Emerging issues in demographic methodology[1]

Nico KEILMAN

During the 1980s, the most important innovations in demographic methodology took place in two fields: event history analysis, and non–linear models. Of these two, the latter has had (and will have) the greatest impact on demographic thinking, because it involves a shift in paradigm from analysing the predictable behaviour of linear models, to the investigation of the dynamics of non–linear models, some of which may show unpredictable equilibrium behaviour, even when they are completely deterministic. Event history analysis became widely accepted in demography in the last ten years, and the treatment of unobserved heterogeneity, and that of simultaneously interacting multiple states, may be considered important new contributions to demographic methodology. In addition to event history analysis and non–linear models, this brief review deals with a number of other developments that took place in demographic methodology during the previous decade: age–structured models, models of union formation and household dynamics, and projection methodology. We also discuss an important field which showed no progress: translation methods. The period which is covered in this paper is largely that of the 1980s, and the focus is on quantitative purely demographic models and methods.

1. Helpful comments from Øystein Kravdal, John Dagsvik, Lars Østby, Didier Blanchet and Jan Hoem are gratefully acknowledged. The usual disclaimer applies.

EVENT HISTORY ANALYSIS

Event history analysis became widely accepted in demography in the 1980s. Demographic literature on this topic was virtually non-existent ten years ago, although the technique (also known by such different labels as survival, renewal, hazard, time-to-failure, reliability, life-testing, and intensity regression analysis) had been applied in other fields by then for a number of years already. The article on proportional hazards written by Cox in 1972 is often taken as a starting point, and connections to counting processes were made by Aalen in 1978. In demography, pioneering articles were published by Menken et al. (1981) and Trussell and Hammerslough (1983). Many applications and methodological papers appeared since then, and the technique is now common practice among demographers. Recent overviews of issues of methodology and application that are relevant in demography have been compiled in book form not only in English (e.g. Trussell et al., 1992), but also in French (Courgeau and Lelièvre, 1989) and German (Blossfeld et al., 1986).

The key advantage of event history analysis is the possibility of classification by criteria which apply equally to the individual under investigation, and to a series of individual events, rather than focusing on one event, or on aggregated data. Another advantage is that appropriate techniques exist for the estimation of the parameters of the model (e.g. the effect parameters in a proportional hazards model) in the presence of right censored data. In recent years, two methodological innovations took place: the treatment of multistate models, and the analysis of unobserved heterogeneity. In the review below, we shall deal with continuous-time models only. For a review of issues of unobserved heterogeneity in the context of discrete-time models, see Wrigley (1990).

Multistate models

Many practical applications of event history analysis deal with one particular event: childbearing, death, migration, divorce, entrance into the labour market etc. Such models are known as single-spell models. But in demography, one often encounters situations in which one would wish to analyse two or more events simultaneously. Competing risks models (multiple decrements), for instance those describing cause-specific mortality can be mentioned here, but also more general multistate (or increment-decrement) models, such as those for interregional migration, in which an individual can jump ("migrate") from any of a set of n regions to any other of the (n−1) regions. Such multistate models include competing risks models and single-spell models as particular cases, and the results obtained for the multistate case apply equally well for the simpler models.

In the beginning of the 1980s, the demographic practice was such that multistate models were largely analysed by reducing them to single-spell models, using very restrictive assumptions (e.g. effects of covariates do not differ between events) or focusing on very special cases (semi-Markov models). But recently, more general multistate models have been investigated. Blossfeld et al. (1989, 60–79), Hamerle (1989), Petersen (1991), and Gill (1992) review this issue. The focus is on Cox–type of models which trace the effect of an individual's characteristics on his or her hazards to experience the various events defined as jumps in the state space of the multistate model. For instance, we might wish to compare the hazards for moving between two pairs of states, after correcting for covariates; or compare the effects of the same covariate on different event hazards. The key problem is: how can we estimate a multistate model with a set of covariates specified for each of the events?

The results that have been obtained for multistate models can be summarized as follows. It can be shown that in the case of competing events, the partial likelihood to be evaluated is the product of the partial likelihoods for each of the events (and its accompanying effect vectors) separately. In the case of two events, say, for which the covariates are not linked in any way (i.e. state–specific), this means that just two separate partial likelihoods have to be evaluated; when they are connected, there is still one single joint partial likelihood. Another result is that in the parametric case (when the baseline hazard is modeled in some way), the partial likelihood can be derived, and that it is not much more complicated than in the case of one single event. Finally, expressions for the partial likelihood in the case of an increment–decrement model have been derived, again assuming that the covariates have no elements in common. Hamerle (1989) investigates these matters in the context of time– invariant covariates, and he contends that most of the results he obtains can be generalized to allow for time–dependent covariates as well.

Thus the usual Cox model and its partial likelihood analysis can be used in the more general multistate model rather straightforwardly, although there is an important assumption of independent censoring (Gill, 1992). It should be noted that as an alternative, one can maximize the full likelihood function, for a model much more general than the multistate model – for example one which accounts for unobserved heterogeneity (cf. below) and time–varying covariates as well – see Heckman and Singer (1984).

It will be clear, that with an increasing number of states, the number of model parameters grows exponentially, and this will decrease the precision of the estimates in a given data set. Therefore, in any multistate application the number of states and the number of covariates will be modest in general.

Unobserved heterogeneity

A second innovation that emerged in the 1980s is the treatment of unobserved heterogeneity. Unlike standard regression models, the usual Cox model has no error term. So when we omit important covariates, we can still have very small standard errors for the effect estimators, and we get no warning of the misspecification. When no control is made for unobserved variables, the estimated hazard rate becomes biased towards negative duration dependence.

There are various ways to deal with this unpleasant property, see, for example, Gill (1992), Trussell and Rodriguez (1990), and Yamaguchi (1986). The one most used by demographers is simply to add an "error term" to the (antilog of the) right–hand side of the standard Cox model. Next there are basically two strategies: a non–parametric approach, in which no assumptions on the distribution of ε are required; and one in which one assumes some particular distribution for ε. In the latter case, a gamma, a normal, and a lognormal distribution have been used. However, we often have not much a priori knowledge as to the specific form of the distribution for ε, and the results can be extremely sensitive to one particular choice (Heckman and Singer, 1982, 1984; Hobcraft and Murphy, 1986)[2]. Therefore, the non–parametric approach may be favoured.

Although the potential dangers of neglecting heterogeneity in event history models are well recognized, there is no general agreement as to which of the two approaches mentioned here, or which estimation technique should be preferred. For instance, it is always possible to find many models incorporating unobserved heterogeneity, and one model not incorporating unobserved heterogeneity, that fit equally well (as judged by the unconditional distributions of the endogenous variable), see Trussell and Rodriguez (1990, 118), and, for a proof, Hoem (1990c). External information is necessary to choose between different approaches. Indeed, Petersen (1991, 310) reports successful fitting of a parametric heterogeneity model on the basis of simulated data, for which he knew the correct distribution of ε, whereas Hoem (1990a) has less positive experiences (inclusion of unobserved heterogeneity didn't provide much additional insight, numerical problems with the CTM program especially written for the analysis of unobserved heterogeneity) with a non–parametric approach with relatively few observations and little a priori knowledge about his area of application, i.e. union dissolution of childless women in Sweden. The size of Hoem's data set contrasts with that of Aaberge et al. (1989), who investigate the divorce behaviour of 51,000 Norwegian women (at least ten times as many, approximately, as the number of observations in most other studies) – yet the latter authors share Hoem's conclusion that the covariate effect estimates are rather insensitive to the omission of unobservables[3]. As to the *distribution* of ε, there seems to be some agreement that as long as the baseline hazard is correctly specified, the choice for a particular distribution does not have much impact on the parameter estimates – thus very

2. Part of this sensitivity may be attributed to an incomplete specification of the model, see the remarks below concerning a correct specification of the baseline hazard.
3. Gill (1992) notes that although adding heterogeneity does not much influence parameter estimates, it can dramatically increase their estimated standard errors, especially when the data is not well balanced. Thus, taking account of the fact that not all covariates are included shows that one has much less precise information about the effects of the covariates which have been included. Indeed, a comparison of estimated standard errors the models estimated by Aaberge et al. (1989) reveals that adding heterogeneity increases standard errors for nearly all parameter estimates. In this particular example however, the increase is not so dramatic as Gill suspects: roughly between 10 and 30 per cent.

general baseline hazards may be preferred (Hoem, 1990a, 138; Trussell and Rodriguez, 1990, 117).

Challenges

In spite of recent methodological progress in methods for event history analysis, a number of methodological problems have not been resolved yet in a satisfactory manner. Three issues will be discussed here: the quality of retrospective data, the problem of left censoring, and the treatment of interacting careers.

Most of the data that have been analysed with event history techniques were collected in a retrospective survey. Such an approach may have a profound impact on the *quality of the data*. In particular, it makes these data subject to errors of omission and misplacement – in particular for events which the respondent experienced as unpleasant (divorce, death of an infant), or for which the timing is not clear–cut (start of a consensual union, leaving the parental home), or for which the occurrence took place in the remote past. In addition, to collect reliable retrospective information on individual norms and values is next to impossible.

Poulain et al. (1991) tested the reliability of retrospectively collected data on the timing of marriage, birth of children, children's leaving home, and migration. Three strategies were used: (i) independent interviews of both spouses in a household; (ii) comparison of the two questionnaires for the two spouses in each household, and correction, by the spouses, of possible inconsistencies; and (iii) checks against population register data. Some 500 married persons aged 41–55 in 1986, who had always lived with their first marriage partner, were interviewed. As could be expected, the timing of migration and of the dates at which children leave home was severely misreported (compared with the register), with a bias of more than one year (ante–dates and post–dates taken together) in 8 to 24 per cent of the cases. Marriage dates and birth dates were reported much more correctly. Yet Courgeau (1991) found little sensitivity for his model estimates of migration behaviour as a result of timing misspecification (see also Courgeau and Lelièvre, 1989, 19). Nevertheless, it may be recommended to take the dates at which events are reported in a retrospective survey not too literally, and to recognize that some of the timings may be very fuzzy. Several strategies to deal with such "fuzzy time" concept may be used. One is to introduce fixed (over individuals) time periods in which the event of interest took place, and to study the effects of varying the length of these periods (1, 3, 6, 12, ... months) – see, for example, Klijzing et al. (1988). A second strategy (less ad hoc and statistically more satisfactory than the former) is to write down the probability that an event takes place between two discrete time points t_1 and t_2, on the basis of the hazard rate model. As Petersen (1991, 312) argues, estimation of such a *model with grouped data* often involves maximization of the complete likelihood function, instead of separate factors. An obvious extension of the latter model would be to make the length of the time period t_2-t_1 depend on individual covariates, as recall lapse differs by sex (women report demographic events more accurately than men, see Poulain et al, 1991) and most probably also by age and length of period until interview date.

If one would want to improve the quality of the data, one could rely on a prospective data collection strategy, such as a panel. This reduces memory effects clearly and it facilitates the collection of data on norms and values. Drawbacks, however, are that a panel is relatively costly, that panel drop-out may introduce a bias in the estimates, and that repeated interview may distort the answers, or even the actual behaviour of the respondents (for an overview of these and related issues, see the volume edited by Kasprzyk et al., 1989). It should be noted that the problem of selective panel drop-out may be handled with appropriate weighting procedures for each wave, provided that enough information at the population level is available. However, such *dynamic weights* greatly complicate the estimation of parameters of aggregate change and models of individual behaviour. It should be noted that in retrospectively collected data we have a similar problem of dynamic weights, because, even when the sample is representative for the population at the time the interview was taken, it is not necessarily representative for the population as of five years ago, or ten years ago. The reason is that selective mortality, migration, and other processes of entrance into and exit from the target population may have introduced a bias. Thus this calls for a modelling strategy in event history analysis which takes time varying weights into account.

Another possibility is to analyse population register data. There is no selectivity connected to this approach, but there are other problems. For instance, the *de jure* picture that the register reflects is sometimes only a crude approximation of the *de facto* situation that one is interested in. Furthermore, the number of variables is usually very limited – information on norms and values is clearly lacking altogether. File matching may be helpful for the latter problem.

A second problem connected to event history analysis is that of *left censoring*. In many data collection strategies, we know the entire life history of the individuals. However, this is not always the case, for instance when measurement only applies to a fixed period in the past. In this situation, the state the individual is in at the first date to be recorded is known, but not the length of time that this person spent in that initial state (unless the latter duration was explicitly asked for). Usually it is not possible to calculate the effects of the unknown event history data upon future events, and therefore one often assumes that the previous history of state occupancies is irrelevant for the process to be analysed, or, equivalently, that the hazard function is time-invariant (e.g. Blossfeld et al., 1989, 29; Courgeau and Lelièvre, 1989, 52; Tuma and Hannan, 1984, 131). But frequently this assumption is not very satisfactory, for example when the sample is selected in terms of the endogenous variable – e.g. mortality analysis based on a sample of surviving persons only. Generally, left censoring is considered as a difficult problem.

A very useful iterative method handling left–censored data was recently suggested by Courgeau and Lelièvre (1989, 52–56). These authors estimated survival functions for two processes (birth of a first child, and a second migration) on the basis of an artificially left–censored sample, and compared their results with survival functions estimated for the whole sample (with complete fertility and migration histories). The idea, stemming from Turnbull (1974) is to estimate an expected time of entry into the risk set for left–censored individuals, on the basis of the model which is initially estimated for event histories that are not censored to the left. Application of this technique to the artificially censored data set led Courgeau and Lelièvre to conclude that agreement with results obtained for the complete data set is very close, for various censoring times between 1943 and 1952 (the data were right–censored at 1965). In the case of birth of the first child, estimation on the basis of the sub–sample of complete life histories only, introduces a substantial bias.

The technique proposed by Courgeau and Lelièvre to handle left censored data is, in fact, a special case of the EM–algorithm for computation of maximum likelihood estimators with missing data, introduced by Dempster et al. (1977). The latter algorithm may be applied in parametric and semi–parametric models (Gill, 1992; Little, 1988). Among the advantages of the EM–algorithm are that it produces maximum likelihood estimates, that estimates of standard errors may be obtained in many practical cases, and that convergence properties are known (see the discussion following the article by Dempster et al.).

A final problem to be taken up in the future is that of *parallel life courses and interacting careers*. Individuals experience events in different arenas of life: the family (childbearing, union formation and dissolution), the labour market, the educational system, etc. The life course in one arena may influence that in another one, and vice versa. The best–known example is the reciprocal relationship between childbearing and a woman's labour market behaviour.

In traditional event history models, the focus is on a single event (or perhaps a sequence of similar events in a multistate set–up), and the impact of certain (possibly time–dependent) covariates on the occurrence of this event. The pattern of causation is from the covariates to the event hazard. However, to investigate parallel individual life courses would require a much more symmetric approach. It would imply treating events of type A as covariates for events of type B, and vice versa. Such models, for which a multistate perspective would be a natural way to proceed, are very rare in demography. The empirical studies in Courgeau and Lelièvre (1989, chapters V, VI, and IX) are a notable exception.

The non–parametric model proposed by Aalen et al. (1980) is a useful starting point. These authors investigated, for a medical application, the reciprocal influence between two events A and B by means of a four–state Markov process, where the four states are defined according to whether the two events have occurred or not (not A, not B; A, but not B; B, but not A; both A and B). This framework facilitates such statements as "occurrence of event A accelerates occurrence of event B, but not vice versa". Klijzing et al. (1988) applied this method to the interaction between female labour force participation and childbearing. A next step is to construct a (semi-)parametric model for two or more events. Klÿzing (1991) formulated a semi-parametric hazard model for each of four processes: start of cohabitation, marriage, conjugal union dissolution, and childbearing. Dummy variables were included as covariates in each hazard to account for the possible occurrence of the other three processes. In spite of these and other models which treat interactions, it should be noted that relationships between two or more events for an individual may be so complex that these cannot be reduced to simple interactions (Courgeau and Lelièvre, 1991).

The methods described above take account of the life courses of one and the same individual. However, two or more persons may also influence each other's life courses. For example, the labour market careers of two spouses may be interrelated; the housing and residential careers of a male and a female who intend to start a consensual union clearly depend on each other. Investigations of how the life course(s) of one person are related to that (those) of another person may have great potential for a better understanding of human behaviour (see also Courgeau and Lelièvre, 1991).

NON-LINEAR MODELS

Non-linear models are not a very recent phenomenon in demography. Lee (1974) introduced a formal model which is able to describe, among others, the feedback mechanisms specified by Easterlin. For some parameter values (i.e. when the feedback mechanisms are strong enough), the model produces sustained cycles in births. Contrary to linear stable population models, which produce cycles in births and age structures that vanish as the system reaches the stable situation, non-linear models such as the one proposed by Lee generate a persistent oscillating behaviour, even in the fertility rates. However, only recently the link was made between non-linear models and unpredictable behaviour of the endogenous variables.

The idea, which stems from weather forecasting, is that important parts of reality are inherently non-linear. Some non-linear systems behave erratically in certain critical areas of their parameter space. Such systems may display stable equilibrium behaviour, but once their parameters have surpassed so-called bifurcation points, the behaviour becomes chaotic, identification becomes impossible, and hence the models of such systems cannot be used for prediction purposes.