
mdiss	Month of first union (cohabitation or marriage) disruption	2
ypartn2	Year of second union	4
mpartn2	Month of second union	2
ymarr2	Year of second marriage	4
mmarr2	Month of second marriage	2
ych1	Year of first child's birth	4
mch1	Month of first child's birth	2
ych2	Year of second child's birth	4
mch2	Month of second child's birth	2
ych3	Year of third child's birth	4
mch3	Month of third child's birth	2
ych4	Year of fourth child's birth	4
mch4	Month of fourth child's birth	2
sex	Sex	1
edu	Level of education (ISCED)	1

In table 5, we show for each possible transition the events that define episodes, i.e. the events that cause the entry into the period at risk, the transition itself and the events that imply the exit from observation (censoring). Given these information we can specify the dates required for the analysis of each transition. In any case the date of interview and the birth date of respondents are necessary.

Table 5. Episodes and dates required for each transition.

TRANSITION	Episode starts at	Events that cause transitions	Events that cause censoring	Dates required[(1)]
TR1 never-married → married (1st marriage)	respondent's birth	1st marriage	interview	(***ymarr***,*mmarr*)
TR2 married (1st marriage)→ divorced	1st marriage	divorce	death of spouse, interview	(***ymarr***,*mmarr*) (***ydiv***,*mdiv*) (***yved***, *mved*)
TR3 married (1st marriage)→ widowed	1st marriage	death of spouse	divorce, interview	(***ymarr***,*mmarr*) (***ydiv***,*mdiv*) (***yved***, *mved*)
TR4 divorced→ married (2nd marriage)	divorce	2nd marriage	death of spouse, interview	(***ymarr***,*mmarr*) (***ydiv***,*mdiv*) (***yved***,*mved*) (***ymarr2***,*mmarr2*)
TR5 widowed→ married (2nd marriage)	death of spouse	2nd marriage	interview	(***ymarr,***,*mmarr*) (***ydiv***,*mdiv*) (***yved***,*mved*) (***ymarr2***,*mmarr2*)
TR6 at parental home (never in union) → first union	date of birth	exit from parental home for union	exit from parental home for other reasons ,interview	(***ypartn***,*mpartn* ()***yexit***,*mexit*)
TR7 at parental home→ alone/with others (never in union)	date of birth	exit from parental home for other reasons	exit from parental home for union, interview	(***ypartn***,*mpartn*) (***yexit***,*mexit*)
TR8 alone/ with others (never in union) → first union	exit from parental home	1st union	interview	(***ypartn,***,*mpartn*) (***yexit***,*mexit*)
TR9 first union→ separated (after 1st union disruption)	1st union	1st union dissolution	interview	(***ypartn***,*mpartn*) (***ydiss***,*mdiss*)
TR10 alone or with other persons (after the 1st union disruption)→ with a partner (2nd union)	1st union dissolution	2nd union	interview	(***ydiss***,*mdiss*) (***ypartn2***,*mpartn2*)
TR11 childless → 1 child	respondent's birth	1st child's birth	interview	(***ych1***,*mch1*)
TR12 1 child → 2 children	1st child's birth+ 9 months	2st child's birth	interview	(***ych2***,*mch2*) *(****ych1***,*mch1*)
TR13 2 children → 3 children	2nd child's birth+ 9 months	3rd child's birth	interview	(***ych3***,*mch3*) (***ych2***,*mch2*) (***ych1***,*mch1*)
TR14 3 children → 4 children	3rd child's birth	4th child birth	interview	(***ych4***,*mch4*) (***ych3***,*mch3*) (***ych1***,*mch1*) (***ych2***,*mch2*)

(1) Date of births, date at the interview and sex are always needed.

3.3 Status variables

The *status* variable indicates if an event has been experienced or not before the interview. It is computed internally, according to the availability of the event's year. The rule is the following: when the year of a date is missing, the event is considered as not having been experienced (yet). For example, let us consider transition TR11 (from "first child" to "second child"). If the year of birth of the second child is missing, we assume that the respondent has only one child at the interview and the corresponding status variable is 0. If the year of second child's birth is not missing, the status variable is 1 (event occurred). Anyhow, if the year of first child's birth is missing, we assume that the respondent is still childless at the interview: the case is "not applicable" in the analysis of second birth because the individual has never been at risk to live the event and the status variable is fixed at 9.
As a general rule, for a generic transition TR*X* the *status* variable have the following values:

- *0* if the individual has never experienced TR*X* at the time of the interview (the case is *censored* at the interview*);*
- *1* if the individual experienced TR*X* before the interview;
- *9* if the case is not applicable, i.e. the individual has never been at risk to experience TR*X*.

The only status variable that does not follow this rule is the one that is associated to the event *leaving parental home*. There are two possible destinations: union (marriage or cohabitation) or other reasons (single living or with other persons). Moreover, there are no "not applicable" cases since everybody is considered as beginning their life in their parents' household. We do not know the reason for leaving home but considering the date at exit from parental home (*yexit, mexit*) and the date at first union (*ypartn,mpartn*), we can compute the status variable with the following categories:

- *0* no exit (no *yexit*);
- *1* union (*yexit*≥*ypartn*);
- *2* other reason (*yexit*≤ypartn or no *ypartn*).

The internal computation of status variables strictly requires that unknown dates are not written as missing in the initial dataset. A missing year simply means that the associated event did not occur. This is in line with indications given by other authors. For example, Matsuo & Willekens (2003) specify that a missing year means that the event did not occur even when the respondent indicated, in another item, that the event did occur. Therefore, the user must pay attention to missing values in the dataset, check possible missing dates and exclude ambiguous cases from the dataset.

3.4 Missing months

It is desirable that the user specifies months using all available information. However, if a missing year is critical and should probably lead to the exclusion of a case, a missing month (when year is known) can be overcome without jeopardizing analyses using simple hypotheses.
Two circumstances are conceivable:

- months are totally missing: for a specific event, the month of occurrence is not available because this information is not gathered in the questionnaire.
- months are partially missing: month was asked in the interview but respondent did not answer.

In both cases, MAPLES estimates missing months through the application of Uniform distribution, i.e. it inputs a random number between 1 and 12. For example, if we do not have the month at the first marriage (*mmarr*) but we have the year (*ymarr*), date of marriage (and age of marriage) can be computed by setting month of first marriage as a random number form 1 to 12. In some cases, we have additional information that can be used as constraints in the estimation of missing months. All the criteria used with this aim can be read in table 6. Missing months are not allowed for date of birth and interview.

Table 6. Constraints used for the imputation of missing months

Missing month	Not missing month	Condition	Input missing month as month at
exit	1st union	year of exit = year of 1st union	1st union
1st union	1st marriage	year of 1st union= year of 1st marriage	1st marriage
1st union	exit	year of 1st union= year of exit	exit
1st union disruption	1st union	year of union disruption = year of 1st union	random: 1st union to 12
1st union disruption	2nd union	year of union disruption = year of 2nd union	random: 1 to 2nd union
2nd union	1st union disrupt.	year of 2nd union=year of union disruption	random: 1st union disrupt. to 12
1st marriage	1st union	year of 1st marriage= year of 1st union	1st union
1st marriage	exit	year of 1st marriage= year of exit	exit
divorce	2nd marriage	year of divorce = year of 2nd marriage	random: 1 to 2nd marriage
death of spouse	2nd marriage	year of death of spouse = year of 2nd marriage	random: 1 to 2nd marriage
death of spouse	1st marriage	year of death of spouse=year of 1st marriage	random:1st marr. to 12
2nd marriage	divorce	year of divorce = year of 2nd marriage	random: divorce to 12
2nd marriage	death of spouse	year of death of spouse = year of 2nd marriage	random: death of sp. to 12
2nd child	1st child	year of 2nd birth = year of 1st birth	1st birth
2nd child	1st child	year of 2nd birth = year of 1st birth +1 and month of 1st birth>4	random: (1st birth -3) to 12
3nd child	2st child	year of 3rd birth = year of 2nd birth	2nd birth
3nd child	2st child	year of 3rd birth = year of 2nd birth +1 and month of 2nd birth > 4	Random: (2nd birth -3) to 12
4nd child	3st child	year of 4th birth = year of 3rd birth	3rd birth
4nd child	3st child	year of 4th birth = year of 3rd birth +1 and month of 3rd birth > 4	Random: (3rd birth -3) to 12

3.5 Consistency checks

The focus on dates can reveal several inconsistencies that may remain hidden otherwise. In particular, some sequences of dates cannot be real (e.g. second marriage experienced before the end of first marriage, second child born before first child) or dates of some events are clearly not reported even if they clearly occurred (e.g. second marriage is reported while information about the end of first marriage are missing).
Inconsistent sequencing and/or timing of events may be due to typing errors made by interviewer or during the data capture. The user should use all available information in order to correct these inconsistencies, but some of them may remain in the dataset. Table 7 shows the list of consistency checks that MAPLES executes on the initial dataset. When an inconsistent date is detected for a specific case, MAPLES assigns value 9 to the correspondent status variable. This means that the case is dropped from the dataset every time that we want to study a transition that requires the date

reported as inconsistent. This means that the case is dropped from the calculations when we study a transition that requires the date reported as inconsistent. If an inconsistency emerges for another event that is not currently required, the case is included in the analysis. For example, if the *j*-th individual shows inconsistency in the date of exit from parental home, when we analyze first child's birth (TR11), he/she is included in the analysis. Otherwise, if we focus on transition to first union (TR6), the individual is dropped.

Table 7 Criteria used in order to identify inconsistent cases.

Year at:		Year at:
exit from parental home	<	birth
first union	<	birth +14
first union disruption	<	first union
marriage	<	birth +14
divorce (1st marr.)	<	marriage
death of spouse (1st marr.)	<	marriage
second union	<	first union
second union	<	first union disruption
second marriage	<	death of spouse (1st marr.)
second marriage	<	divorce (1st marr.)
first child	<	birth + 14
second child	<	first child
third child	<	second child
fourth child	<	third child

Years of events must have lower or equal to the year of interview.

3.6 Computation of decimal dates and ages

A generic date (*year, month*) is transformed in a continuous expression through the formula

$$date\_event = year - 1900 + \frac{month - 0.5}{2}$$

A date is pointed to the middle of the specified month *mdate*. The correspondent age is computed as:

$$age\_event = date\_event - date\_birth$$

where *date_birth* is the decimal date of birth.

4. Transition-specific data matrix

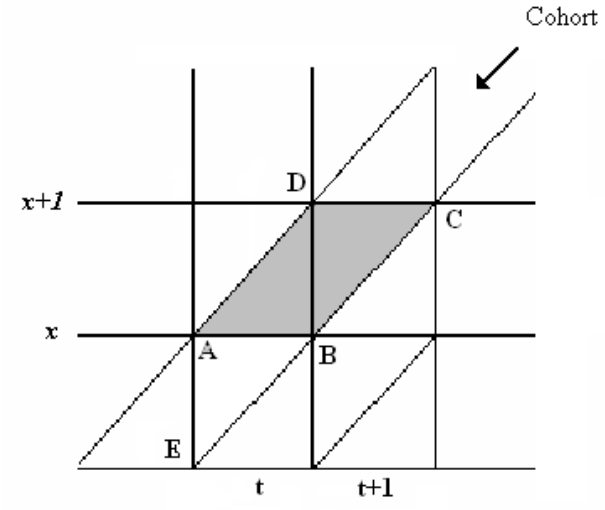
After the first step, all relevant data at the individual level are available. Using these data, our aim is to create a matrix containing, for any combination of covariates, the number of events experienced by individuals and their exposure time at each age included in the windows of observation. In order to do so, we have to specify the type of rates, episode and window of observation considered, the time varying covariates retained, and some additional computation criteria. From this point onward, the procedure is transition-specific, i.e. it has to be repeated for each transition.

4.1 Cohort-period rates and cohort-age rates

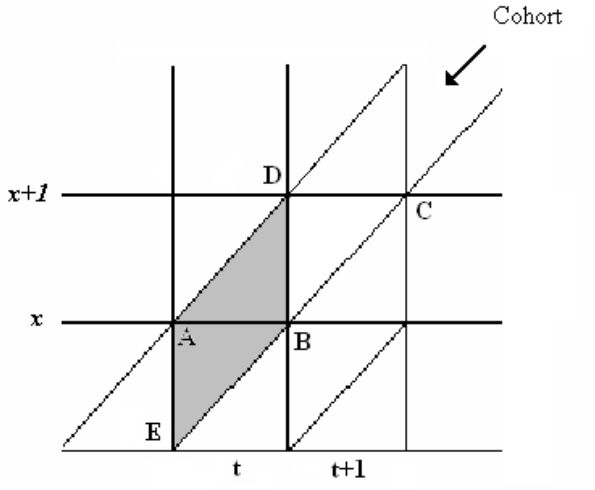
The first point that we have to fix is the kind of rates we want to estimate. In demographic analysis there are basically three kinds of rates, depending on how events are classified: 1) period-age, 2) cohort-age, and 3) cohort-period rates. Given that we refer to retrospective data, which are strictly related to a specific cohort, we have no interest in period-age rates. Cohort-age rates (figure 1a) are the best choice in order to define age profiles: they relate to events that occurred to a specific cohort at age *x* (between the *x*-th and the *x*+1-th birthday). Cohort-period rates (figure 1b) take into account events occurred to a specific cohort during the *t*-th calendar year, then, referring to two different years of age *x-1* and *x*. It is well known in the literature that the latter are the best choice when rates are used in demographic projections. MAPLES can estimate age profiles using both kinds of rates. In the first case (cohort-age rates), the time scale is based on the individual age (episodes are defined by ages at different events) whereas, in the second case (cohort-period rates), the time scale is based on calendar time (episodes are defined by dates of different events). Given that an age profile is described as a vector of rates at different ages, the generic cohort-period rate at time *t* (covering age *x-1* and *x*) is referred to as the rate at age *x*.

Fig. 1 Area of interest in the Lexis diagram according to the kind of transition rate

a. Cohort-age rate



b. Cohort-period rate



4.2 Window of observation and episodes

Here, information in the living arrangement and fertility fields is collected through surveys based on respondents aged at least 18 years at the interview. This is consistent with the dynamics of such behaviors in contemporary Europe. Given that the focus of MAPLES is the estimation of age profiles, especially as used for forecasting purposes, we refer explicitly to the most recent behaviors. A plausible period could be the last five years before the interview but in general we may refer to a generic length of *wl* years. For the *j*-th individual the window of observation is the time interval $\left(t_j^{WIN\_start}, t_j^{INTERVIEW}\right)$ defined as follows:

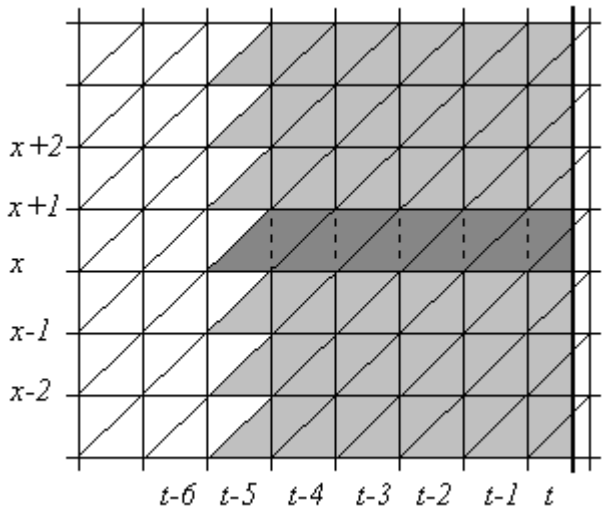
- for cohort-period rates (the time axis is calendar time):
$$t_j^{WIN\_start} = \text{trunc}(date\_int) - wl$$
$$t_j^{INTERVIEW} = date\_int$$
- for cohort-age rates (the time axis is age):
$$t_j^{WIN\_start} = \text{ceiling}(\text{trunc}(\text{date}\_int) - wl - date\_birth)$$
$$t_j^{INTERVIEW} = age\_int$$

where *date_int* and *age_int*, respectively the decimal date and the age at the interview, *date_birth* is the decimal date at birth; the operator *trunc(x)* gives the integer part *x* and *ceiling(x)* returns the smallest integer not less than *x*; finally, *wl* is the window length.

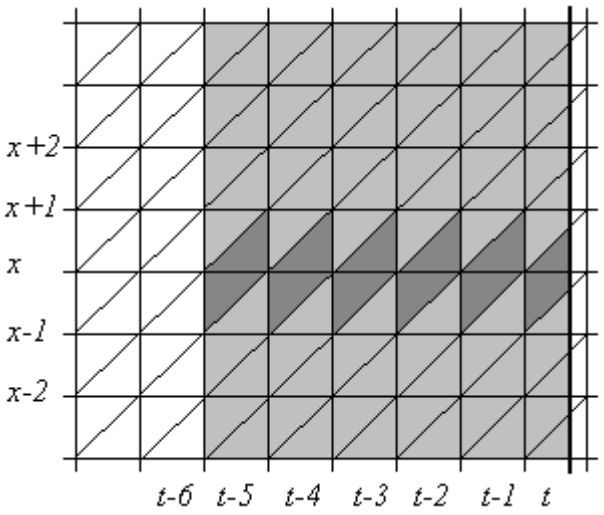
As a general rule, we consider only the events experienced in this window. Taking *wl*=5, the window of observation in the Lexis diagram is the light gray area in figure 3a (cohort-age rates) and in 3b (cohort-period rates). Transition rate at age *x* is computed taking into consideration the black gray area.

Fig. 3. Window of observation in the Lexis diagram (window length=5 years) according to the kind of transition rates. Individuals are interviewed in a precise point in time during year t.

a. Cohort-age rate



b. Cohort-period rate



In section 3.1 and 3.2 we saw that a generic transition TR*X* represents the passage from state *A* to state *B* caused by the experience of event *E*. We need to define the related episode for each individual *j*. Generally speaking, an episode starts in t_j^S (the point in time when the individual enters in state *A, i.e.* starts to be at risk of experiencing *E*) and ends in t_j^F (the point in time when *E* occurs or when the observation is censored). The list of events that cause transitions or censoring for each specific TR*X* is shown in table 4.
For a window of observation defined as the interval from $t_j^{WIN\_start}$ to $t_j^{INTERVIEW}$, we have that:

if $t_j^F <= t_j^{WIN\_start}$ episode is not considered
if $t_j^S > t_j^{WIN\_start}$ then $t_j^S = t_j^{WIN\_start}$

In other words, episodes are reduced to their intersection with the window of observation.

4.3 Time-varying variables

Life courses are usually segmented into domains of life that exist in parallel and generally interact (e.g., Blossfeld and Rohwer, 2002). One of the most interesting aspects of MAPLES is the opportunity to consider time-varying covariates, i.e. variables that identify changing status throughout life course. Each domain may be divided further into discrete states of existence. The set of possible states a person can occupy is known as the *state space*. The states are mutually exclusive (only one state can be occupied at a time) and exhaustive (at in any point of time each individuals must be in one of the states.

Given the dates included in the initial dataset it is possible to consider the following time varying covariates

Marital status (MAR):

a. Never married
b. First marriage
c. Second marriage.
d. Divorced or widowed

Own children ever born (CHI):

a. childless
b. one child
c. two children
d. three or more children.

Living arrangements (LIV):

a. Living in the parental home
b. Living alone or with other persons
c. With a partner

We assume that the parental home may be left only once (see, e.g., Matsuo and Willekens, 2003). We also suppose that a married individual lives with his/her partner.

Table 8. Combination of categories for each variable.

Available dates	Not available dates	Categories	Code	Number of categories
MAR (Marital Status)				
None	Fist marriage; divorce; death of spouse; second marriage	None	**0**	1
First marriage	Divorce; death of spouse; second marriage	1. Never married (a) 2.Ever married (b,c,d)	**1**	2
Fist marriage; divorce; death of spouse, second marriage	None	1.Never married (a) 2.First marriage (b) 3. Second marriage (c) 4. Divorced or widowed (d)	**2**	4
CHI (Own children ever born)				
None	First child; second child; third child	None	**0**	1
First child	Second; third child	1. Childless(a) 2. With children (b,c,d)	**1**	2
First child; second child	Third child	1. Childless (a) 2. One (b) 3. Two or more (c,d)	**2**	3
First child; second child; third child	None	1. Childless (a) 2. One (b) 3.Two (c) 4. Three or more (d)	**3**	4
LIV (Living arrangements)				
None	Exit, first union; union disruption; second union	None	**0**	1
Exit	First union; union disruption; second union	1. Parental home (a) 2. Exit (b,c)	**1**	2
Exit, first union	Union disruption; second union	1.Parental home (a) 2. Alone or with others (b) 3. First union (c)	**2**	3
Exit, first union; union disruption; second union	None	1. Parental home (a) 2. Alone or with others (b) 3. With a partner (c)	**3**	3
EDU (Education)				
None	Education	None	**0**	1
Education	-	1.Primary-Lower sec.(a,b) 2. Upper sec.-tertiary (c,d)	**1**	2
Education	-	1. Primary- Lower sec.(a,b) 2. Upper sec. (c) 3. Tertiary (d)	**2**	3
Education	-	1. Primary (a) 2. Lower sec.(b) 3. Upper sec.-tertiary (c,d)	**3**	3
Education	-	1. Primary (a) 2. Lower sec.(b) 3. Upper sec.(c) 4. Tertiary (d)	**4**	4

Each time-varying variable can be specified when a set of specific dates is available. Generally speaking, the level of detail of covariates depends on the dates included in the initial dataset. If all dates are collected, covariates have the maximum number of categories; otherwise, limited information implies one or more aggregations. For example, considering the variable CHI, if we have only the date of the first child's birth but not the date of following births, we can distinguish between the childless period and the period with one or more children in the biography of an individual but we cannot differentiate between category b, c, and d. In table 8, we show a set of rules for each time-varying variable: depending on available dates, we can have a certain number of categories each one coded as an integer from 0 to 3.
Besides, it is also useful to specify combinations of categories for the time-fixed variable EDU. We will understand why in section 5.6 where we will refer again to the scheme in table 8.
Some covariates make no sense for specific transitions. For example, in TR1, MAR is constantly "never married" by definition. Table 9 shows for each transition which time-varying covariates are allowed.

Table 9. Allowed covariates for each transition

TRANSITION	Allowed covariates
TR1 never-married → married (1st marriage)	EDU, LIV, CHI
TR2 married (1st marriage)→ divorced	EDU,CHI
TR3 married (1st marriage)→ widowed	EDU,CHI
TR4 divorced→ married (2nd marriage)	EDU, CHI
TR5 widowed→ married (2nd marriage)	EDU, CHI
TR6 at parental home (never in union) → first union	EDU, CHI*
TR7 at parental home→ alone/with others (never in union)	EDU, CHI*
TR8 alone/ with others (never in union) → first union	EDU, CHI*
TR9 first union→ separated (after 1st union disruption)	EDU, MAR, CHI,
TR10 alone or with other persons (after the 1st union disruption)→ with a partner (2nd union)	EDU, MAR,CHI
TR11 childless → child	EDU, MAR, LIV
TR12 1 child →2 children	EDU, MAR, LIV
TR13 2 children → 3 children	EDU, MAR, LIV
TR14 3 children → 4 children	EDU, MAR, LIV

* "Own children ever born" is always coded in only two categories: "childless/with children".

The effect of a time-varying variable is modelled by "splitting" the individual episode at the point the change occurs (see Blossfeld and Rohwer, 2002). Each sub-episode which results from splitting is then characterized by a unique value of this variable. For example, let us consider the transition TR2 and an episode that ends with the divorce event. This episode may be described with the vector $(t_j^S, t_j^F, 0, 1)$ where the first value is the starting time, the second is the final time and the last two values indicate that the *j*-th individual starts the episode with a status 0 (married, not divorced) and ends it with the status 1 (divorced). Now, we can imagine that at times t_j^{CH1} and t_j^{CH2} (where $t_j^S < t_j^{CH1} < t_j^{CH2} < t_j^F$) the *j*-th individual experienced respectively first and second child's birth. The original episode is thus split into the three sub-episodes: $(t_j^S, t_j^{CH1}, 0, 0)$, $(t_j^{CH1}, t_j^{CH2}, 0, 0)$, and

(t_j^{CH2}, t_j^F, 0, 1). The covariate CHI is fixed at 0 (childless) in the first sub-episode, at 1 in the second (1 child) and at 2 (2 children) in the third.
In general, an episode is split according to all possible covariates allowed for the specified transition.

4.4 Events and exposure time

For age x varying from 0 to 100, let E_x be the total number of events experienced by all the individuals at age x, and PY_x be their total duration of exposure at the same age. The *transition rate* at age x (r_x) is then the ratio between the number of events E_x and the amount of time spent in the initial state PY_x. In order to calculate it, we first compute E_x and PY_x for every age x.
Considering x_j^S and x_j^F the age at the beginning and at the end of the episode, we have seen in section 4.2 that the episode is included in the window of observation:

$$x_j^{WIN\_start} \le x_j^S \le x_j^F \le x_j^{INTERVIEW}$$

Let us call N the number of episodes. For the j-th individual we have:

$$E_x = \sum_{j=1}^{N} E_{j,x}$$

$$PY_x = \sum_{j=1}^{N} PY_{j,x}$$

and

$$E_{j,x} = \begin{cases} 1 & \text{if } x_j^S < x < x_j^F \text{ and the transition occured at age x} \\ 0 & \text{otherwise} \end{cases}$$

$$PY_{j,x} = \begin{cases} 1 & \text{if } x_j^S < x < x_j^F \text{ and neither transition nor exit from observation are experienced at the age x} \\ \delta_{j,x} & \text{if } x_j^S < x < x_j^F \text{ and transition or exit from observation occured at the age x} \\ 0 & \text{if } x < x_j^S \text{ or } x > x_j^F \end{cases}$$

where $\delta_{j,x}$ is the fraction of year spent in the initial state by the j-th individual at the exact age at which he experienced the event or the exit from observation.[1] For example, let us suppose that we are interested in the transition TR2 (married → divorced) and that the j-th individual's episode starts at age 41 and ends at age 42.31 with a divorce. At age 41, his contribution is 0 for event and 1 for exposure time. At age 42 he contributes with 1 event and with 0.31 years for exposure time.

[1] Fomulas and examples are valid not only for cohort-age rates but also for cohort-period rates under the assumption that the rate at time t (covering age x-1 and x) is referred to as rate at age x (see fig. 1).

We can also include individual post-stratification weights w_j in the computation. The formulas become:

$$E_x = \sum_{j=1}^{N} E_{j,x} \cdot w_j$$

$$PY_x = \sum_{j=1}^{N} PY_{j,x} \cdot w_j$$

In order to take into account categorical covariates, it is sufficient to count events and exposure time separately for any allowed combination of their levels of categories. In other words, we select sub-intervals (defined for each combination of covariates levels) to which the previous calculations apply. The resulting data matrix will have one row for each combination of age and levels of covariates. For example, if we consider an age range from 15 to 49 (35 age classes) and four covariates (EDU: 4 levels; MAR: 4 levels; CHI: 4 levels; LIV: 3 levels), the matrix will have a number of rows equal to

$$35 \cdot 4 \cdot 4 \cdot 4 \cdot 3 = 6720$$

Rows with a zero exposure time are dropped from the matrix.

5. GAMs and transition rates

In the final step, we start from the transition-specific data matrix. Now, each row constitutes a specific combination of events, exposure time and covariates (EDU, MAR, CHI, LIV). Transition rates are estimated by using Generalized Additive Models, which lead to both obtain a smoothed age profile and asses the (multiplicative) effect of one or more covariates. Moreover, MAPLES overcomes the proportional assumption by estimating covariate effects separately for different sub-interval of ages. Finally, smoothing and tail-flattening procedures ensure the continuity of the final age profile.

5.1 Generalized Additive Models

If we consider the transition rate for a specific event as the dependent variable, we should model it as a function of age and a set of covariates. However, age profiles for a specific transition should never be considered as a linear function. Smoothing or graduating rates, or more specifically the age profile of rates, has been a traditional issue in various disciplines, including demography and actuarial science. Traditional approaches based on polynomials have been criticized in the literature for a long time, authors proposing to use spline functions as a solution (see, e.g., McNeil et al., 1977); recent developments include Smith *et al*. (2004) and, on age-specific fertility rates, Schmertmann (2003).
For our purpose the so-called family of *Generalized Additive Models* is a suitable solution (Hastie & Tibshirani, 1990; Chambers & Hastie, 1992; Hastie *et al*, 2001). GAMs constitute a generalization of linear model where the dependent variable *Y* can be modeled as a sum of non-linear (smoother) functions.
The model structure is as follows:

$$g(\mu) = \beta_0 + f(age) + \sum_k \beta_k X_k \quad (1)$$

where $\mu = E(Y)$; $g(.)$ is the link function; Y is the response variable (distributed as some exponential distributions); X_k is a generic covariate and β_k the corresponding parameter; β_0 is the intercept; $f(age)$ is the smoothing function of age.
Since transition rates at age x for a specific event are given by the ratio between number of events (*Events*) and the time of exposure (*Exp.time*), considering natural logarithm as link function, for each i-th row of data matrix[2] we can write:

$$\ln\left(\frac{Events_i}{Exp.time_i}\right) = \beta_0 + f(age_i) + \sum_k \beta_k X_{ki} + \varepsilon_i$$

where ε_i is a random error term. Then,

$$\ln(Events_i) = \ln(Exp.time_i) + \beta_0 + f(age_i) + \sum_k \beta_k X_{ki} + \varepsilon_i$$

or, considering the expected value

$$\ln(E[Events]) = offset[\ln(Exp.time)] + \beta_0 + f(age) + \sum_k \beta_k X_{ki} \quad (2)$$

where

$$Events \sim Poisson$$

It is important to underline that the term $ln(Exp.time)$ has no coefficient to be estimated.
In our dataset, *Events* are calculated starting from individual weighted information. As a consequence, number of events and time of exposure are not integers. Since the *Poisson* distribution is defined only for integers, we need to round the number of weighted events. Empirical analyses (not shown here) suggest that this approximation appears acceptable.
The smoothing function f is a *piecewise cubic spline*, a curve made up of sections of cubic polynomial joined together so that they are continuous in value, as well as first and second derivatives. The points at which the sections join are known as the *knots* of the spline, that are placed at quantiles of the distribution of unique x values. The number of knots defines the *degree of smoothness* of the f (i.e. number of knots + 2). In order to avoid the choice of parameter, that is essentially arbitrary, the degree of smoothness is estimated by Generalized Cross Validation[3] (Wood, 2006).
Calling $\hat{y}$ the fitted values of this model, the transition rate is estimated as:

$$\hat{r} = \frac{\hat{y}}{Exp.time} \quad (4)$$

[2] We remember that each row of the data matrix is given by a specific combination of age x and the levels of categorical covariates.
[3] The way to control smoothness by altering the basis dimension, is to keep it fixed at a size a little larger than the one that could reasonably be necessary, but to control the model's smoothness by adding a "wiggliness" penalty to the least squares fitting objective (penalized regression spline) (Wood, 2006). The *mgcv* R package contains a GAM implementation in which the degree of smoothness of model terms is estimated as part of fitting.

5.2 Multiplicative effects of covariates

The effect of covariates is considered as a multiplicative change to be applied to the grand mean, i.e. to the mean risk for the whole sample. This is pursued by applying the "deviation coding" system that compares the mean of the dependent variable for a given level to the overall mean of the dependent variable. If we consider, for example, the categorical covariate *Education* with 4 levels, the deviation coding is accomplished by assigning value "1" to level 1 for the first comparison (because level 1 is the level to be compared to all others), to level 2 for the second comparison (because level 2 is to be compared to all others), and to level 3 for the third comparison (because level 3 is to be compared to all others). The value "-1" is assigned to level 4 for all three comparisons (because it is the level that is never compared to the other levels). The value "0" is assigned to all other levels (see table 10).
Given that the expected values of the dummies specified in such a way are always zero[4], we can obtain the baseline transition rate as:

$$baseline_i = e^{\beta_0 + f(age_i)}$$

The estimated coefficients related to covariates express multiplicative changes to be applied at the baseline age profile in order to evaluate the estimated risk for each year of age. In other words, the effect of a covariate can be seen as a vertical shift throughout the whole range of age. For example, figure 4 shows the multiplicative effect of the level of education for an unspecified transition.

Table 10. Deviation coding for level of education

EDU	Dummy 1 (Primary vs. mean)	Dummy 2 Low. sec. vs mean	Dummy 3 Upp. sec vs mean
Primary	1	0	0
Lower secondary	0	1	0
Upper secondary	0	0	1
Tertiary	-1	-1	-1

However, very often the effect of a covariate shows a combination of vertical and horizontal shifts. Therefore, the proportional assumption on the whole age rank appears too simplistic. MAPLES adopts the following solution: the multiplicative effect of a covariate is estimated separately within three different sub-intervals of ages delimited by two knots. These knots are fixed systematically at the 33rd and 67th percentiles of the event distribution (i.e. at the ages x_1 and x_2 at which, respectively, 33% and 67% of all the events are already experienced). In other words, each sub-interval contains one third of the total number of events.

With this new configuration, the model contains, other than the baseline transition rate, a set of dummy variables, one for each combination of independent variable and the three age sub-intervals. For example, the effect of education (with 4 levels) is estimated by including 12 dummy variables in the model equation.
In figure 5 we can see coefficient estimates and transition rates for the effect of EDU on a generic transition TR*X*. In this example, knots are computed at age 30 and 34.

[4] More precisely, the expected values are zero if the number of cases is (approximately) the same for each level. In our analysis this condition is satisfied given the structure of our data-matrix (similar number of rows for each combination of levels of covariates).

Fig. 4 Multiplicative effects of covariates estimated with additive model.

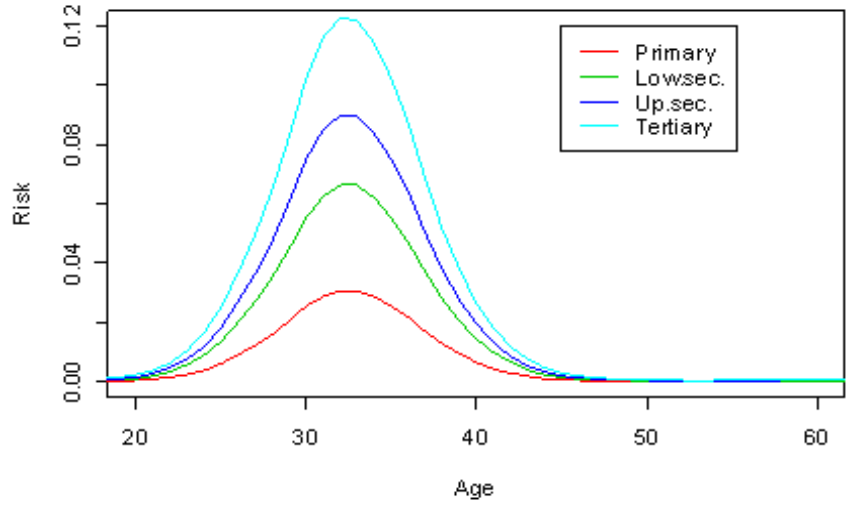
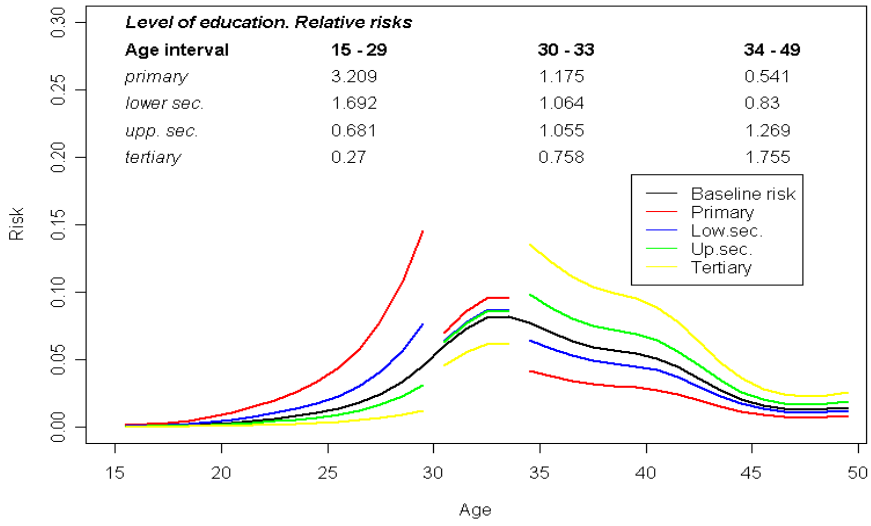


Fig. 5 Proportional effects of education on the transition TR1.



5.3 Smoothing procedure

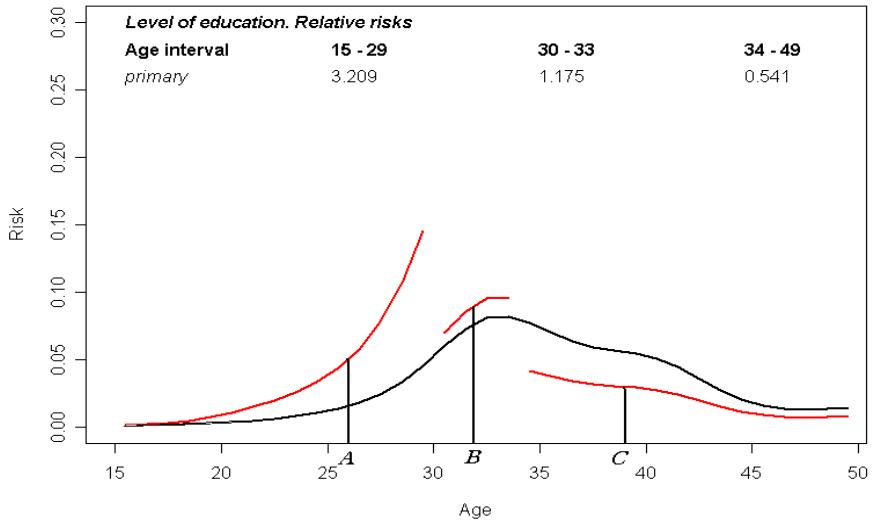
Transition rates for a sub-sample characterized by a specific value of covariates, result in non-continuous curves. The next step is the specification of a smoothing procedure that permits to obtain continuous curves which remain consistent with the estimated relative risks for each value of covariates and for each sub-interval of age.

In order to do so, let us consider the interval of age that starts at the midpoint of the first subinterval (point *A* in figure 6: age at which at least 16% of the events have been experienced) and ends at the middle point of the second sub-interval (point *B*: median age). At point *A*, the transition rate is the product of the baseline at *A* by the relative risk associated to the covariate level for the first sub-interval (β_1). At point *B* the transition rate is the baseline at *B* multiplied by the relative risk for the second sub-interval (β_2). When we proceed over the age axis from *A* to *B*, the continuous transition rate is obtained by multiplying the baseline by a weighted means of β_1 and β_2. The weight of β_1 is decreasing from a value close to 1 (at *A*) to a value close to 0 (at *B*) whereas the weight of β_2 is

increasing in the opposite way. The trend of weights is not linear but follows a logistic curve. The same procedure can be applied to the interval (*B*, *C*).
In the example of figure 6 we focus on the effect of primary level of education on transition's risks (all the other effects are not shown). Age at point *A* is 26 years, and 28 years at point *B*. We have that:

$$r_A = baseline_A \cdot 3.209$$
$$r_B = baseline_B \cdot 1.175$$

Fig. 6. Smoothing procedure. Mid-points fixed according to 16th, 50th and 84th percentiles.



For each age $x \in (A, B)$ the transition rate is

$$r_x = baseline_x * \left(3.209 * (1 - wgt_x) + 1.175 * (wgt_x)\right)$$

Weights *wgt* follow a logistic curve and they are computed as follow:

$$wgt_x = \frac{1}{1 + Ke^{-h \cdot (x - A)}} \quad \text{for } x = (A+1) \text{ to } (B-1)$$

and

$$K = e^{\frac{(B-A)*h}{2}}$$

where *h* is the growth rate and it is computed as

$$h = e^{\frac{5}{B-A}}$$

This means that *h* is equal to 1 when the length of the interval (*A,B*) is 5. In this way, the shape of the logistic curve remains the same independently from the interval's length (see fig. 7).

The same procedure could be applied to the second jump of the transition rate curve. Focusing on points *B* and *C* we have

$$r_B = baseline_B \cdot 1.175$$
$$r_C = baseline_C \cdot 0.541$$

Fig 7. Logistic curve with h=1 showing weights for a specific point x

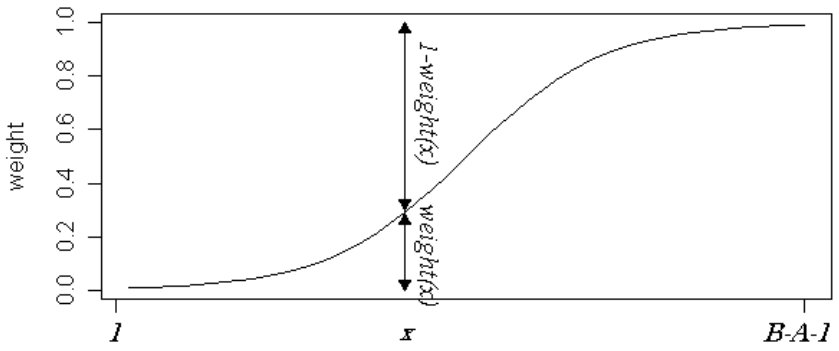
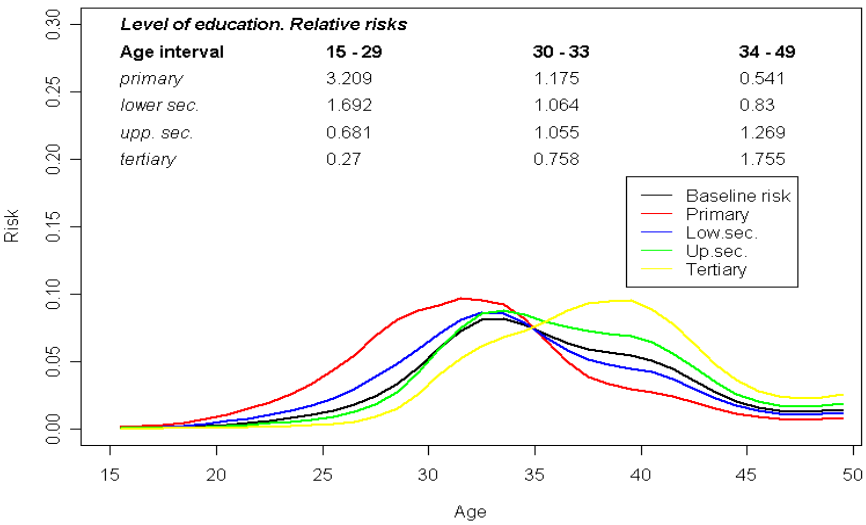


Fig. 8 Smoothed curves



For each age $x \in (B, C)$ the transition rate is

$$r_x = baseline_x * \left(1.175 * (1 - wgt_x) + 0.541 * (wgt_x)\right)$$

where

$$wgt_x = \frac{1}{1 + Ke^{-h\cdot(x-B)}} \quad \text{for } x = (B+1) \text{ to } (C-1)$$

and

$$K = e^{\frac{(B-A)*h}{2}}.$$

This procedure may be repeated for all levels of covariates. The resulting smoothed curves appear as in figure 8.

5.4 Tail-flattening procedure

Age profiles are obtained for the age interval (age_{min}, age_{max}). Working with retrospective data we can fix the limits respectively to 15 and 100 but we must face with non-zero exposure time age classes. Besides, data may be not available for the older classes when in a survey individuals older than a certain age have not been interviewed. Generally speaking we can define:

age_{min}: the lowest available age (≥15) with non zero exposure time
age_{max}: the highest available age (≤100) with non zero exposure time

By applying MAPLES procedure, we may have non-zero baseline risk at age_{min} and/or at age_{max}. Thus, if we can assume that outside (age_{min}, age_{max}) the baseline is zero, there are two jumps in the edges of the specified age interval. For example, in figure 9 we see two discontinuities in age_{min}=20 and age_{max}=63. MAPLES can avoid this situation by "flattening" the risk in the tails of the age profiles through the application of logistic weights to the baseline. Let us call D the age at which 5% of the events have been experienced, in the left tail the weights are:

$$weight_x = \frac{1}{1 + Ke^{-h\cdot(x-age_{\min})}} \quad \text{for } x = age_{\min} \text{ to } (D-1)$$

where

$$K = e^{\frac{(D-age_{\min})*h}{2}}$$

and h is fixed arbitrarily to

$$h = e^{\frac{10}{B-A}}$$

If E is the age at which 95% of the events have been experienced, in the interval right tail the weight to be applied to baseline are:

$$weight_x = 1 - \frac{1}{1 + Ke^{-h\cdot(age_{\max}-x)}} \quad \text{for } x = (E+1) \text{ to } age_{\max}$$

In fig. 9 we can see the resulting flattened tails (in red). This procedure could also be useful in order to avoid odd values caused by few events.
Given that the effect of covariate is a multiplicative change to be applied to the baseline, age profiles for a specific level of one covariate will automatically flattened in the tails.

When the hypothesis that transition rates are zero outside the interval (age_{min}, age_{max}) is not applicable, it has no sense to flatten age profiles in one or in both tails. For example, for transition TR3 (1st marriage→death of spouse) rates are not decreasing at the older ages. In general, the tail-flattening procedure is optional and it can be excluded in one of the two tails or in both of them. In figure 10, the baseline risk has flattened left tail and non-flattened right tail. In this case, we do not know transition rates at the right of age_{max} and we need to complete the shape of age profiles for older ages using, for example, extrapolation methods.

Fig.9 Baseline with discontinuities (black line) and with flattened tails (red line). Rates are assumed to be zero outside the interval (*20, 63*).

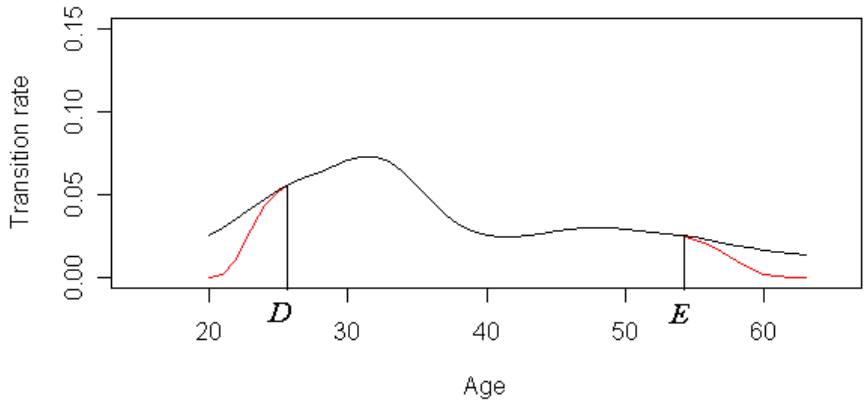
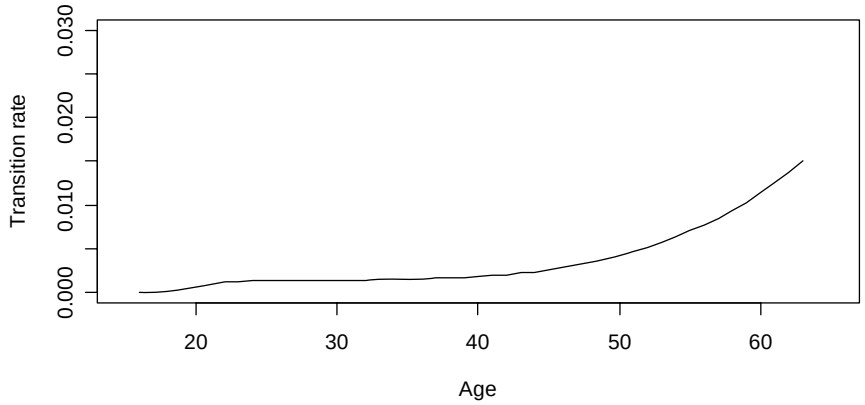


Fig.10 Baseline with flattened left and non-flattened right tail.



5.5 The MAPLES output

Estimates are calculated separately for men and women. Moreover, our approach is to introduce covariates (EDU, MAR, CHI, LIV) one by one under the independent hypothesis between couples of covariates. A single covariate, however, implies that in model equation (2) the number of included dummy variables (k) is equal to the number of levels multiplied by 3 (number of age subintervals).
The final stage of the procedure provides:

- a vector of ages within a selected age range;

- a vector containing the baseline age profile. It is obtained through the estimation of the baseline in a model without covariates (i.e. by fixing k=0 in the model equation 2);

- a set of vectors containing relative risks for each year of age and for each level of allowed covariates. For example, transition TR2 has a maximum of 4 vectors for EDU (one for each level of education) and a maximum of 4 vectors for CHI (one for each number of children ever born). Data availability and low numbers of events may group levels and, therefore, may reduce the number of vectors. Relative risks for each year of age are computed as ratio between the smoothed transition rate (see section 2.3.4) and the relative baseline rate. An example, concerning transition TR1 in Italy, is reported in table 11.

As additional feature, MAPLES tests the statistical significance of the additional covariate X in the model by dropping it and noting the change in the deviance. The fitted models are compared using an analysis of deviance table. The tests are usually approximated, unless the models are un-penalized (Wood, 2006). Therefore, for each variable we have a *pvalue* relating to the comparison between base model (without covariates) and model with covariate X.

Table 11. Baseline and relative risks (TR1. Fss Italy 2003. Women)

age	baselin	prim	lowsec	uppsec	tert	noch	1+ch	par_hom	no_part	partner
15	3e-04	2.4896	1.7982	0.6402	0.3489	0.5977	1.6731	0.8793	0.5446	2.0882
16	9e-04	2.4896	1.7982	0.6402	0.3489	0.5977	1.6731	0.8793	0.5446	2.0882
17	0.0021	2.4896	1.7982	0.6402	0.3489	0.5977	1.6731	0.8793	0.5446	2.0882
18	0.0048	2.4896	1.7982	0.6402	0.3489	0.5977	1.6731	0.8793	0.5446	2.0882
19	0.0092	2.4896	1.7982	0.6402	0.3489	0.5977	1.6731	0.8793	0.5446	2.0882
20	0.0153	2.4896	1.7982	0.6402	0.3489	0.5977	1.6731	0.8793	0.5446	2.0882
21	0.0228	2.4896	1.7982	0.6402	0.3489	0.5977	1.6731	0.8793	0.5446	2.0882
22	0.0316	2.4019	1.7521	0.6601	0.3754	0.623	1.6276	0.9045	0.5476	2.0369
23	0.0423	2.2787	1.6875	0.688	0.4125	0.6585	1.5638	0.94	0.5518	1.9649
24	0.0555	2.0532	1.5691	0.7391	0.4806	0.7235	1.447	1.0049	0.5595	1.833
25	0.0711	1.77	1.4205	0.8032	0.566	0.8051	1.3003	1.0865	0.5692	1.6675
26	0.0873	1.5445	1.3022	0.8543	0.6341	0.8701	1.1834	1.1514	0.5769	1.5357
27	0.101	1.4213	1.2375	0.8822	0.6713	0.9056	1.1196	1.1869	0.5811	1.4637
28	0.1096	1.3336	1.1915	0.902	0.6977	0.9309	1.0742	1.2121	0.5841	1.4124
29	0.1127	1.2926	1.1683	0.9305	0.7293	0.9323	1.0727	1.2064	0.5923	1.4016
30	0.1114	1.235	1.1357	0.9705	0.7738	0.9341	1.0706	1.1983	0.6039	1.3865
31	0.1071	1.1295	1.076	1.0436	0.8552	0.9376	1.0667	1.1835	0.6251	1.3589
32	0.1001	0.9971	1.0011	1.1355	0.9573	0.9419	1.0618	1.1649	0.6517	1.3242
33	0.0901	0.8917	0.9415	1.2087	1.0387	0.9453	1.0579	1.1501	0.6728	1.2965
34	0.0771	0.8341	0.9089	1.2487	1.0831	0.9472	1.0558	1.142	0.6844	1.2814
35	0.0626	0.7931	0.8857	1.2771	1.1148	0.9485	1.0543	1.1362	0.6926	1.2707
36	0.049	0.7931	0.8857	1.2771	1.1148	0.9485	1.0543	1.1362	0.6926	1.2707
37	0.0382	0.7931	0.8857	1.2771	1.1148	0.9485	1.0543	1.1362	0.6926	1.2707
38	0.0309	0.7931	0.8857	1.2771	1.1148	0.9485	1.0543	1.1362	0.6926	1.2707
39	0.0269	0.7931	0.8857	1.2771	1.1148	0.9485	1.0543	1.1362	0.6926	1.2707
40	0.0253	0.7931	0.8857	1.2771	1.1148	0.9485	1.0543	1.1362	0.6926	1.2707
41	0.0255	0.7931	0.8857	1.2771	1.1148	0.9485	1.0543	1.1362	0.6926	1.2707
42	0.0261	0.7931	0.8857	1.2771	1.1148	0.9485	1.0543	1.1362	0.6926	1.2707
43	0.0258	0.7931	0.8857	1.2771	1.1148	0.9485	1.0543	1.1362	0.6926	1.2707
44	0.0231	0.7931	0.8857	1.2771	1.1148	0.9485	1.0543	1.1362	0.6926	1.2707
45	0.0182	0.7931	0.8857	1.2771	1.1148	0.9485	1.0543	1.1362	0.6926	1.2707
46	0.0123	0.7931	0.8857	1.2771	1.1148	0.9485	1.0543	1.1362	0.6926	1.2707
47	0.0074	0.7931	0.8857	1.2771	1.1148	0.9485	1.0543	1.1362	0.6926	1.2707
48	0.004	0.7931	0.8857	1.2771	1.1148	0.9485	1.0543	1.1362	0.6926	1.2707
49	0.0021	0.7931	0.8857	1.2771	1.1148	0.9485	1.0543	1.1362	0.6926	1.2707

5.6 Rare transitions or few events in the dataset

Our effort in developing MAPLES is to take into account the greatest possible number of transitions. However, it is possible that for some transitions, the number of events is very low. This is more frequent when the dataset has a limited number of cases. With a small number of events, it is possible that relative risks could be biased. In order to control for such situations, we introduced specific limitations. We may specify a number *nmin* in such a way that, for a generic transition TR*X* and for each sex:

1. the total number of events must be higher than *nmin*.

2. the number of events for each level *k* of a given covariate *X* and within each subinterval of age (defined by knots and containing at least one third of the total number of events) must be higher than *nmin/3* (this implies that the number of events for level *k* must be higher than *nmin*). In other words, considering a covariate *X* with *K* levels, we impose that

$$\text{events}(X = i \text{ and age.int} = j) >= nmin \text{ for } i = 1..K \text{ and } J = 1..3$$

The most extreme case is when condition 1 is not fulfilled: estimates can be hardly considered reliable. In this situation a possible solution may be to extend the observation window. When we increase the window length, we take into account longer segments of life and, therefore, more events. However, we refer to behaviors that are experienced in a more distant past.
When the total number of events is sufficiently high but for level *k* of covariate *X* the condition 2 is not satisfied, the strategy is to group level *k* with a nearby level. If condition 2 is still not satisfied when only two levels remain for *X*, the covariate is excluded from the analysis.
As an example, let us consider the situation depicted in table 12 (transition TR1, women). In case a. we have only level of education (EDU). This means that we do not have enough information to consider CHI whereas LIV is not allowed in TR1. If we fix *nmin*=30, condition 1 is satisfied given that the total amount of events are 297, but condition 2 is not satisfied for level "tertiary" because in the first age interval we have only 4 events (<*nmin*/3=10). This level is grouped with "uppsec" and variable EDU will have 3 levels ("primary", "lower secondary", and level 3 "uppsec+" given by the aggregation of "upper secondary" and "tertiary"). With this new configuration, condition 2 is always fulfilled.

Table 12. Number of events for the transition TR1.

Case a.

Number of events - WOMEN				
	int1	int2	int3	tot
prim	11	28	23	62
lowsec	43	57	30	130
uppsec	12	33	18	63
tert	4	20	18	42

Case b.

Number of events - WOMEN				
	int1	int2	int3	TOT
prim	10	6	6	22
lowsec	2	4	5	11
uppsec	7	6	0	13
tert	2	0	4	6
no ch	2	2	6	10
1+ ch	13	16	13	42

In case *b*, with the same value of *nmin*=30 condition 1 is satisfied (52 events) but both covariates (EDU and CHI) are excluded because there no enough events to justify at least two levels. In fact, for EDU if we group "tert" and "uppsec" we have 9, 6 and 4 events respectively in the first, second and third subinterval, condition 2 is never satisfied; if we group "tert","uppsec and "lowsec"

together, condition 2 is not satisfied in the third interval. For CHI, condition 2 is not fulfilled for level "no ch".
The progressive aggregation of levels within each covariates follow the scheme presented in table 8: when condition 2 is not satisfied, MAPLES reduces the code number by one. As a consequence, we may have covariates with 2, 3, or 4 levels.
These checks on number of events permit to avoid age profiles estimated from very few events, which gives stability to the estimates.

5.7 Extension: combination of covariates

Let us call $r(x, c_1, c_2, c_3)$ the transition rate at age x for a specific transition TRX, for individuals with values c_1, c_2, c_3 respectively for covariates C_1, C_2, C_3.

Through the application of MAPLES we have for a given transition TRX and a given sex, a set of relative risks for each age x and for each level of covariates:

$$\begin{aligned}
&rrisk_{edu}(x, l_{edu}) && x:\ \min(x)\ to\ \max(x) && and\ l_{edu} = 1\ to\ N_{edu}\\
&rrisk_{mar}(x, l_{mar}) && x:\ \min(x)\ to\ \max(x) && and\ l_{mar} = 1\ to\ N_{mar}\\
&rrisk_{chi}(x, l_{chi}) && x:\ \min(x)\ to\ \max(x) && and\ l_{chi} = 1\ to\ N_{chi}\\
&rrisk_{liv}(x, l_{liv}) && x:\ \min(x)\ to\ \max(x) && and\ l_{liv} = 1\ to\ N_{liv}
\end{aligned}$$

where N_{edu}, N_{mar}, N_{chi} and N_{liv} are the number of level of the relative covariate.
These four sets are estimated in separated models. Given the hypothesis of independence between covariates, we can easily compute an age profile for every combination of levels of covariates. Generally speaking, transition rate at age x for individuals characterized by a specific combination is:

$$r(x, l_{edu}, l_{mar}, l_{chi}, l_{liv}) = baseline(x) \cdot rrisk(x, l_{edu}) \cdot rrisk(x, l_{mar}) \cdot rrisk(x, l_{chi}) \cdot rrisk(x, l_{liv})$$

where $baseline(x)$ is the baseline transition rate estimated in the model without covariates and expresses the grand mean of transition rates for all possible combinations of covariates.
For example, if we consider transition TR1 we have three possible covariates (EDU, CHI, and LIV) and a number of combination equal to 4*4*3=48 for each age x in the age range (e.g. from 15 to 49). Let us suppose that we want to know transition rates to first marriage for women aged x=30 with a tertiary level of education, childless and cohabiting. The application of MAPLES gives us the following rates (see section 5):

$$\begin{aligned}
&baseline(x) = 0.1114\\
&rrisk_{edu}(x = 30, l_{edu} = "tertiary") = 0.7738\\
&rrisk_{chi}(x = 30, l_{chi} = "noch") = 0.9341\\
&rrisk_{liv}(x = 30, l_{liv} = "partner") = 1.3865
\end{aligned}$$

The required rate is, then:

$$\begin{aligned}
r(x = 30, l_{edu} = "tertiary", l_{chi} = "noch", l_{liv} = "partner") &=\\
= 0.1114 \cdot 0.7738 \cdot 1 \cdot 0.9341 \cdot 1.3865 &= 0.1116419
\end{aligned}$$

6. An application to Italy

We conclude the paper by describing an application to real data from Italy. These data come from the multipurpose survey called "Famiglia e soggetti sociali (FSS-IT)", the survey associated with the Generations and Gender Programme (Vikat et al., 2007). Carried out at the end of 2003, this survey contains wide retrospective information on life course trajectories and transition to adulthood, including data on the history of marital unions, cohabitations (followed by a marriage or not) and marital disruption, for a large sample of the resident population. The retrospective nature of the survey makes it possible to update the collected information and to follow the same individual over time.

Table 13. Missing data in FSS-IT

```
> chkfile("ITALY.dat")
[1] ____________________________________________
[1] Check available data
[1] WARNING:mdiv missing
[1] WARNING:mved missing
[1] WARNING:meit missing
[1] WARNING:ydiss missing
[1] WARNING:mdiss missing
[1] ____________________________________________
```

Table 14 Transitions that can be analyzed with Italy FSS-IT

TR1 never-married → married (1st marriage)
TR2 married (1st marriage)→ divorced
TR3 married (1st marriage)→ widowed
TR4 divorced→ married (2nd marriage)
TR5 widowed→ married (2nd marriage)
TR6 at parental home (never in union) → first union
TR7 at parental home→ alone/with others (never in union)
TR8 alone/ with others (never in union) → first union
~~**TR9** first union→ separated (after 1st union disruption)~~
~~**TR10** alone or with other persons (after the 1st union disruption)→ with a partner (2nd union)~~
TR11 childless → child
TR12 1 child →2 children
TR13 2 children → 3 children
TR14 3 children → 4 children

Table 15. Consistency check on FSS-IT.

```
> consistency("ITALY.dat")
[1] Consistency check. File: ITALY.dat
[1] ____________________________________________
[1] 1st union<birth+14 - noc: 8
[1] 1st marriage<birth +14 - noc: 13
[1] divorce<=1st marriage - noc: 1
```

```
[1] death of spouse<=marriage - noc: 18
[1] 2nd union<=first union - noc: 2
[1] 1st child<birth + 14 - noc: 26
[1] 2nd child<1st child -  noc: 18
[1] 3rd child<2nd child - noc: 11
[1] 4th child<3rd child - noc: 5
[1] ____________________________________________________
```

In the Italian dataset, the main limit is the lack of dates relating to first union disruption (ydiss and mdiss) (table 13). This implies that TR9 and TR10 cannot be analyzed (see table 14). Consistency check (table 15) shows a lower number of inconsistent cases.

As an example of application, we consider transition TR1 (bever married->1st marriage) and TR12 (1st birth->2nd birth). The estimated age profles for women are plotted in fig 11. The window of observation is fixed at 5 years before the interview for both the transitions. The minimum number of events (*nmin*) is equal to 30 for TR1. This constraint reduces the number of categories for variable CHI to two categories (childless and with one or more children) whereas EDU and LIV have the maximum number of categories, respectively 4 and 3. For TR12 *nmin* is fixed at 10 in order to maintain a comparable number of categories. Nevertheless, upper secondary and tertiary level of education are grouped together as well as all the categories that refers to ever married women (first marriage, second marriage, and divorced/widowed)

As far first marriage is concerned, we can see that in Italy, the higher risk is experienced by women around 30 years of age and a low level of education (primary or lower secondary) tends to increase transition rates at lower ages. After 35 years of ages the effect of education reduces substantially. Moreover, women with a partner and/or with children show higher risks of marriage especially between 20 and 30 years of age.
Transition rates for the second child show increases rapidly after the 20th birthday and they remains high up to 35-36 years of age. The effects of covariates show a clear anticipation for the calendar of second births among women with a primary level of education and a strong negative effect given by the condition “not-married” on second birth rates, which confirms the low diffusion of births out-of-wedlock in Italy.

Fig.11a Estimated age profiles for Italy (survey held in 2003). Transition TR1 (never married->1st marriage) according to level of education, children ever born and living arrangements.

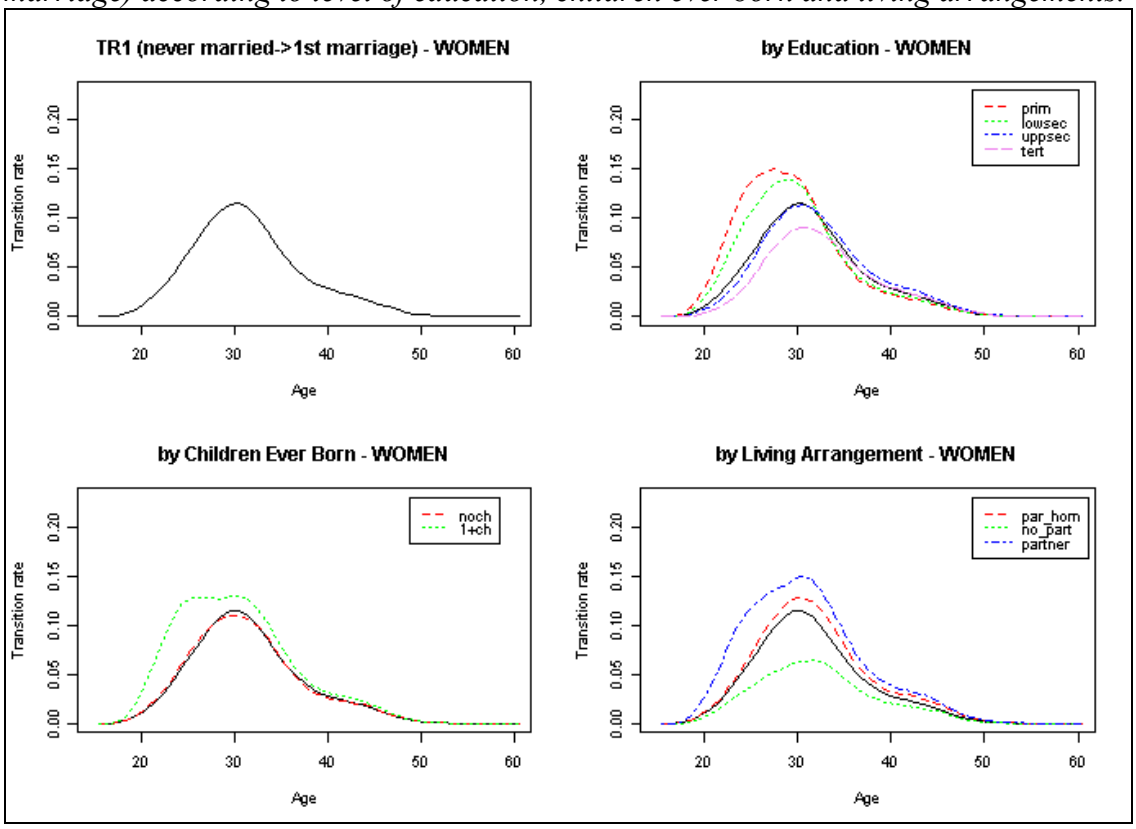
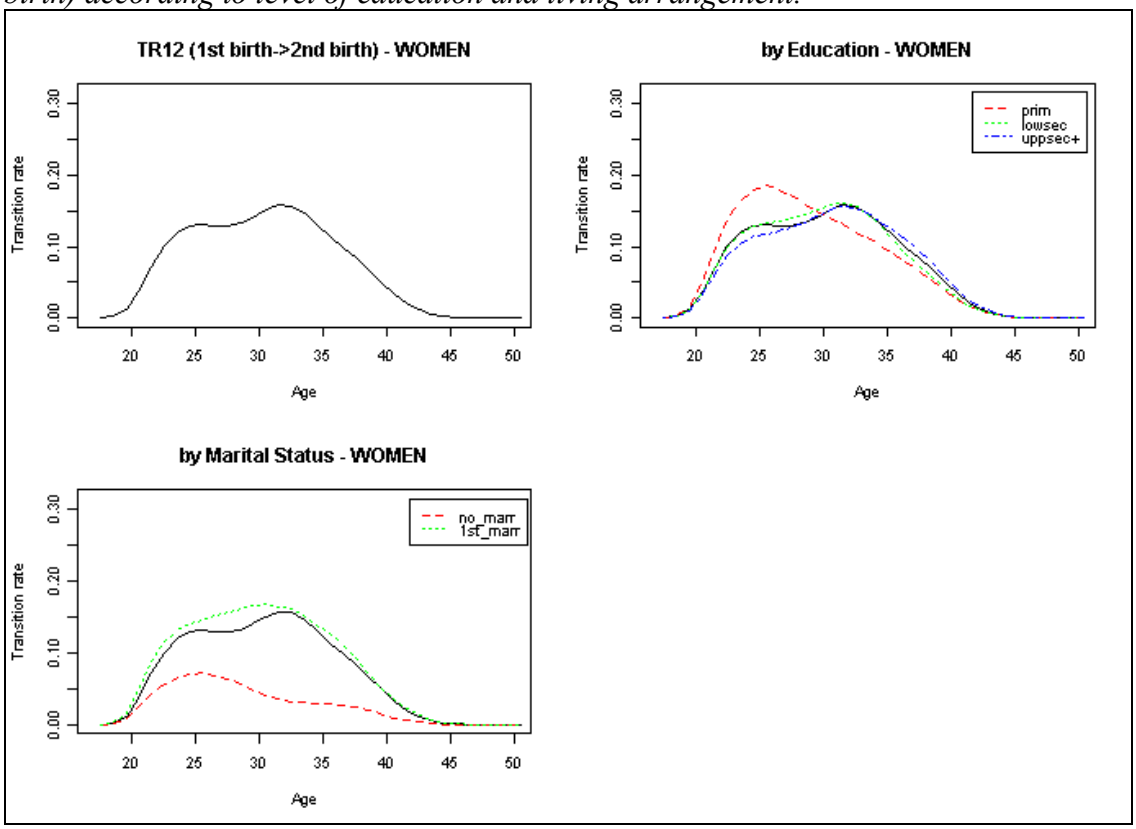


Fig.11a Estimated age profiles for Italy (survey held in 2003). Transition TR12 (1st birth → 2nd birth) according to level of education and living arrangement.



References

Andersen P.K., Borgan Ø., Gill R.D., Keiding N. 1993. *Statistical Models Based on Counting Processes*. Berlin-New York: Springer.

Blossfeld H.P., Rohwer G. 2002. *Techniques of Event History Modelling*. Mahwah, NJ: Lawrence Erlbaum Associates.

Chambers J.M. and Hastie T.J. (eds.), 1992, *Statistical models in S*, New York: Chapman and Hall.

de Beer J., van der Gaag N., Willekens F., 2006, *Report on Input Data Requirements of MAC*, Deliverable D2, MicMac Project "Bridging the micro-macro gap in population forecasting"

Hastie T.J and Tibshirani R.J., 1990, *Generalized additive models*, London: Chapman and Hall.

Hastie T.J., Tibshirani R.J. and Friedman J., 2001, *The elements of statistical learning. Data mining, inference and prediction*, New York: Springer-Verlag.

Matsuo H. and Willekens F., 2003, *Event histories in the Netherlands Fertility and Family survey, 1998. A technical report*. Population Research Centre, Research Report 03-1, February 2003.

McNeil D.R., Trussell T.J., Turner J.C., 1977, "Spline interpolation of demographic data", *Demography*, 14 (2): 245-252.

Schmertmann C., 2003, "A system of model fertility schedules with graphically intuitive parameters", *Demographic Research*, 9 (5): 81-110.

Smith L., Hyndman R.J., Wood S.N., 2004, "Spline interpolation for demographic variables: the monotonicity problem", *Journal of Population Research*, 21(1): 95-98.

van der Gaag N., de Beer J., Ekamper P., Willekens F., 2006, *Using MicMac to project living arrangements: an illustration of biographic projections*, paper presented at the European Population Conference, Liverpool.

Venables W.N. and Ripley B.D., 1997, *Modern applied statistics with S-plus*, New York: Springer.

Vikat A., Spéder Z., Beets G., Billari F.C., Buhler C., Désesquelles A., Fokkema T., Hoem J., MacDonald A., Neyer G., Pailhé A., Pinnelli A. and Solaz A., 2007. "Generations and Gender Survey (GGS): Toward a better understanding of relationship and processes in the life course", *Demographic research*, 17 (14): 389-440.

Wachter K.W., Blackwell D., Hammel E.A., 1998, *Testing the Validity of Kinship Microsimulation: An Update,* University of California, Berkeley, CA.

Willekens F., 2005, "Biographic forecasting: bridging the micro-macro gap in population forecasting". *New Zealand Population Review* 31 (1): 77-124.

Matsuo, H. & F. Willekens, 2003, *Event histories in the Netherlands Fertility and Family Survey 1998: A technical report. PRC Research Report 2003-1*, Groningen: Population Research Centre, University of Groningen.

Wood S.N, 2006, *Generalize Additive Models. An introduction with R*, Chapman and Hall/CRC

Appendix

BOX 1. Rules for the preparation of the initial dataset

1. The record structure of the data file must be as shown in table 4. The same variable names specified in this table must be used. Other names are not recognized and the meaning of the dates strictly follows the indication given in table 4. However, the order of variables is not important (variables could be sorted in a different way).

2. The initial dataset must contain id, weight, date of birth (*ybirth*, *mbirth*), date of interview (*yint*, *mint*), sex and education. All other dates are optional.

3. When the individual has not experienced an event, the date (*year*, *month*) must be coded as empty cells (blank). Other codes like "na", "999999", "mv", etc are not accepted. R will read empty cells as "Not available" information and it will call them as "NA" in the internal dataset.

4. A missing year means that the related event has not been experienced. If the individual has experienced a specific event but the year is not available, the case must be dropped from the initial dataset.

5. Dates may contain missing months (totally or partially). In that case, virtual months are computed as random numbers with the constraints specified in table 6.

BOX 2 The `chkfile` utility

The utility **`chkfile()`** provides basic information for the specified dataset. In particular, it specifies missing months and missing years.
The syntax is:

```
chkfile(filename)
```

The application to the Italian dataset gives the following output

```
> chkfile("ITALY.dat")
[1] ____________________________________________
[1] Check available data
[1] WARNING:mdiv missing
[1] WARNING:mved missing
[1] WARNING:mexit missing
[1] WARNING:ydiss missing
[1] WARNING:mdiss missing
[1] ____________________________________________
```

Given this output and referring to table 5, we are able to identify which transition can be studied.

BOX 3 The `consistency` utility

The utility **`consistency()`** is a tool included in MAPLES library that executes all the consistency checks presented in table 7 for a specified data file. It permits the user to take a first glance to the quality of the initial dataset. The syntax is

```
consistency(filename,showid=T)
```

Option `filename` Input datafile (with path)

The application of the utility *consistency* to the Italian dataset FFS 2003 called *ITALY.dat* shows the following output (noc means "number of cases"):

```
> consistency("ITALY.dat")
[1] Consistency check. File: ITALY.dat
[1] ____________________________________________
[1] 1st union<birth+14 - noc: 8
[1] 1st marriage<birth +14 - noc: 13
[1] divorce<=1st marriage - noc: 1
[1] death of spouse<=marriage - noc: 18
[1] 2nd union<=first union - noc: 2
[1] 1st child<birth + 14 - noc: 26
[1] 2nd child<1st child -  noc: 18
[1] 3rd child<2nd child - noc: 11
[1] 4th child<3rd child - noc: 5
[1] ____________________________________________
```

Option `showid=T` shows the ID (identification number of inconsistent cases. This could help the user to take a look at the original dataset. The output becomes:

```
> consistency("ITALY.dat",showid=T)
[1] Consistency check. File: ITALY.dat
[1] ____________________________________________
[1] 1st union<birth+14 - noc: 8
[1] Cases ID: 54301 292501 394902 609301 731601 1081202  1512301 1689401
[1] 1st marriage<birth +14 - noc: 13
[1] Cases ID: 54301  174901  92501  311501 394902  480702 609301  731601
[10] 1081202 1512301 1689401 1837801 1876001
[1] divorce<=1st marriage - noc: 1
[1] Cases ID: 901903
[1] death of spouse<=marriage - noc: 18
[1] Cases ID: 4901   339401 347101 613003 616501 723604  907002 959301
[10] 990701  1112901 1189801  1418001 1452901 1481901  1605601  1605602 1792803
[19] 1802501
[1] 2nd union<=first union - noc: 2
[1] Cases ID: 486401   1876001
[1] 1st child<birth + 14 - noc: 26
[1] Cases ID: 54301  75202 122201 150804  164902 312102  330602  480702
[10] 486202  731601 757502  774202  979301   1081202  1223202   1252402
1452602
[19] 1467402  1480401  1512301  1592802  1689401 1713702 1836802 1876001
1902402
[1] 2nd child<1st child -  noc: 18
[1] Cases ID: 39601  390001  462002 578001  609301 609302  724201 726702
[10] 963801  963802 1001501 1001502 1417401 1417402 1497002 1540301 1540302
[19] 1550302
[1] 3rd child<2nd child - noc: 11
[1] Cases ID: 90001 322402 390001  609301 609302 963801  963802  1001501
[10] 1001502  1550301 1897001
[1] 4th child<3rd child - noc: 5
[1] Cases ID: 143401  322401  528002  609301  609302
```

BOX 4. The `dataset` function

The function **`dataset()`** prepares data for the estimation of age profiles. In details, it loads the initial data set (filename), checks for missing dates, detects inconsistent dates, inputs missing months and computes decimal dates, ages and status variables. The output is a data frame that is ready to be processed by the function `ageprof()` (see below)
The syntax is

```
dataset(filename)
```

An example of application:

```
> d<-dataset("ITALY.dat")
[1] Dataset extracted from ITALY.dat is ready.
```

BOX 5. The `ageprof()` function.

`ageprof()` is the main function in the MAPLES package. It computes age profiles for a specific transition and for a given data.frame.

The syntax is

```
ageprof(d,tr,wl=5,minage=15, maxage=100,cpa=T,outf=F,nmin=30
```

Arguments are:

Option `d`	`d` is the data.frame containing initial data (No default value). It must be prepared through the function `dataset()`
Option `tr`	Specifies which transition have to be studied. Allowed values: integer from 1 to 14 (No default value).
Option `wl`	Specifies the length of the observation window (number of years before the interview). Only events and exposure times referring to this window will be considered in the analysis. Allowed values: integer from 3 to 30 (default = 5)
Option `minage`	Defines the lower limit of age range to be considered (default = 15)
Option `maxage`	Defines the upper limit of age range to be considered (default = 100)
Option `outfile`	Creates a text file containing all the standard output (No default value).
Option `cpa`	Specifies the kind of transition rates. If `TRUE` cohort-period transition rates are computed; otherwise ageprof() computes cohort age transition rates (default = `TRUE`)
Option `outf`	`If TRUE,` creates a text file with standard output (default = `TRUE`)
Option `nmin`	Specifies the minimum number of events for each level of covariates (to be considered separately for each sex): a number of events lower than `nmin` causes the aggregation of proximate levels. The same effect is done if for each level and in each subinterval of age (defined by knots and containing at least one third of the total number of events) we have lower than `nmin/3` events. If the total amount of events (independently by levels of covariates) is lower than `nmin` a warning message appears in the standard output.

Option `lft`	If `TRUE` age profiles are flattened in the left tail of the age interval, i.e. before the age at which 5% of the events have been experienced (default= `TRUE`)
Option `rgt`	If `TRUE` age profiles are flattened in the right tail of the age interval, i.e. after the age at which 95% of the events have been experienced (default= `TRUE`)

The standard output of `ageprof()` is a list containing the following objects:

`$name`	String containing the name of the considered transition
`$minage`	value of parameter minage
`$maxage`	value of parameter maxage
`$outf`	value of parameter outf
`$cpa`	value of parameter cpa
`$nmin`	value of parameter nmin
`$lft`	value of parameter left
`$rgt`	value of parameter right
`$knot_m`	Vector of 2 elements containing knots (Men).
`$cov_m`	Matrix containing information about covariates (code as defined in table 8; number of allowed levels; pvalue: anova test that compares model without covariates and model with the specified covariate) (Men).
`$numev_m`	Matrix with number of events according to age sub-intervals and covariate categories (Men).
`$rrisk_f`	Baseline transition rates and relative risks for each levels of allowed covariates (Women).
`$knot_f`	Vector of 2 elements containing knots (Women).
`$cov_f`	Matrix containing information about covariates (code as defined in table 8; number of allowed levels; pvalue: anova test that compares model without covariates and model with the specified covariate) (Women).
`$numev_f`	Matrix with number of events according to age sub-intervals and covariate categories (Women).
`$rrisk_f`	Baseline transition rates and relative risks for each levels of allowed covariates (Women).

However, focusing on more recent events is crucial when we want to use rates for population forecasts. However, when we face transitions with few events, we can extend the length of the window of observation in order to gather more events and, then to obtain more stable estimates. The user can explicitly specify the value of `wl` (default value is 5 years).
It could also be useful to reduce the value of `nmin`. This may give the opportunity to have more covariates and/or less aggregated levels. However, a low the value for this parameter requires a stronger accuracy in the evaluation of the output. As a general rule, it is advisable to run ageprof with an high value of `nmin` (>=30) in order to have more stable estimates and only in second instance, try to reduce `nmin` and check if results remain similar to the previous ones.
NOTE: the automatic aggregation of categories may be excluded by setting `nmin`=1

BOX 6. `Plot.ageprof()` utility.

After the execution of `ageprof`, we can obtain a graphical representation of age profiles for a specified sex through the utility `plot.ageprof`. This command takes as main argument the standard output of `ageprof().` This utility takes into account the significance of covariates as well: when the *pvalue* of *anova* test between model without covariate and model with the specified covariate is higher than 0.05, curves are plotted in shaded gray. This permits to have an immediate glance on significant covariate.
The syntax is:

```
plot.ageprof<-function(tab,sex,edu=T,mar=T,chi=T,liv=T)
```

Arguments are:

Option `tab`	`tab` is a list containing the standard output of `ageprof()`(No default value)`.`
Option `sex`	Specifies for which sex transition rates have to be plotted. (No default value)`.`
Option `edu`	If `false` excludes plots for covariate “Level of education (EDU)”
Option `mar`	If `false` excludes plots for covariate “Marital status (MAR)”
Option `chi`	If `false` excludes plots for covariate “Own children ever born (CHI)”
Option `liv`	If `false` excludes plots for covariate “Living arrangements (LIV)”