The Plenum Series on Demographic Methods and Population Analysis

Series Editor: Kenneth C. Land, *Duke University, Durham, North Carolina*

ADVANCED TECHNIQUES OF POPULATION ANALYSIS
Shiva S. Halli and K. Vaninadha Rao

CURBING POPULATION GROWTH: An Insider's Perspective on the Population Movement
Oscar Harkavy

THE DEMOGRAPHY OF HEALTH AND HEALTH CARE
Louis G. Pol and Richard K. Thomas

FORMAL DEMOGRAPHY
David P. Smith

HOUSEHOLD COMPOSITION IN LATIN AMERICA
Susan M. De Vos

HOUSEHOLD DEMOGRAPHY AND HOUSEHOLD MODELING
Edited by Evert van Imhoff, Anton Kuijsten, Pieter Hooimeijer, and Leo van Wissen

MODELING MULTIGROUP POPULATIONS
Robert Schoen

THE POPULATION OF MODERN CHINA
Edited by Dudley L. Poston, Jr. and David Yaukey

A Continuation Order Plan is available for this series. A continuation order will bring delivery of each new volume immediately upon publication. Volumes are billed only upon actual shipment. For further information please contact the publisher.

Household Demography and Household Modeling

Edited by

Evert van Imhoff

Netherlands Interdisciplinary Demographic Institute
The Hague, The Netherlands

Anton Kuijsten

University of Amsterdam
Amsterdam, The Netherlands

Pieter Hooimeijer

Utrecht University
Utrecht, The Netherlands

and

Leo van Wissen

Netherlands Interdisciplinary Demographic Institute
The Hague, The Netherlands

Plenum Press • New York and London

8. EVENT HISTORY ANALYSIS OF HOUSEHOLD FORMATION AND DISSOLUTION

Daniel Courgeau

Institut National d'Etudes Démographiques
27, rue du Commandeur
F-75675 Paris Cédex 14
France

Abstract. This chapter shows how traditionally individual-oriented event history analysis may be generalized to model household history. These models raise an important distinction between different time scales. A first scale uses a common time origin for all episodes, while the second one introduces a waiting time for each change in household size. These two clocks lead to different results when applied to a retrospective life history survey undertaken in France. The discussion of these results leads to a preference for the waiting time model for household history analysis.

8.1. INTRODUCTION

Event history analysis, using individual data sets, has become widely accepted in demography during the last ten years. It has enabled the analysis of *interactions* between events and taking into account the *heterogeneity* of the studied population (Courgeau & Lelièvre, 1989 and 1992), by focussing on several different demographic events simultaneously which the researcher considers to be particularly important ones.

Some of these individual data sets, like the French "3B" survey, which reconstructed three biographies (family, profession, and migration), were collected to study *individual* life courses. However, these data give very detailed information on *household* changes (Klijzing 1988) and also permit a focus on household histories, even if some elements of these histories are missing.

We will take this household perspective to show how traditionally individual-oriented event history analysis may be generalized to analyse

Household Demography and Household Modeling
Edited by E. van Imhoff *et al.*, Plenum Press, New York, 1995

household histories. To support such a theoretical approach and to show its limits, we will use data from our French "3B" survey. This data set consists of retrospective life histories from a random sample of individuals between the ages of 45 and 69 inclusive, all of whom were living in France in 1981. From this survey, we obtained 4,602 completed questionnaires (2,050 for males and 2,552 for females), giving a response rate of 89 per cent. These individual histories contain information from the birth of the individual until the observation year 1981, and on a number of variables concerning parents' origin, childhood history, marriage and childbearing history, and detailed job and residence history.

We will first present a general event history model for the analysis of household histories (Section 8.2). This model raises an important distinction between different time scales. We will distinguish two alternatives here. Firstly, we can introduce the time since the first establishment of the household as a common time horizon for the household history. Secondly, we can introduce a waiting time for each change in household size, with the clock reset to zero after each household change. These two clocks lead to different results. In Section 8.3 we try to apply such a model, from the more simple situation to more and more complex ones. This will give an idea of the model's usefulness in unravelling the interrelationships between household formation or dissolution and the various event histories of each household member.

8.2. A GENERAL MODEL FOR HOUSEHOLD FORMATION AND DISSOLUTION

We define a household as a primary group identified by the co-residence of its members. This group may consist of one or more individuals and may be both familial and non-familial (Kuijsten & Vossen 1988). However, such a definition is a cross-sectional one, and when one tries to give a proper longitudinal definition of a household many problems occur (McMillan & Herriot 1986; Duncan & Hill 1985). All possible longitudinal household definitions contain one or more decision rules which are to a large extent arbitrary, as well as different continuity rules which may produce different findings for the same situation (Keilman & Keyfitz 1988).

For these reasons, we prefer to follow an *individual-based* approach, which is also consistent with our "3B" survey. The time origin is the moment when the household is established with at least one member, the surveyed individual. Since the survey contains the occupation of each place of residence, we are able to detect a leave from the parental home. However, some of these departures may not be considered as the beginning of a household. For example, departures for military service or stays in student residences, in hospital (for long duration illness), etc., are not to be considered. In that case, we have to make a

preliminary choice to be able to precisely define a departure from the parental home which leads to a new household. For our survey, we have considered the residential history and the professional one simultaneously, for which every period can be characterized by many possible states (studies, military service, war periods, etc.). The household may then follow a path of formation and dissolution, with successive increases or decreases in its size and changes in its composition.

Let us first focus on size changes. For a given household, we may define successive periods of time during which its size remains the same. Each of these periods has an end point, the time at which an event occurs, with the introduction of one or more newcomers or the departure of other members. From one given size, such a change may lead to any other size. For example, it may be the simultaneous arrival of a great number of individuals (a man marrying a widowed woman with many children all entering his household) or the departure of many individuals (a woman divorcing and leaving a household with all her children). Even in some cases, for example, when we are working in single-year discrete time, the resulting change in size may be zero as the death of a child may occur very close to the birth of another one.

The death of the household marker does not happen in our data set because the survey is retrospective. When using prospective data, such a death leads to a right censored observation of the household. As this censoring scheme is not independent of the household history, it creates problems that are difficult to solve. This issue will not be pursued in the present paper.

Let us now formalize such an approach. We are interested here in the following times at which changes in household size occur, considered as ordered random variables $0=T_0 \leq T_1 \leq T_2 \ldots$, and the sizes of the household in successive episodes k between size changes characterized by a series of random variables $\{S_k;\ k=0,1,2\ldots\}$, $S_k \in \{0,1,\ldots,m\}$, where m is the maximum of the observed household sizes.

The corresponding stochastic process which describes the state of a household by a continuous time, discrete state process, can be decomposed into two related processes. First, a *duration process* which governs the time elapsed since the occurrence of some specific event. We will see later how it is possible to undertake different analyses by changing the definition of such a specific event. Second, a *transition process* which governs the moves between household sizes (Hamerle 1989). As stated above, such changes may lead to any other possible size.

Let us suppose that we have a vector $z_k(t)$ of relevant covariates measured at time t for every individual, including characteristics of the other household members, and episode k. Such a vector may contain dummy or metric variables, or both. These covariates may be time invariant or time dependent. In the latter case, the covariates may change during episode k. We also have to consider the previous history of the process until t_{k-1}. This history is collected in a vector

$U_{k-1}=\{s_0, z_0, t_1, s_1, z_1, \ldots, t_{k-1}, s_{k-1}, z_{k-1}\}$, where z_l is always a vector of covariates that may be time dependent in the interval (t_{l-1}, t_l); s_l is the size of the household; and t_l the time of change.

There are different ways to model the time dependence of event-specific hazard rates. We can first use a common process time for all episodes with a *common time origin*. As previously stated, the study of household formation and dissolution begins when the household is established with at least one member: this may be the time origin. In such a case, all subsequent starting and ending times of an individual's episodes are the durations between the time origin and the occurrence of household changes. Then, both the duration and the transition process can be simultaneously characterized by the following event-specific hazard rate:

$$h_k^{s_{k-1}\, s_k}(t;\ z_k(t),\ U_{k-1}) =$$

$$= \lim_{dt\to 0} \frac{P(T_k<t+dt,\ S_k=s_k \mid T_k \geq t \geq t_{k-1},\ S_{k-1}=s_{k-1},\ z_k(t),\ U_{k-1})}{dt} \quad (1)$$

for the k^{th} move from household size s_{k-1} to s_k. Such a rate will be identically zero for $t<t_{k-1}$. There is a possibility that some household sizes are not attainable from certain sizes. This can be accommodated in equation (1) by restricting it to be identically zero for the appropriate (k, s_{k-1}, s_k) values. However, even with this possibility, the number of rates to estimate is so large that the observation of a very large sample is required.

This common time origin supposes that the time from the first establishment is the crucial variable affecting the future evolution of the household. In this case, we assume that the successive events occurring during the process of household formation and dissolution do not alter the rates for subsequent episodes: these are mainly directed by the duration since initial household establishment. Such an assumption does not permit a picture showing the dependence on time since the beginning of the k^{th} episode, without introducing a multidimensional time scale. An alternative possibility is to consider that experiencing an event once alters the rates for subsequent episodes. In this case the 'common time horizon' loses importance, while we are taking a 'waiting time approach'. We will consider here that the clock is reset to zero after each event. The crucial variables are now the waiting times from successive household changes, each new position developing its own time scale.

Following this alternative approach, let us model the hazard rates in terms of *waiting times*, starting with time zero whenever a new episode begins. For the first episode, the two approaches lead to the same result as the starting time

is identical. For the following ones, we will have to reformulate equation (1). For the k^{th} change in household from size s_{k-1} to size s_k, the hazard rate may be written:

$$h_k^{s_{k-1}\, s_k}(t;\ z_k(t),\ U_{k-1}) =$$

$$= \lim_{dt\to 0} \frac{P(T_k-T_{k-1}<t+dt,\ S_k=s_k \mid T_k-T_{k-1}\geq t,\ S_{k-1}=s_{k-1},\ z_k(t),\ U_{k-1})}{dt} \quad (2)$$

The time t is now measured *from the beginning of each episode*, the other characteristics being the same as for the previous equation (1). Such a model would be suitable if the process of household formation and dissolution leads to hazard functions that can be more parsimoniously expressed in terms of gaps between failure times, rather than in terms of the total observation time (Kalbfleisch & Prentice 1980).

Let us see now in more detail how to include covariates and past history. As the household size changes, the covariates may depend on the various members which are present at each time. If the household has only one member, these covariates refer to him and his relatives: educational level, parents' occupation, past history, etc. If it is composed of more than one member, we can introduce the previous characteristics for each member of the household in the same way. It is also useful to introduce more complex characteristics that are related to the household as a whole: household composition, members becoming home-owners, etc.

It is often necessary to reduce the number of changes in households sizes. As we previously stated, it is not possible with survey data to consider every pair of sizes (s_{k-1}, s_k), as the number of such pairs is too large to permit a useful analysis of household changes. We will consider only two types of movements: to lower sizes and to sizes equal to or larger than the current one. As we are working in single-year discrete time, some changes in size may be zero. However, as such changes are very rare occurrences, we will not consider them separately and will include them arbitrarily with 'larger sizes'. In that case, we only have to estimate the two following kinds of rates:

$$h_k^1(t;\ z_k(t),\ U_{k-1}) = \sum_{s_k \geq s_{k-1}} h_k^{s_{k-1}\, s_k}(t;\ z_k(t),\ U_{k-1}) \quad (3)$$

and

$$h_k^2(t;\ z_k(t),\ U_{k-1}) = \sum_{s_k < s_{k-1}} h_k^{s_{k-1}\ s_k}(t;\ z_k(t),\ U_{k-1}) \qquad (4)$$

It is easy to generalize these estimations to any other kind of more complex moves.

It is also necessary to formulate the hazard function with more precise specifications to be able to estimate its parameters. We will use a semi-parametric approach, which is a generalization of the Cox proportional hazard model. We assume that the hazard rates have a shape function that depends arbitrarily on the index number k of the event and that the various covariates act multiplicatively on these rates. We will introduce some of the characteristics of the previous history U_{k-1} that are considered as useful predictors of the hazard rate; they will be included in the $z_k(t)$ covariates.

If we follow a household from its first establishment, ignoring the waiting times between each step, we can write the following hazard rate, with $i=\{1,2\}$:

$$h_k^i(t;\ z_k(t),\ U_{k-1}) = h_{k0}^i(t) \times \exp\left(z_k^i(t)\beta_k^i\right) \qquad (5)$$

where $h_{k0}^i(t)$ is an unspecified baseline hazard function and β_k^i is a column vector of regression parameters. To estimate these regression coefficients, a partial likelihood approach may be used (Cox 1972; Kalbfleisch & Prentice 1980). Such a partial likelihood[1] may be derived as:

$$PL = \prod_k \prod_i \prod_{l=1}^{d_k^i} \frac{\exp\left(z_k^{il}(t_k^{il})\beta_k^i\right)}{\sum_{r \in R(t_k^{il})} \exp\left(z_k^{ir}(t_k^{il})\beta_k^i\right)} \qquad (6)$$

where t_k^{i1}, t_k^{i2}, ... are the d_k^i durations in which the k^{th} episode ended by a transition to state i, and where $R(t_k^{il})$ is the set of households that are at risk of their k^{th} transition just before time t_k^{il}. For the common time horizon approach, such a risk set may increase as time goes on. For example, in a certain time interval, a large number of individuals can have their $(k\text{-}1)^{th}$ household change and then become at risk of their k^{th} change, while those who have their k^{th} change in the same interval may be less numerous. This may lead to an increase in the risk set.

[1] To make this estimation, we used the computer program TDA (Transition Data Analysis) written by Götz Rohwer (European University Institute).

In the second case, we follow the successive waiting times of an individual. The expression for the hazard rate is the same as (5), except that time t is defined in a different way. A similar partial likelihood like (6) may be written to estimate the regression parameters. But in this case the risk set is always decreasing as time goes on, as at each episode everybody begins at time zero. The two partial likelihoods are not directly comparable, and cannot be used to tell which approach is the better one. To do that, we will have to use some other methods.

Until now, we only dealt with *changes in the size* of the household. We will now have to introduce *changes in composition*. According to the household position of the entering or departing members, we may observe very different changes in composition for a given size change. A divorce and a child's death may lead to the same decrease in household size, but have quite different consequences for subsequent household development. As we consider household changes, we may introduce the occurrence of one or more different events: marriage, a child's birth, a divorce, etc. A split of the destination state according to the events leading to it permits such an introduction. This leads to a generalization of previous hazard rates without introducing new difficulties. For example, rate h_k^1 may be split in h_k^{1m}, if the change is due to a marriage, h_k^{1b}, if the change is due to a child birth, etc. We can then introduce comparative rates on household composition changes, according to the destination state.

8.3. AN EXAMPLE OF ANALYSIS OF HOUSEHOLD FORMATION AND DISSOLUTION

As previously stated, the "3B" survey gives detailed information on some changes in household composition. But as it was undertaken with an individual perspective, some elements may be missing. Let us see first how to approach a household history.

8.3.1. The "3B" Survey in a Household Perspective

Since the survey contains the residential history and the professional one, we are able to define very precisely a departure from the parental home in the way we gave in Section 8.1 of this chapter.

When leaving his parental home, an individual may already be married, with children, and the initial size of his household may be greater than one. Also, once in a new household, an individual may come back to his parental home, and we have to consider the possibility of several successive households for the same individual. In every case, we have to follow the increase or the decrease of this household size, according to marriage, birth of children, divorce, departure of children from the parental home, etc. The major part of these events are to be found in the part of the questionnaire on the matrimonial history and on children of the surveyed person.

However, when taking a household perspective, some events are missing from our survey. Cohabitation was not taken into account, but since it was a rare situation for the surveyed generations during their youth, this will not introduce an important bias. It is evident that if we study more recent generations, cohabitation will have to be considered. Also, only the children or the adoptive children of the surveyed individual have been considered. We may miss some children from previous marriages of the spouse that are living in the household. We also lose some ancestors, siblings, cousins, etc., or other non-kin members of the household.

If we want to have a complete event history of households, we must undertake a survey with a household perspective, as the individual perspective misses some events. However, we will consider the household histories obtained from our "3B" survey, even though this is incomplete.

Before undertaking any analysis, it may be useful to observe the evolution through time of different kinds of households, in order to illustrate our individual-based household concept. Figure 8.1 gives four household histories from our sample.

The first household has a very simple history: from its beginning (a male leaving his parents' home at age 22), it is characterized by regular increases due to marriage and birth of children, to reach its maximum size of 8 members after a duration of 20 years. After that, a very quick decrease occurs due to the departure of children and the wife's death, to end with the male and one child at survey time.

The second household has a somewhat more complex history. First, a household with a single male member from 19 to 32 years old is formed, followed by a return to the parental home for one year. He again leaves his parental home in order to settle independently with a wife, has two children followed by the death of the first one, quickly followed by the birth of two more children. Then the usual decrease follows due to departure of the children.

The third household is started by a marriage, quickly followed by the birth of a child. But eleven years later a divorce occurs, the child remaining with his father. This is followed by a remarriage seven years later, with the birth of a new child, etc. The fourth household is even more complex: twice there is a return to the parental home after the first household formation. Next, a first marriage followed by divorce and departure of the first wife with her child, a remarriage, etc.

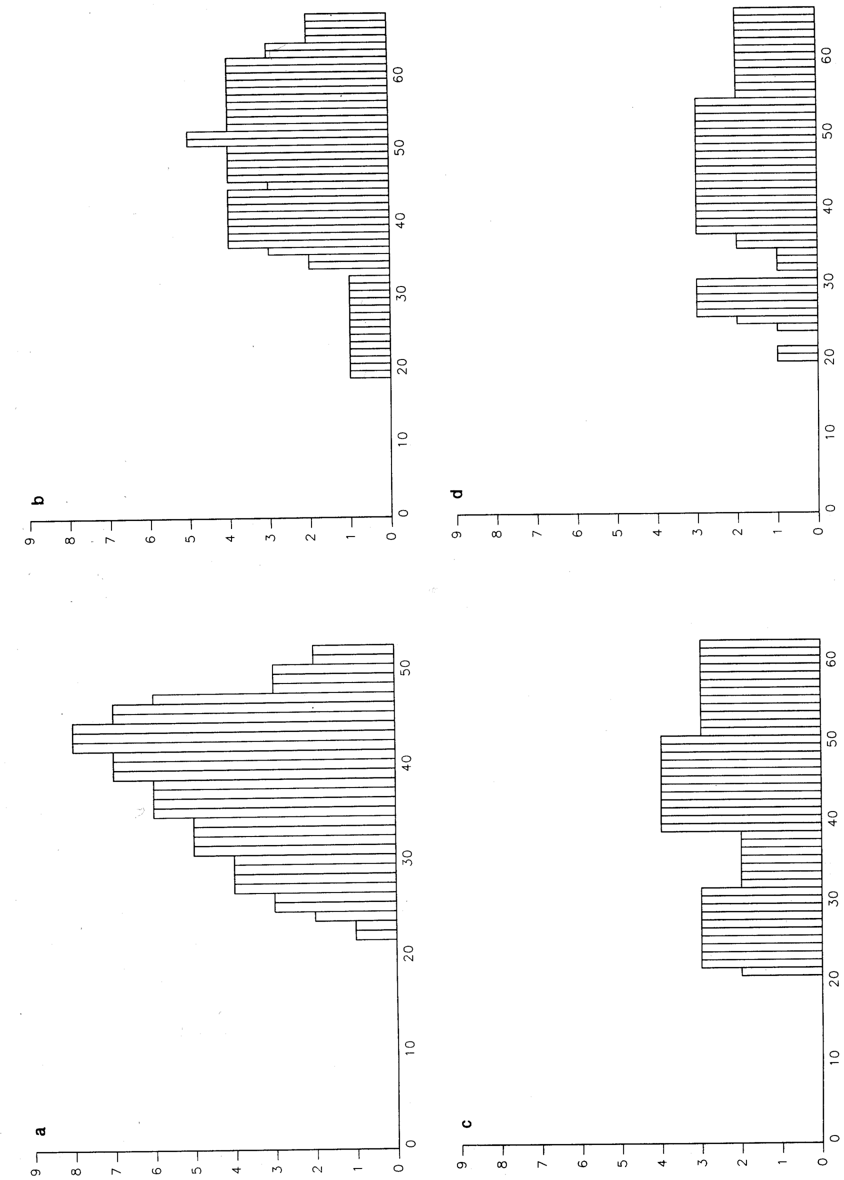

Figure 8.1. Four illustrative household histories
a. Household 1 b. Household 2 c. Household 3 d. Household 4

8.3.2. A Nonparametric Approach to Household History

The maximum number of spells for one household history is 24 for the male sample and 27 for the female sample, leading to a total number of spells of 10,211 for males and 13,596 for females. However, the major part of these histories lies in a more restricted limit: less than 7.5% of the spells are of serial number greater than eight for males (10.0% for females).

With this limit size of eight, we will now consider what are the main transitions. Table 8.1 gives the destination in per cent for each size of origin. As the household instigator may return to the parental home (size zero) and leave it again afterwards, we also have a line and a column with this number. We also have to consider those in a household of a given size who remain in this situation until censoring occurs.

We first observe that the major part of household changes consists of the addition or the subtraction of one member. For households with an origin size of less than four members, the addition of a new member is much more important than subtraction, for both sexes. Afterwards the reverse is true. Differences between sexes appear for origin sizes equal to zero and one. First the females returning to their parents' home are proportionately less numerous than males, and afterwards the probability of attaining size 1 is smaller than the probability of attaining any of the sizes 2, 3, and 4 (0.198, 0.273, 0.233 or 0.201, respectively), which cannot be observed for males. Secondly, the proportion of censored women in household size equal to one is much larger than that for men (46.4% against 23.7%). This is related to the more important number of women observed alone in our survey (mainly widowed ones). These observations lead us to conclude that it is necessary to observe the surveyed individuals according to their sex.

Let us now observe the cumulative hazard rates to higher (or similar) and to lower household size, for males and females separately. If we first consider *waiting times*, we can use formulae (3) and (4) without rank distinction to give an overall view of changes in household size. Figure 8.2 gives these results.

Again, differences between males and females can be observed. When we observe the rates to higher size, the male plot is above the female one, the reverse being true for rates to lower size. However, the main difference between these plots lies in their shape. For movements to higher household size, the rate regularly decreases with duration, for both sexes. It remains near zero from 25 years onward. For movements to lower household size, the plot is more complex. It begins with a low but nearly constant rate until a duration of 15 years. Afterwards this rate increases from 15 to 25 years, with a new decrease coming afterwards. These first results clearly show that we have to consider these two kinds of movements separately.

We can also consider the *duration from first establishment of the household*. Figure 8.3 gives these results.

Table 8.1. Destination sizes of uncensored households for each size of origin (percentages)

Origin size	Destination size 0	1	2	3	4	5	6	7	8	Number censored	Total number
					Males						9841
0	0.0	35.4	25.2	20.9	11.7	3.9	2.4	0.5	0.0	0	206
1	12.9	0.4	75.2	11.2	0.3	0.0	0.0	0.0	0.0	321	1353
2	1.3	12.8	2.8	81.6	1.5	0.0	0.0	0.0	0.0	964	2373
3	0.9	1.4	42.1	1.5	53.3	0.8	0.0	0.0	0.0	378	2416
4	0.5	0.7	3.1	51.4	1.8	42.0	0.5	0.0	0.0	200	1726
5	0.1	0.5	0.5	6.3	53.2	0.9	37.9	0.6	0.0	80	948
6	0.2	0.4	0.9	1.5	6.5	50.4	2.6	37.0	0.5	31	491
7	0.0	0.0	0.4	0.4	1.7	9.4	48.3	3.0	36.8	14	248
8	1.4	0.0	0.0	1.4	0.0	0.0	12.1	79.7	5.4	6	80
					Females						13522
0	0.0	19.8	27.3	23.3	20.1	7.9	0.8	0.8	0.0	0	253
1	13.2	0.7	74.0	12.1	0.0	0.0	0.0	0.0	0.0	755	1626
2	2.3	27.4	2.6	66.5	1.2	0.0	0.0	0.0	0.0	1204	3459
3	1.4	1.2	47.3	1.8	47.7	0.6	0.0	0.0	0.0	356	3342
4	0.9	0.1	4.4	55.0	1.6	37.3	0.7	0.0	0.0	108	2357
5	0.7	0.0	0.4	6.5	54.8	2.0	35.1	0.5	0.0	50	1319
6	0.0	0.1	0.0	0.3	6.3	55.0	2.4	35.6	0.3	18	684
7	0.0	0.0	0.0	0.8	1.1	9.9	51.6	1.1	35.5	5	360
8	0.0	0.0	0.0	0.0	0.0	0.8	14.3	79.8	5.1	3	122

Source: French "3B" survey.

The patterns in Figure 8.2 and 8.3 are quite different. First, the differences between men and women only begin to appear after a duration of 15 years. But afterwards they are similar to those found for the waiting-time approach. Secondly, the movements to lower household sizes do not exhibit a decrease in hazard rates after 25 years, as for waiting times. The rates remain quite constant from 20 to 40 years after the first establishment, and afterwards only show a slight decrease.

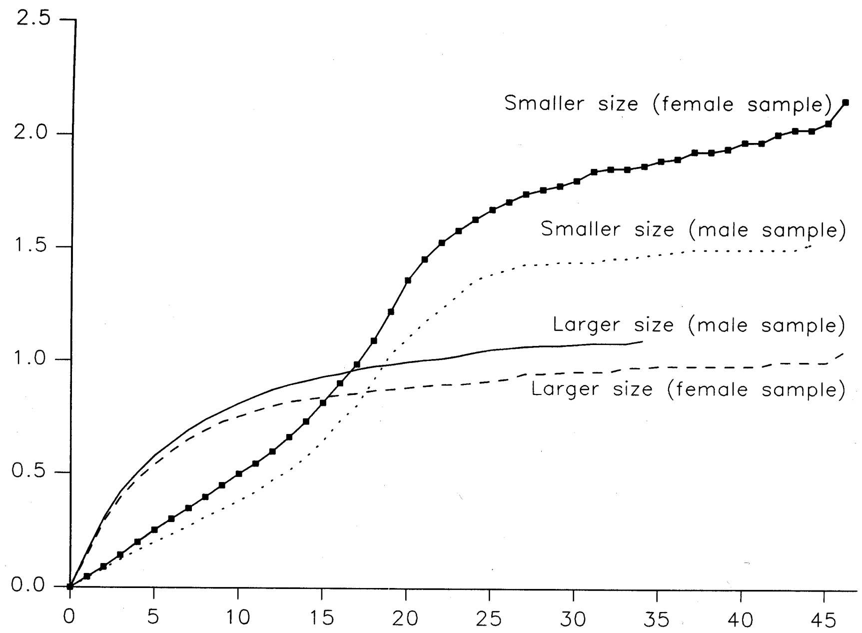

Figure 8.2. Cumulative transition rates to larger (or similar), smaller household size: Waiting time approach

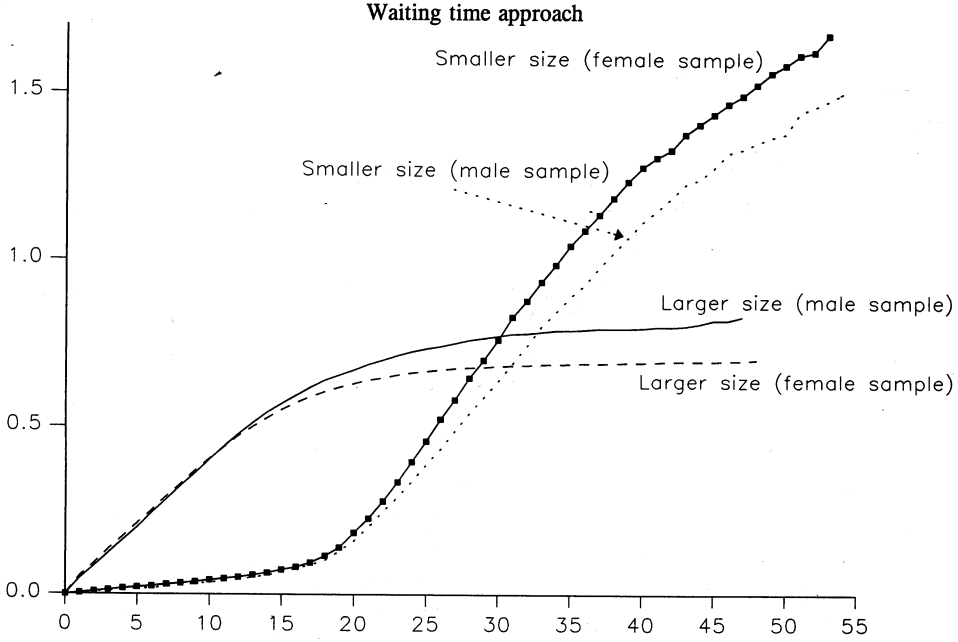

Figure 8.3. Cumulative transition rates to larger (or similar), smaller household size: Time since first establishment approach

The two approaches give us different perspectives on the same phenomena, but for the moment we have no reason to prefer one to the other.

Let us go further and observe, for each successive change the cumulative rates to larger (or similar) and to smaller household size, for males only, considering the different waiting times. Figures 8.4 and 8.5 give these rates.

We see that these hazard rates are highly dependent on the order of the studied transition and that they do not show any typical shape for all groups. For example, the cumulative hazard rates to smaller household size show a linear trend for the first change, an S-shaped trend for the second change, a J-shaped trend for the third and fourth change, and again a linear trend with a steep slope for the higher order changes.

8.3.3. A Semiparametric Approach to Household History

Such a dependence on the order of the studied transition does not allow us to use a simple parametric model for each order. It is the reason why we prefer to use a semiparametric model, which allows an arbitrary baseline hazard for each event order.

To give a simple but clear example of such a semiparametric analysis, we only investigate the effect of two covariates. The first one is time invariant: it is a binary variable of value 1 when the surveyed individual has no diploma, 0 if they do. The second one is time dependent and is a characteristic of the household as a whole: it is also a binary variable of value 1 when the household is home owner.

The β parameter estimates are given in Figure 8.6, for each order of the transition, k, up to eight, for the two destination states ($i=1$, if movements are to lower size households, $i=2$, if they are to sizes equal or higher) and for the two considered models (waiting times or times from the first establishment of the household). We have used square markers where the results were significantly different from zero (at the 5% level).

We can see from these plots that the model which uses waiting times gives much more clear and significant results than the model which uses time from first establishment. This result is particularly clear when we look at waiting times to larger household size.

Being a home owner reduces the probability of increasing ones household size already from the second step, and this reduction increases with the order of the step, when we look at waiting times. The other approach shows significant results only for step 3 and 4, with a less important effect. To have no diplomas increases the probability of moving to a larger household size also from the second step. In that case, the differences between the two models are less clear, with only a non-significant result at step 7 for the model with time from first establishment.

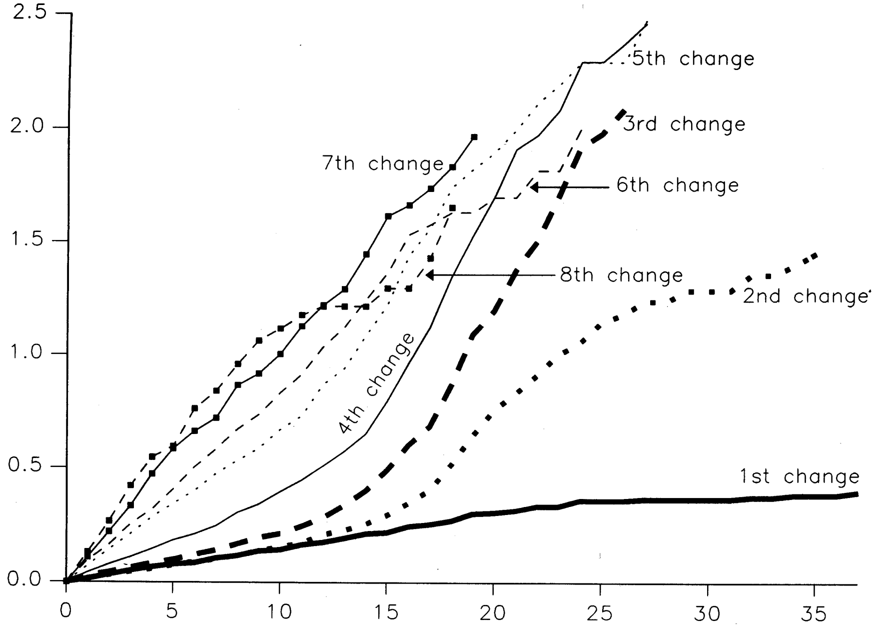

Figure 8.4. Cumulative transition rates to smaller household size: Waiting time approach

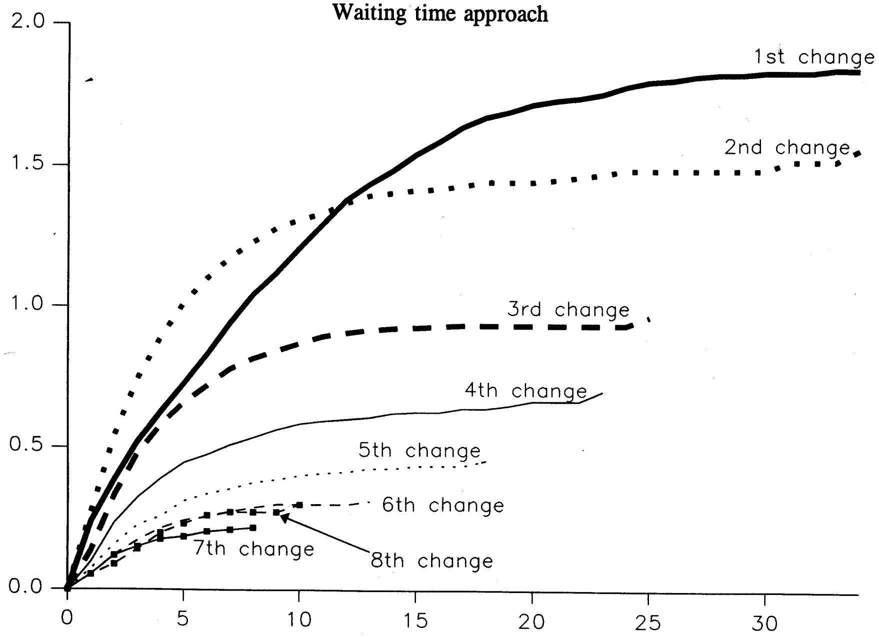

Figure 8.5. Cumulative transition rates to larger household size: Waiting time approach

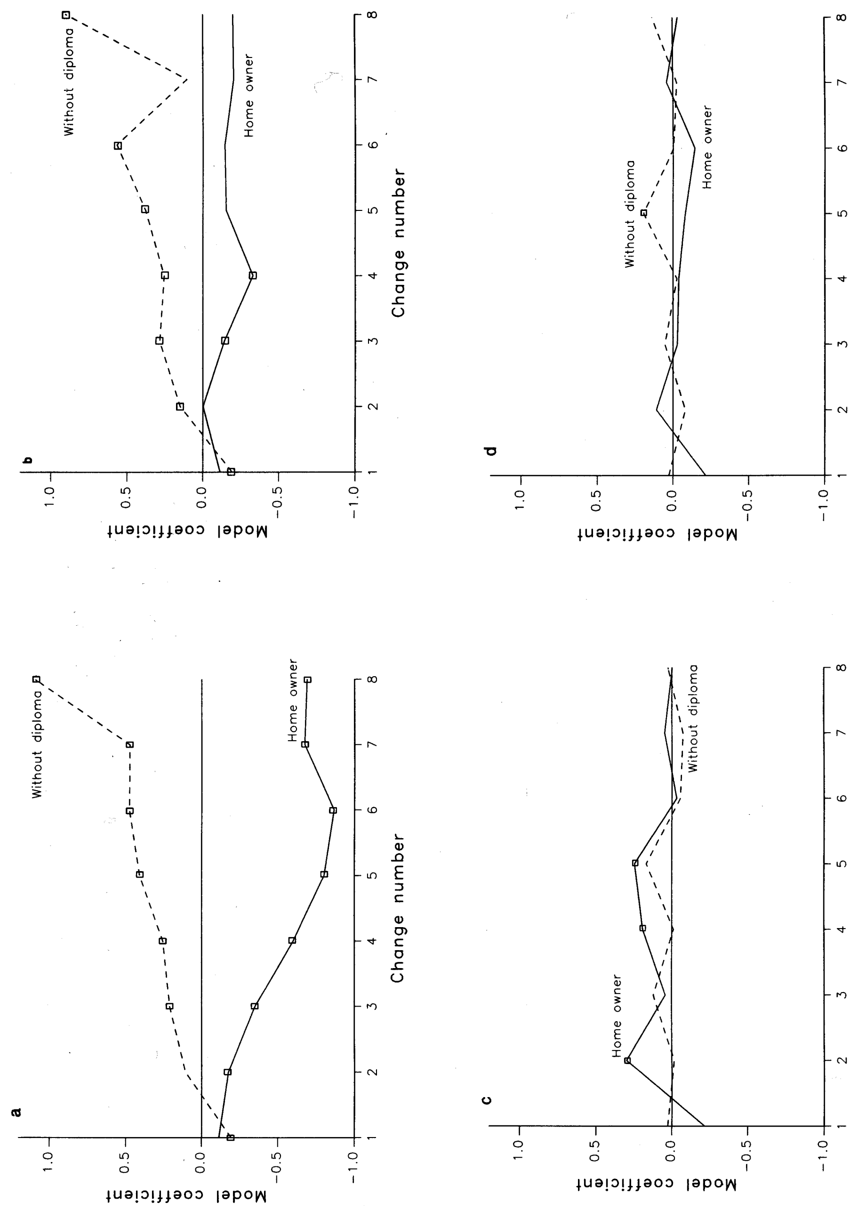

Figure 8.6. Parameter estimates for the two models
a. Waiting time to larger household size b. Time from first establishment to larger household size
c. Waiting time to smaller household size d. Time from first establishment to smaller household size

For movements to smaller household sizes, the waiting time model always gives better results, even if these are less significant than for movements to larger household sizes.

Let us conclude this section with a more detailed discussion of the reasons for preferring the waiting-time model to the common time-origin model for the study of household formation and dissolution. When using a common clock for all episodes, some difficulties of interpretation appear. How can we interpret, for example, the hazard rates for households leaving their i^{th} state after a duration, t, from its formation? Such rates are estimated on the whole population entering the i^{th} state, but only on a part of these households in their i^{th} state at duration t. This risk set has a complex definition, and the interpretation of the corresponding rates is not clear. This difficulty is increased as each household at risk has a different duration of stay in the i^{th} state. Thus, it is not possible to have a clear picture showing the dependence on time of these rates from the beginning of the i^{th} episode. Such an approach, expressed in terms of total observation time, seems to be too narrow to understand such a complex phenomenon as household formation and dissolution.

A more flexible solution is to use a waiting time approach. First, it replaces the common clock by a number of waiting times for each of these episodes. We include all the households entering the i^{th} state before the survey, and we follow them from the beginning of each interval. Such hazard rates have a clear meaning, as the duration variable is conditional on the time of entry in the i^{th} state. It is also easy to introduce a dependence of these rates on the past history, such as the duration since household formation, the occurrence of some related events, etc. Finally, this last approach gives statistical results that are more easy to interpret than the previous ones, and also more significant with the variables used in this paper.

8.4. CONCLUSIONS

In order to study household formation and dissolution, we have developed an event history approach, with an application to our "3B" survey. In this chapter, we tried more precisely to follow a complex entity over time, which may change its size and its composition. We have not considered changes in localization, even if we consider this topic an important one. A further paper will have to take this issue into account.

Such an approach has led us to explore multiple-spell models for duration data. This chapter provides a general formulation of hazard rates. However, for a complete specification, we must define a time scale which measures the time dependence of these rates. Two different clocks have been used which lead to different underlying stochastic processes.

If time is measured as duration of stay since the establishment of the household formation, such a time scale leads to nonhomogeneous Markov processes. A Markov process assumes that the state at some future time depends on its state at the present time, but not on its past states nor on the duration spent in the current state. However, as the process is nonhomogeneous, it may depend on other relevant aspects of its history. We have given a general formalization of the hazard rate in this case (formula (1)). If time is measured as waiting time in the current state, the clock is reset to zero after each transition. Such a time scale leads to a different underlying stochastic process which is semi-Markov. A semi-Markov process also assumes that the state at some future time depends on its state at the present time, but not on its past states. However, it also allows dependence on the duration spent in the current state. Again we have given a general formalization of the hazard rate in this case (formula(2)).

The application of these two approaches to our data set showed that the second model is more suitable to show the dependence of rates on covariates, particularly time-dependent variables. Obviously this is only a first step. We will have to study, in greater detail, the means to test the most suitable model using more precise criteria.

It seems to us that we will have to introduce models with multiple time scales: a first time scale from the beginning of household formation with secondary time scales from the beginning of each transition. Further research will tell us if such a complex model may be estimated from our data sets.

REFERENCES

COURGEAU, D., & E. LELIÈVRE (1989), *Analyse Démographique des Biographies*, INED: Paris.

COURGEAU, D., & E. LELIÈVRE (1992), *Event History Analysis in Demography*, Oxford University Press: Oxford.

COX, D.R. (1972), Regression Models and Life Tables, *Journal of the Royal Statistical Society B*, Vol. 34, pp. 187-220.

DUNCAN, C.J., & M.S. HILL (1985), Conceptions of Longitudinal Households: Fertile or Futile?, *Journal of Economic and Social Measurement*, Vol. 13, pp. 361-375.

HAMERLE, A. (1989), Multiple-Spell Regression Models for Duration Data, *Applied Statistics*, Vol. 38, pp. 127-138.

KALBFLEISCH, J.D., & R.L. PRENTICE (1980), *The Statistical Analysis of Failure Time Data*, John Wiley and Sons: New York.

KEILMAN, N., & N. KEYFITZ (1988), Recurrent Issues in Dynamic Household Modelling. *In*: KEILMAN, KUIJSTEN & VOSSEN.

KEILMAN, N., A. KUIJSTEN & A. VOSSEN (eds.) (1988), *Modelling Household Formation and Dissolution*, Oxford University Press: Oxford.

KLIJZING, E. (1988), Household Data from Surveys Containing Information for Individuals. *In*: KEILMAN, KUIJSTEN & VOSSEN.

KUIJSTEN, A., & A. VOSSEN (1988), Introduction. *In*: KEILMAN, KUIJSTEN & VOSSEN.

MCMILLAN, D.B., & R. HERRIOT (1985), Toward a Longitudinal Definition of Households, *Journal of Economic and Social Measurement*, Vol. 13, pp. 349-360.

9. COMPETING RISKS AND UNOBSERVED HETEROGENEITY, WITH SPECIAL REFERENCE TO DYNAMIC MICROSIMULATION MODELS

Heinz P. Galler

Department of Econometrics and Statistics
Martin-Luther University of Halle-Wittenberg
Grosse Steinstrasse 73
D-06099 Halle
Germany

Abstract. The idea of competing risks provides a useful conceptual basis for multistate models in general, and microsimulation models as a special case. However, some methodological problems arise when the different risks are not independent. Besides causal relations, such dependencies may be caused by unobserved heterogeneity of the units. In this case, not all relevant explanatory variables are included in the model, which may cause stochastic dependencies between different partial processes. Approaches for dealing with this problem are discussed both in a continuous time and a discrete time framework.

9.1. THE NOTION OF COMPETING RISKS

In dynamic microsimulation models, an attempt is made to simulate socio-demographic processes at the level of elementary micro units. Based on a stochastic model of the relevant processes, trajectories are generated for the state of each individual unit in a representative sample of the population considered, using Monte Carlo techniques. The structure of the population and its dynamics are then inferred from the structure of the sample generated by the simulation model. Since comparatively rich information is available on the state of each individual unit in the sample, rather complex model structures can be used and detailed analyses of the processes under consideration become feasible.

On the micro level, many of the characteristics of the units are of a qualitative nature and are represented by qualitative variables. Such variables do not change smoothly over time, but jump at some point in time from one

Household Demography and Household Modeling
Edited by E. van Imhoff *et al.*, Plenum Press, New York, 1995