Modelling Household Formation and Dissolution

EDITED BY

Nico Keilman

Anton Kuijsten

Ad Vossen

CLARENDON PRESS·OXFORD
1988

11

Estimation of transition rates in dynamic household models

D. Courgeau and E. Lelièvre

ABSTRACT

This chapter deals with the estimation of instantaneous transition rates in nonparametric, semiparametric, and parametric models. The bivariate and multivariate problems are introduced and discussed using data from a retrospective life history survey, undertaken in France, 1981.

11.1. Introduction

THE construction of dynamic household models requires comprehensive knowledge of the demographic events which lead to household formation and dissolution. The interplay between these events in individual cases may be observed by using prospective or retrospective data sets. We will here show how such data sets allow the estimation of transition rates between states of the family cycle. Such transition rates can be applied in macromodels (as discussed by Keilman in chapter 9) as well as in micromodels (examples of which are given by Willekens in chapter 7 and by Galler in chapter 10).

There are three types of interactions between demographic phenomena: interactions in which one phenomenon prevents the occurrence of another phenomenon, those that create the preconditions for new phenomena, and those that neither impede nor bring about a phenomenon (Courgeau, 1979).

An example of the first type of interaction is the selection which occurs by virtue of survival when studying marriage. In order correctly to estimate probabilities for each event in the pure (undisturbed) state, one usually makes the hypothesis of 'non-dependence', namely, that an individual who died single would have had the same behaviour, had he lived, as individuals who did not die at the same age (Henry, 1972).

The second type of interaction concerns events that enable the occurrence of other events which would not have been possible without the occurrence of the initial event. For instance, marriage allows legitimate reproduction.

The third type is intermediate: the disturbing phenomenon neither impedes nor brings about another phenomenon. The concept of local dependence may be introduced (Schweder, 1970) to study these interactions. This concept

formalizes the intuitive notion that a stochastic process may influence the development of another process. Hence, local dependence is a dynamic property developing with time. It always has a certain direction since the first process may be locally dependent on the second one, while this second process is locally independent of the first one (Aalen *et al.*, 1980). Such a situation may be extended to more than one type of event for each interacting phenomenon.

Different types of observational plans may be available to collect the information, and since the measurement of the interactions is dependent on these plans, we must introduce the two principal ones: prospective and retrospective observational plans.[1]

In a prospective observational plan the individuals are sampled randomly before the events of interest take place. Such a sample may be a cohort of individuals born in a given year followed prospectively by a follow-up survey. Some of the individuals may die before the events take place, others may experience these events. Such a follow-up design is rarely used in demographic studies because it is expensive and often not cost-effective, given the often considerable amount of drop-out. Moreover, the time needed to get sufficient information on household formation and dissolution prevents prompt reporting. However, the accuracy of such reports may be excellent.

In a retrospective observational plan, data are collected only from survivors, which induces a selection by virtue of survival; reliability and validity of retrospectively reported information is not assessed here (Hoem, 1983); nor are register and survey data compared (Lyberg, 1983). However, such an observational plan, which needs only one interview, is an appealing alternative. It is mainly used for demographic studies in countries where register data either do not exist or are unable to provide an answer to specific research questions.

These two kinds of observational plans create problems of censoring which are important for this paper. Right censoring, which occurs in both prospective and retrospective observations, is not too problematic. We will see later how we can obtain unbiased estimates of transition rates. Left censoring, which may occur when the observation starts at an arbitrary point in time, presents more complicated problems.[2] In such cases we are unable to estimate the effects of past history, occurring before this arbitrary time, without making far-reaching assumptions. We may assume either that the process studied begins at the first date of observation, or that the history of the process prior to this date does not affect the future of the process. Such an assumption is not always realistic. Yet it is necessary because of the extremely complicated analytical problems that would arise if it had not been made.

This paper deals with the estimation of instantaneous transition rates, which generalize hazard functions for multivariate event histories of individuals. We will restrict our discussion to right-censored event histories since, as we have seen, left censoring is very problematic.

The survey data used here record to the nearest month the dates of status changes. Such detailed event histories will allow us to work with continuous time models (Tuma and Hannan, 1984, pp. 82–8). However, it seems important to go beyond this concept. It appears that, even if the exact timing of each event is correctly registered, such a precise time may be far removed from the time generated by the relational systems within which an individual lives. These events do not occur in a linear and continuous time but in a more 'fuzzy time'. It thus seems important to introduce this 'fuzzy time' concept. Let us see how this can be done.

We will consider here the interaction between leaving the agricultural sector and marrying. In a number of cases these two events may take place almost simultaneously for the same individual, some marrying shortly before leaving the agricultural sector, others leaving the agricultural sector shortly before marrying. Under these conditions it is not significant which event takes place first. The two events can therefore be considered to occur simultaneously, even if there is a certain time-lag between them. The introduction of a 'fuzzy time' allows us to analyse these almost simultaneous events. We will see later how this concept of fuzzy time may be defined.

We will present here different methods of analysing these interactions. The first approach (section 11.2) is a nonparametric one, which generalizes the usual demographic methods of longitudinal analysis to the case of complex interactions. A new impetus was recently given to the development of the mathematical apparatus needed to estimate hazard rates and to construct confidence intervals for such nonparametric models (Aalen, 1978; Johansen, 1983). The second approach is semiparametric or even fully parametric (section 11.3). The introduction of (semi)parametric models in demography took place only recently, but their use is rapidly becoming more widespread. Here, we will consider the instantaneous transition rate as a dependent variable, and introduce a relationship between this rate and a variety of observed variables.

The purpose of this paper is to develop methods for estimating transition rates, using a detailed data set. Rather than presenting general techniques, we prefer to discuss a few concrete examples, since they illustrate appropriately the problems that arise when estimating these rates. However, the methods presented here are quite general and they may, with slight modifications, be used in very diverse situations.

The data used here consist of the retrospective life histories of a random sample of individuals aged 45–69, living in France in 1981. These data were collected from a nationwide sample and for this survey we obtained 4,602 questionnaires, with a response rate of approximatively 89 per cent. Its purpose was to study the migration, family, and work histories of the French population (compare Courgeau, 1985).

11.2. Nonparametric methods of analysis

The first type of method is a classical longitudinal analysis, generalized to deal with more complex dependencies between life history events. Let us first consider the very simple case of an independent sample from a homogeneous population with a single event.

Let T be a non-negative random variable indicating the time at which a particular event occurs. The survivor function is the probability that T is at least as great as a value t:

$$F(t) = Pr(T \geqslant t).$$

The key nonparametric approach to estimating the survivor function is derived from Kaplan and Meier's (1958) paper giving a product-limit estimator for the survivor function where no assumption is made about its functional form.

Another concept closely related to the survivor function is the hazard function. It specifies the instantaneous failure rate at $T = t$, conditional upon survival to time t, and is defined as:

$$h(t) = \lim_{\Delta t \to 0} \frac{Pr(T < t + \Delta t \mid T \geqslant t)}{\Delta t} = -\frac{d \log F(t)}{dt}.$$

Thus the logarithm of the Kaplan–Meier estimator can be used to estimate the hazard function. This will give good asymptotic estimates of the hazard function for large samples. However, for small samples the Kaplan–Meier approach is biased and therefore Aalen (1978) used martingale theory to derive a better estimator. These functions may arise in the same way for continuous and discrete cases where the Dirac delta functions handle the discrete distributions.[3]

When some individuals are censored at time t they are usually included in the number of individuals under consideration until, and at, time t. After this time they will no longer appear in this population.

Let us now first consider the bivariate case, before trying to handle the multivariate case.

11.2.1. The bivariate problem

We will here consider the relationship between getting married and leaving the agricultural sector. But as we said earlier, such a model is applicable to any bivariate situation.

Here, we work in terms of the state space diagram, given in Figure 11.1. We now have two types of failure time represented by random variables T_1 and T_2: T_1 is the age at leaving the agricultural sector; T_2, the age at marriage. In our sample everybody starts in state 0. However, an individual in state 1 may

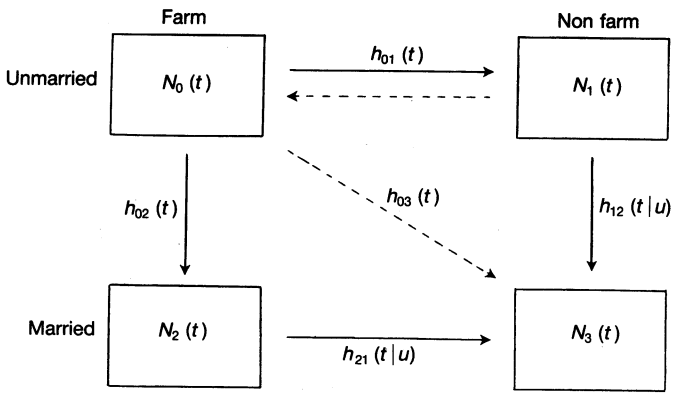

FIG. 11.1. State space diagram for the bivariate case

return to state 0 if he is not married. On the other hand, we consider only first marriage.

The previous hazard function defined in the univariate case can be generalized into four hazard functions:

$$h_{01}(t) = \lim_{\Delta t \to 0} \frac{Pr(T_1 < t + \Delta t | T_1 \geqslant t, T_2 \geqslant t)}{\Delta t}$$

with a similar one for $h_{02}(t)$, while

$$h_{12}(t|u) = \lim_{\Delta t \to 0} \frac{Pr(T_2 < t + \Delta t | T_1 = u, T_2 \geqslant t)}{\Delta t}, \quad t \geqslant u$$

with a similar one for $h_{21}(t|u)$.

It is also possible to introduce simultaneous occurrence of the two types of failure $h_{03}(t)$.

Though a fully satisfactory nonparametric procedure of estimation has not yet been found (Cox and Oakes, 1984, p. 163), we will present some approximated estimates. To do this, we have to gain insight into the studied phenomena.

As mentioned above, these events do not occur in a linear and continuou time but in a 'fuzzy time'. It therefore seems more important to introduce this fuzziness than to give a nonparametric procedure of estimation.

Let us first consider, for example, the marriage behaviour of men who are on the verge of leaving the agricultural sector. Such behaviour may be closer to that of men who have already left this sector, than to that of men who remain in it. It therefore seems better to consider such individuals as having left the agricultural sector.

On the other hand, let us consider the attitude towards the labour market of men who are on the verge of marrying. Again, such behaviour may be

closer to that of men who are already married than to that of bachelors. It thus seems better to consider such individuals as being married. The problem is now how to introduce such a 'fuzzy time'.

Different possibilities are open to us. We may focus on the decision-making process. Let us assume a time-lag t_1 between actually leaving the agricultural sector and the decision to leave it, and another time-lag t_2 between marriage and the decision to marry. Although we realize that taking such decisions is a gradual process, we may introduce different time-lags, from one month to one year, for example, and observe how the results differ.[4]

We used a similar procedure for our French survey, taking a one-year time-lag. This procedure, less accurate than the previous one, is easier to implement, as it introduces a time-discrete version of the nonparametric model. Let us have a closer look at the estimates to which it leads. Censored individuals will not be taken into account.

Let $N_i(t)$ $(i = 0,1,2)$ be the population in state i at the beginning of year t. Let $n_{ij}(t)$ be the number of occurrences of type j in the population in state i. Let $r_{10}(t)$ and $r_{32}(t)$ be the number of people returning to the agricultural sector.

We assume that the simultaneous events occurring to the same individual during the same year (this occurs only for 5 per cent of the observed population) are related to the corresponding population at risk.[5] As we have a limited number of observations (668 men and 519 women) we assume that the behaviour of the observed individuals will depend only on their age and not on the time the previous event occurred. With this assumption we can estimate the following hazard rates by observed occurrence/exposure rates using the total time at risk during year t and assuming that both types of events occur uniformly throughout the interval:

$$\hat{h}_{01}(t) = \frac{n_{01}(t)}{N_0(t) - \frac{1}{2}(n_{01}(t) + n_{02}(t) - r_{10}(t))},$$

$$\hat{h}_{21}(t|u) = \frac{n_{21}(t)}{N_2(t) - \frac{1}{2}(n_{21}(t) - n_{02}(t) - r_{32}(t))},$$

$$\hat{h}_{02}(t) = \frac{n_{02}(t)}{N_0(t) - \frac{1}{2}(n_{02}(t) + n_{01}(t) - r_{10}(t))},$$

$$\hat{h}_{12}(t|u) = \frac{n_{12}(t)}{N_1(t) - \frac{1}{2}(n_{12}(t) + r_{10}(t) - n_{01}(t))}.$$

We are then able to compare $h_{01}(t)$ with $h_{21}(t)$ (respectively, $h_{02}(t)$ with $h_{12}(t)$) to see whether or not the behaviour of unmarried individuals (of individuals working in the farming or nonfarming sector, respectively) is different. To do this, we may use test statistics given in Hoem and Funck Jensen (1982).

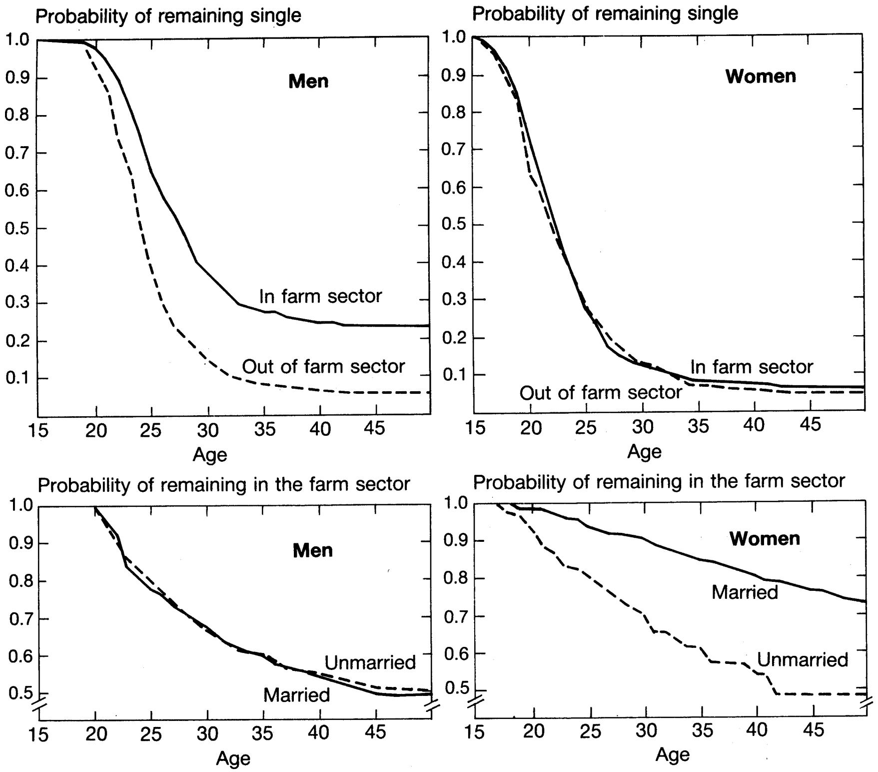

FIG. 11.2. Probabilities of remaining single inside or outside the farm sector, and probabilities of remaining in the farm sector, by marital status

Figure 11.2 shows the results for the estimated survivor functions in each of the four defined states. The behaviour of men and women is very different. For men, the probability of remaining single is significantly higher for those in the agricultural sector than for others, while for women it is the same in both categories. On the other hand, for men there is no significant difference in the probability of leaving the agricultural sector between those who are unmarried and those who are married, while married women remain in the farming sector significantly more often than unmarried women. Hence, in terms of local dependence, we find the following: for men, being in the farming sector diminishes their probability of marriage, while being married or not does not influence the probability of them leaving the farming sector. For women the reverse is true.

11.2.2. The multivariate problem

We will consider here, as an example of the multivariate problem, the interaction between migration occurring after marriage and marital fertility. The state space is presented in Figure 11.3. We can also introduce failure times represented in each of the two dimensions by random variables T_1, $T_2 \ldots T_k \ldots$ for migration and $T^1, T^2, \ldots T^n \ldots$ for childbirths.

A state is specified as an ordered pair (k,n) where k is the number of moves previously experienced and n the number of children previously born. We have hazard functions between all these states of the following kind (for migration):

$$h^n_{k,k+1}(t \mid u_1, \ldots, u_k, v_n, \ldots, v_1) =$$
$$\lim_{\Delta t \to 0} \frac{Pr(T_{k+1} < t + \Delta t \mid T_{k+1} \geqslant t, T_1 = u_1, \ldots, T_k = u_k, T^n = v_n, \ldots, T^1 = v_1)}{\Delta t}.$$

Evidently, such a model cannot be estimated with a small number of individuals, as in our survey, and we have to make some further assumptions. Let us assume that a migration hazard function is dependent on its rank, on the time since the last move (u) and on the time since the last childbirth (v). With these assumptions, we can write for the hazard function

$$h^n_{k,k+1}(t \mid u,v) = \lim_{\Delta t \to 0} \frac{Pr(T_{k+1} < t + \Delta t \mid T_{k+1} \geqslant t, T_k = u, T^n = v)}{\Delta t}.$$

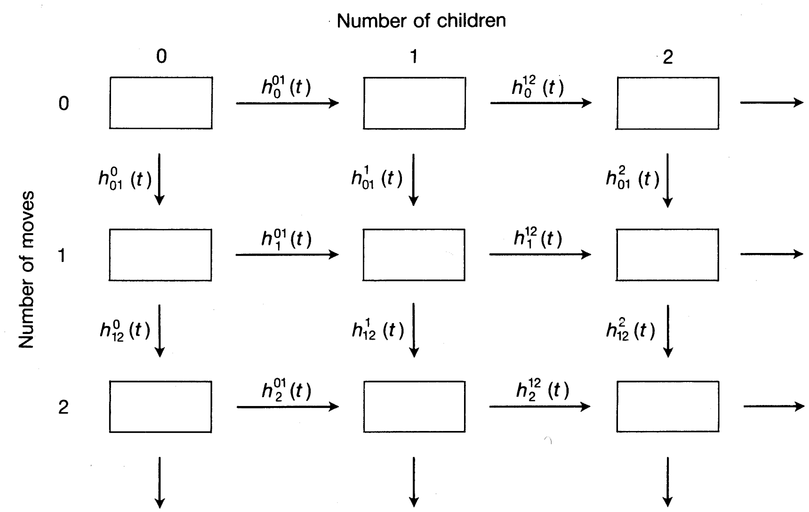

FIG. 11.3. State space diagram for the multivariate case

For the childbearing hazard function we have similar hypotheses:

$$h_k^{n,n+1}(t|u,v) = \lim_{\Delta t \to 0} \frac{Pr(T^{n+1} < t + \Delta t | T^{n+1} \geqslant t, T_k = u, T^n = v)}{\Delta t}.$$

We assume further, that both of the two previous functions are independent of u and v, and that $h^n_{k,k+1}(t)$ depends only on n, whereas $h_k^{n,n+1}(t)$ depends only on k.

Under these assumptions it is possible to estimate the hazard rates. They will allow us to answer the following two questions:

- What will be the effects of n births after marriage on the migration process?
- What will be the effects of having undertaken k moves after marriage on the childbearing intensity?

We can obtain a cumulative spatial mobility index until time t for the hypothetical population of women of parity n (or the cumulative fertility of women with k moves). However, as the risk set may be small for short durations and values of n or k greater than zero, we assume that if the risk set contains less than 50 individuals, the cumulative mobility index (or the cumulative marital fertility) of couples with n children (k moves) is the same as that of couples with $(n-1)$ children ($(k-1)$ moves).

Figure 11.4 gives the results for women born in 1911–25, married to men born in 1911–35.[6] We classify these women according to their age at marriage (15–22, 23–30). For women who married young, there is a clear effect of family size on the cumulative number of moves: the greater the size, the more mobile the subpopulation will be. However, such an effect, although striking for the lower birth ranks, is less perceptible for the higher ones. For the same women we can also see that some moves are undertaken to provide for forthcoming births.

For women married after the age of 22 years, the effect of family size on the number of moves undertaken has entirely disappeared. These women, compared with the younger ones, appear to have a dwelling sufficiently large to anticipate their ultimate family size. But, on the other hand, the cumulative fertility of these women according to their number of previous moves indicates that some of those moves are undertaken to provide for forthcoming births. Thus, in this case, we have a local dependence of fertility on spatial mobility: if no more moves are undertaken after or during the year of a birth in the household, it appears that some moves may be undertaken in expectation of future births.

The assumptions applied here seem very restrictive, because they take into account only the longer term effects on the level of each of the two status dimensions considered (childbearing parity and number of moves experienced). It will then be important to take into account the short-term after-effects of the occurrence of one event on the other. In this case, if the birth of a child modifies the migratory behaviour of his parents, then

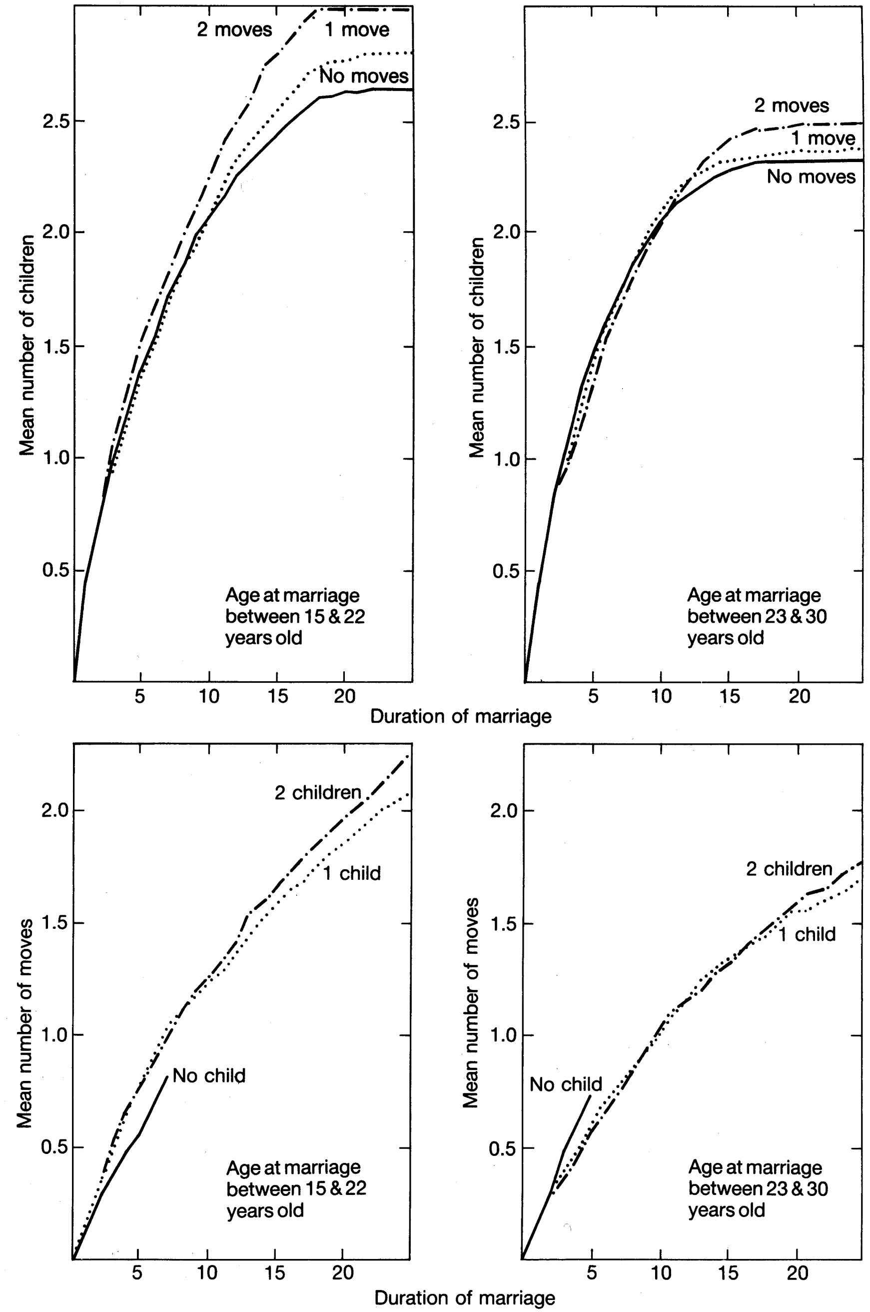

FIG. 11.4. Mean number of moves for women having 0, 1, or 2 children, and mean number of children for women having undertaken 0, 1, or 2 moves, according to age at marriage and marriage duration (women born in 1911–1925)

$h^n_{k,k+1}(t|u,v)$ should depend on v, at least for small v, and this effect will come in addition to its possible dependence on t and we shall have to estimate $h^n(t|v)$. In the same way, if a new spatial move modifies the couple's reproductive behaviour, then $h^{n,n+1}_k(t|u,v)$ should depend on u, at least for small u, in addition to its possible dependence on k. We will have to estimate $h_k(t|u)$. However, the limited number of surveyed persons does not allow us to go further. In order to extend such an analysis, we need parametric methods.

11.3. Semiparametric and parametric methods of analysis

When trying to introduce some exogenous variables into the previous analysis, it seems easier to discard the interaction point of view we took in section 11.2 and to introduce a more causal analysis on the studied phenomena. However, such an analysis is, as yet, not wholly causal. It is only a one-sided point of view of two or more interacting phenomena.

11.3.1. The bivariate problem revisited

We will now attempt to introduce exogenous variables into the previous bivariate analysis. We shall consider a semiparametric model for the hazard function of leaving the farming sector. Suppose, the hazard is affected by marriage but that it is independent of the waiting time u. We use a model considered by Crowley and Hu (1977). It uses proportional hazards for the two rates. Then we have

$$h_{01}(t|z) = h(t)\exp(z\beta_1),$$
$$h_{21}(t|u,z) = h(t)\exp(z\beta_1 + \beta_0 + z\beta_2).$$

The parameters β_0, β_1, β_2 can be estimated by partial likelihood methods (Kalbfleisch and Prentice, 1980). Table 11.1 gives estimates of the regression coefficients for models introducing different kinds of variables.

Recall that for men, leaving the farming sector is independent of marital status, whereas married women remain in the farming sector to a far greater degree than unmaried women. This is consistent with the results given by model 1, where only marital status is considered: this variable appears to be associated significantly with an important decrease in the exit risk for the female population only.

Then the number of siblings (model 2) appears to have an impact on unmarried women only: they leave the farming sector at a rate which increases concomitantly with the number of siblings. On the other hand (model 3), elder women, when unmarried, leave this sector to a lesser extent. Once married, there is no interaction between these two variables, so that their effect will remain at the previous level.

TABLE 11.1. Regression coefficients for a semiparametric model of exit from the farming sector

Model	Variables	Indicator	Main effect (β_1)		Marital status (β_0)		Interactions (β_2)	
			Males	Females	Males	Females	Males	Females
			$\times 10^{-3}$					
1	Marital status	0 unmarried 1 married	—	—	−171	−835[a]	—	—
2	Number of sibs	number	3	16[a]	−170	−835[a]	−0	−3
3	Elder	0 if not 1 if elder	−67	−412	−161	−930[a]	−46	258
4	Father farmer	0 if not 1 if farmer	−563	−950	−483[a]	−1255[a]	490	608[a]
5	Father-in-law	0 if not 1 if farmer	—	—	189	−563[a]	−580[a]	−452[a]

[a] One-sided test significant at the 5 per cent level.

When introducing the father's occupation, a number of interesting results appear (model 4). Unmarried men and women exhibit the same behaviour when their father is a farmer: they leave the farming sector to a lesser extent than married men and women. This may be because they are to inherit their father's farm. Once married, the interaction is significant but in the opposite direction. There is no longer any difference in behaviour between them, irrespective of whether or not their father is a farmer.

Model 5 includes a covariate in the time-dependent part of the model and not in the constant part. It measures whether or not the father-in-law is a farmer. Again, such a variable has a very clear effect: it lowers the degree to which both sexes leave the agricultural sector, once married. However, these rates remain very close to previous ones.

This example shows how such an approach will be important for future demographic research, especially when improving it by introducing interaction effects. We will not consider them here.

11.3.2. A split of the multivariate problem into parametric models

We shall now introduce exogenous variables into the previous analysis of migration related to childbearing. To do this we will split up this complex problem into parametric or semiparametric models. Let us take the migration process as being the main one, and introduce the other variables as explanatory variables.

As before, let T_i^0 be the duration of residence of an individual. We can write his survival probability as a function of a vector x_i of characteristics of the individual at the beginning of the observation period. Hence

$$F(t; x_i, \theta) = Pr(T_i^0 \geqslant t; x_i, \theta)$$

where θ is a parameter vector which we have to estimate.

We can also define a probability density function for migration as:

$$f(t; x_i, \theta) = \lim_{\Delta t \to 0} \frac{Pr(t \leqslant T_i^0 < t + \Delta t; x_i, \theta)}{\Delta t} = -\frac{\partial F(t; x_i, \theta)}{\partial t}.$$

However, as we are working on retrospective survey data we have for some duration of residence a censored observation, so that $T_i < T_i^0$. As censoring times are in this case stochastically independent of each other and of the failure time, we can write

$$Pr(T_i < t + \Delta t; \delta_i = 1, x_i, \theta) = O_i(t) f(t; x_i, \theta)\, \Delta t,$$

$$Pr(T_i < t + \Delta t; \delta_i = 0, x_i, \theta) = q_i(t)\, F(t; x_i, \theta)\, \Delta t,$$

where δ_i is a dummy variable that is equal to zero if the ith item is censored,

and equal to one if it is not censored; $O_i(t)$ and $q_i(t)$ are the survivor and density functions for censoring.

As neither O_i nor q_i are informative about θ, the likelihood of the data is proportional to

$$\prod_{i=1}^{n} f(t_i; x_i, \theta)^{\delta_i} F(t_i; x_i, \theta)^{1-\delta_i}$$

where n is the number of observed durations of residence.

It is then possible to estimate the parameter vector θ with maximum likelihood methods. It has been shown that the asymptotic distribution of θ is multivariate normal with mean $\hat{\theta}$. When using the Newton–Raphson technique to find the maximum of the likelihood function, we also have an estimate of the covariance matrix of these parameters.

We suppose here that the instantaneous migration rate is related to the observable variables in a generalized Gompertz model:

$$h(t; x_i, \theta) = \exp(\theta x_i + \theta_0 t).$$

Previous migration analyses show a very good fit of regression models to Gompertz's duration of residence effect (Ginsberg, 1979). We suppose that the other variables act multiplicatively on the migration rate.

We will not present here the detailed analysis we undertook in another paper (Courgeau, 1985), using 37 variables, introducing first age group variables with duration of residence, then family life cycle variables, tenure status variables, career variables, and finally some more general variables introducing war periods or periods of economic crisis. We will only give as an example the reduction of age effects when introducing different kinds of variables. Figure 11.5 shows the multiplicative effect of age on mobility

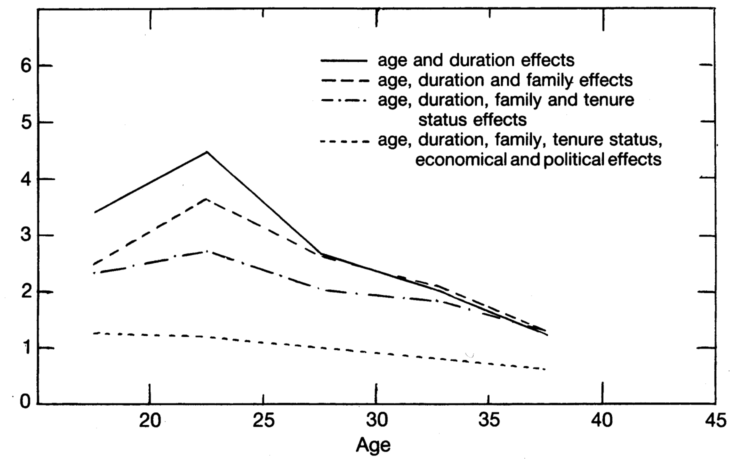

FIG. 11.5. Age-specific multiplicative effect on mobility according to the set of considered variables: males born in 1931–1935

according to the set of considered variables, for the male cohort born in 1931–35.[7]

When introducing only age and duration effects, we have a curve that is very close to the curve we obtain with period data: a maximum migration rate for the age group 20–4 years, and an important reduction for the older age groups. When introducing family variables, this age effect is reduced for the age group 15–24. Introducing tenure status variables yields a new reduction of this age effect for all age groups. Later, the introduction of the whole set of variables cancels out any age effect. Hence, for this cohort we are able to explain all age effects usually noted in the different stages of the individual's family, economic, and political life.

11.4. Conclusion

We have here covered the entire spectrum of analyses, from nonparametric models to fully parametric ones, via semiparametric models. We have also introduced univariate analyses, bivariate ones, and finally more complex multivariate analyses. Through this exploration, some important issues for future research have appeared. We will pin-point them and give some suggestions for future research.

An important problem was that we were not able to estimate complete models because this would require a huge sample with many observations in each cell. Such detailed information was not available. Therefore, we introduced some assumptions on the main effects. In order to do so we mainly used earlier experiences with demographic data. For example, this applies to the prominent effect of the duration since the last event rather than to the duration since earlier ones. Such hypotheses need to be explored in more detail. With more surveys containing event history data, we will be able to give a more solid basis to these hypotheses. However, to do this we will need larger samples than usual.

Another problem will be to introduce more time-dependent explanatory variables into these analyses. When revisiting the bivariate problem we introduced such a time-dependent variable, namely, being married or not at time t. Such an approach needs to be generalized. In the parametric approach we only allow for the value of the parallel process at the beginning of each event interval. Obviously, these life cycle variables, tenure status variables, and career variables may change in between moves or births. It will be important, in future work, to try to incorporate such time-dependent explanatory processes as well.

A third problem did not appear clearly from the previous analyses but seems very important for further research. We introduced previously different explanatory variables to control their effect. However, other characteristics of the studied individuals interfere with them, while we do not have any possibility of measuring these characteristics. We can incorporate this unobserved heterogeneity (Tuma and Hannan, 1984, pp. 155–86) into the models. However, research on this problem has only just begun, and its possible solutions will most likely appear in future research.

Last but not least, we have introduced a 'fuzzy time' that leads to very important problems which are difficult to deal with. The estimation of transition rates needs a very precise time scale on which events can be placed. However, when studying demographic events, we do not observe such a time scale, but rather a more 'fuzzy time'. It does not seem important, when working with a time interval of six months or even a year, to know which event occurred first (for example a birth or a migration). The analysis on a very precise time scale may lead to inconsistent results with a sociological analysis. One way to avoid such inconsistencies may be to introduce a time which is not ordered linearly. For example, this may occur in a two-dimensional time-space: the first dimension may be our common time; the other, a hazardous time that will be added to or subtracted from the continuous time. This possibility is open to further research.

Acknowledgements

The authors are indebted to Richard Gill for valuable comments and discussions on a preliminary version of this paper and to Stephen Clarke for correcting the English.

Notes

1. We will not consider here another commonly used observational plan, namely panel data. Such data are usually so scarce that it is difficult to know the exact timing of every demographic event.
2. In demographic surveys, such as the ORIN Survey in the Netherlands, left censoring, rather than left truncation, arises (Klijzing, chapter 4). Left truncation arises when individuals come under observation only some known time after the natural time origin of the phenomenon under study. That is, had the individual failed before the truncation time in question, that individual would not have been recorded (Cox and Oakes, 1984). Such a truncation leads to a maximum likelihood fitting of any parametric or nonparametric model.
3. Such a function $\delta(t)$ is defined so that $\delta(t)\,dt = 1$ if $t = 0$, otherwise $\delta(t)\,dt = 0$.
4. Nico Keilman is trying to use this approach to study interrelations between migration and birth events.
5. It is also possible to exclude them and to compute rates as $h_{03}(t)$ in Figure 11.1.
6. For reasons of clarity, we consider only moderate family size and moderate numbers of migration.
7. Such an effect is measured by the antilog of the parameters ($\exp \theta$). When it is equal to one, the behaviour of the considered age group will be the same as the behaviour of the comparison group (40 years and over). To undertake this analysis we use the Fortran computer programme 'RATE', written by N. Tuma and D. Pasta.

References

AALEN, O. (1978). Nonparametric inference for a family of counting processes, *The Annals of Statistics* 6, 701–26.
—— Ø. BORGAN, N. KEIDING, AND J. THORMANN (1980). Interaction between life history events: nonparametric analysis for prospective and retrospective data in the presence of censoring, *Scandinavian Journal of Statistics* 7, 161–71.
COURGEAU, D. (1979). Migration and demographic phenomena in France, in: J. White (ed.), *The Urban Impact of Internal Migration* (Chapel Hill, University of North Carolina Press), 1–32.
—— (1985). Interaction between spatial mobility, family and career life cycle: a French survey, *European Sociological Review* 1, 139–62.
COX, R. AND D. OAKES (1984). *Analysis of Survival Data* (Chapman and Hall, London).
CROWLEY, J. AND M. HU (1977). Covariance analysis of heart transplant data, *Journal of the American Statistical Association* 72, 27–36.
GINSBERG, R. (1979). Timing and duration effects in residence histories and other longitudinal data, II: studies of duration effects in Norway, 1965–1971, *Regional Science and Urban Economics*, 369–92.
HENRY, L. (1972). *Démographie: analyse et modèles* (Larousse, Paris).
HOEM, J. M. (1983). Weighting, misclassification and other issues in the analysis of survey samples of life histories, Stockholm Research Reports in Demography, 11.
—— AND U. FUNCK JENSEN (1982). Multistate life table methodology: a probabilist critique, in: K. C. Land and A. Rogers (eds.), *Multidimensional Mathematical Demography* (Academic Press, New York), 155–264.
JOHANSEN, S. (1983). An extension of Cox's regression model, *International Statistical Review* 51, 165–74.
KALBFLEISCH, G. D. AND PRENTICE, R. L. (1980). *The Statistical Analysis of Failure Time Data* (John Wiley, New York).
KAPLAN, E. L. AND P. MEIER (1958). Nonparametric estimation from incomplete observations, *Journal of the American Statistical Association* 53, 457–81.
LYBERG, I. (1983). The effects of sampling and non-response on estimates of transition intensities: some empirical results from the 1981 Swedish fertility survey, Stockholm Research Reports in Demography, 14.
SCHWEDER, T. (1970). Composable Markov processes, *Journal of Applied Probability* 7, 400–10.
TUMA, N. B. AND M. HANNAN (1984). *Social Dynamics: Models and Methods* (Academic Press, Orlando).

PART IV

Applications