12 An Attempt to Analyse Individual Migration Histories from Data on Place of Usual Residence at the Time of Certain Vital Events

France during the Nineteenth Century

DANIEL COURGEAU
Institut National d'Études Démographiques, Paris

Our knowledge of the geographical mobility of the French population during the nineteenth century is still somewhat fragmentary. Net intercensal mobility between different *départements* can be estimated from census data, but the results are not always reliable, as the quality of successive censuses varies. Answers to questions on place of birth, which have been asked since 1861, provide more useful information, particularly when the results of successive censuses are compared.[1] However, estimates of mobility obtained in this way can only show the joint effects of different migration flows (e.g. the difference between those who had moved for the first time and those who had returned), and are, therefore, difficult to interpret.

Moreover, census data cannot provide even the simplest type of longitudinal information that would throw light on the nature of population movement, e.g. hazard functions of moving, and the distribution of members of different cohorts by the number of times they have moved. The situation is very different in countries such as Sweden, where population registers have been kept for a long time.[2] In France, there is no direct method of estimating the parameters that characterize population mobility.

We must therefore use indirect sources that provide information about the location of individuals at sufficiently frequent intervals during their lives. Birth, marriage, and death registers contain local information which may be used to construct individual mobility histories. In this chapter I shall examine this information and use it to estimate migration probabilities for members of different French cohorts during the nineteenth century.

[1] Y. Tugault, *La Mesure de la mobilité: cinq études sur les migrations internes*, INED Cahiers 'Travaux et Documents' no. 67 (1973); D. Courgeau, *Méthodes de mesure de la mobilité spatiale* (Paris, 1988) J. P. Poussou, D. Courgeau, and J. Dupâquier, 'Les Migrations intérieures', in J. Dupâquier (ed.), *Histoire de la population française*, iii, *De 1789 à 1914* (Paris, 1988).

[2] B. Wendel, *A Migration Schema*, Lund Studies in Geography, Series B, (Human Geography), no. 9 (Lund, 1953).

Vital Registration in France

Vital registration was transferred from Church to State by the decree of 20–25 September 1792, when the registration of baptisms, marriages, and burials gave way to the registration of births, marriages, and deaths. However, the information collected remained the same as in 1736, so that figures for the periods before and after 1792 are comparable.

The registers normally contain information about the place of residence of individuals: the parish or commune of residence of relatives of the newborn and of deceased persons is registered. As the total number of communes in France amounted to some 37,500, it is clear that very detailed geographical information can be obtained from these registers. The figures for different communes can be aggregated into *départements* (of which there are today 95). In some cases the village or hamlet of residence is shown in the register, but as these smaller places were of a very diverse nature, and because it was not compulsory to record them at registration, I have confined myself to information at commune level.

Information obtained at marriage is more difficult to interpret. In general, the communes of residence of both parties are shown in the register, as well as the commune in which the marriage was solemnized. As will become apparent later, the commune of solemnization is not always the most useful piece of information for our purposes.

I shall use some of the data collected in the survey on social, geographical, and property mobility in France during the nineteenth and twentieth centuries, and would like to thank Professors Jacques Dupâquier and Denis Kessler for having allowed me access to their data. Their study consisted of the male descendants of 3,000 couples, whose place of residence during the period of the First Empire was known. The proportion of couples in the sample from each *département* was the same as that which the population in the *département* bore to the population of France in the Census of 1806.[3] I shall use data relating to the first three cohorts in the sample whose members were born during the eighteenth and nineteenth centuries, and for whom the reconstruction of genealogies was well advanced. As my purpose is essentially methodological, I do not need to have information about the whole sample to validate my method. But, obviously, my results cannot be regarded as definitive until all the data have become available. The methods I suggest can also be applied to other types of survey, such as that presented by Bideau and Brunet.[4]

What Type of Mobility Should Be Studied?

We shall need some definitions to specify the conditions in which mobility occurs.

[3] J. Dupâquier, 'Une grande enquête sur la mobilité géographique et sociale du XIXe et XXème siècles', *Population*, 36 (6) (1981).

[4] A. Bideau and G. Brunet, Ch. 7 above.

First, a date must be selected for the beginning of the study. If this date were taken as the date of birth of an individual, his or her mobility during infancy and childhood could, indeed, be traced,[5] but in practice such a study would reflect the mobility of the individual's parents, rather than his own. In contemporary studies of mobility, therefore, the initial date is generally taken as the date when the individual first leaves his or her parental home. This date is not, however, registered in France, and we are therefore forced to use the date of marriage as our starting point. Marriage is always registered and marks an important stage in an individual's life-course. Its date is linked to mobility, because marriage frequently causes one or other of the parties to move. However, this means that individuals who never marry will be excluded from the sample, although unmarried individuals who have children can return to the sample when their first child is born. This has in fact happened for a very small number of men in our sample, as issue in the female line was excluded.

Having defined an initial point in time, we must next define an initial point in space. Scrutiny of a number of mobility histories suggests that marriages were often celebrated at the place of residence of the parents of one of the spouses, but that the newly married couple frequently established their marital home in a different place. I therefore adopted the following rules:

- If the commune of residence of one of the spouses before marriage was the same as that in which their first child was born, this commune was accepted as the commune of residence immediately after marriage.
- If the commune in which the first child was born was different from that in which either of the two spouses lived before marriage, the commune in which the marriage was celebrated was regarded as the commune of residence immediately after the marriage, and the individual concerned was considered as having moved during the interval between marriage and the birth of the first child.
- If a couple had had children before their marriage, or had never been officially married even though they had produced children, the commune in which the first child was born was regarded as the commune of initial residence.

There is a risk that application of the first rule will result in the omission of some moves that did actually occur between the date of marriage and the date of birth of the first child. Application of the second rule, by contrast, may result in an overestimation of the number of moves between the date of marriage and that of the birth of the first child, for example if the couple had moved to the commune in which their first child was born immediately after their marriage. It is possible that these two types of error will cancel out.

Finally, we must define what we mean by a 'move'. When respondents moved from one commune to another during the interval between two vital events, we

[5] D. Blanchet and D. Kessler, 'La Mobilité géographique de la naissance au mariage', unpublished paper, 1991.

assume that they moved once, and only once, during that interval. We take no account of moves within the same commune. However, as a knowledge of the commune of residence implies a knowledge of the *département* in which the commune is situated, we could just as easily have considered moves between different *départements* only.

It is clear that our definitions will result in the omission of some moves, and will underestimate mobility between communes. An individual who moved several times during the interval between two vital events will be recorded as having moved only once during this period, and may even be classified as never having moved at all,—for instance if, by the date of the second vital event, he had returned to the commune in which he lived at the time of the first event. However, the extent of this underestimation is likely to be very small for the nineteenth century, because at the time fertility rates were higher than rates of geographical mobility. Also, the most mobile section of the population tended to be the young, who moved at a time when vital events were taking place relatively frequently in their lives. By contrast, moves of elderly people, for instance those who returned to live with their children, may be missed rather more often. For this reason, I have focused my analysis on moves that occurred during the period immediately after marriage, at the time of family formation.

An Important Analytical Problem

Because this method detects moves only by reference to two different vital events, a much more complex problem arises. This makes it necessary to construct a formal model of the interaction between vital events and mobility, in order to frame more precise hypotheses that make it possible to estimate the mobility of these populations.

Let $T_1, T_2, \ldots, T_n$ be possible dates of moves after marriage. The T are positive ordered random variates. Let $T^0, T^1, \ldots, T^m$ be possible dates of vital events, with T^0 being the date of marriage, T^1 the date of birth of the first child (etc.), and T^m the date of the individual's death. These are again positive ordered random variates. Thus, we need to study the interaction between two series of events, each of which may occur repeatedly.[6]

Our data are not sufficient to estimate the hazard functions of moves of different orders. However, we can estimate the probabilities of more complex events, which involve both births and moves. Consider, for example, the first move after marriage. We know the individual's place of residence at the date of his marriage, but all we know about his mobility is that he moved either during the interval between his marriage and the birth of his first child, or between two births, or between the birth of his last child and his death, or, indeed, that he did not move at all during his lifetime. In this case, we can estimate conditional probabilities.

[6] D. Courgeau and E. Lelièvre, *Analyse démographique des biographies* (Paris, 1989), pp. 89–90.

Thus, if an individual's first move had occurred between the jth and the $(j+1)$th vital event, where $0 \leq j \leq m$, the probability shown below can be estimated:

$$P(t \leq T_1 \leq t' | T^j = t \cap T^{j+1} = t') = \frac{P(t \leq T_1 \leq t' \cap T^j = t \cap T^{j+1} = t')}{P(T^j = t \cap T^{j+1} = t')}$$

$$= P(t \leq T_1 \leq t') \frac{P(T^j = t \cap T^{j+1} = t' | t \leq T_1 \leq t')}{P(T^j = t \cap T^{j+1} = t')}. \quad (1)$$

The second equation is obtained by applying the theorem of compound probability twice. We have thus obtained the factor by which the probability that we wish to estimate, $P(t \leq T_1 \leq t')$, needs to be multiplied to obtain a probability that we can actually estimate. If the probability of occurrence of the two vital events is the same irrespective of whether or not the individual had moved between these two events,[7] our estimated probability is equal to that which we wish to measure.

Similarly, if we knew that the individual had not moved before his death, we estimate the probability

$$P(T_1 \geq t | T^m = t) = P(T_1 \geq t) \frac{P(T^m = t | t \leq T_1)}{P(T^m = t)}. \quad (2)$$

It is clear that the survival function $P(T_1 \geq t)$ is correctly estimated, provided the probability of dying does not depend on whether the individual had or had not moved before his death.

These results can easily be generalized for the case where the kth move occurs between the jth and $(j+1)$th vital events, and the $(k+1)$th between births of order j' and $j'+1$.

Different Models of Duration of Stay

In the demographic analysis of life histories, different methods are used to model the dates of events, and to study the effect of different individual characteristics on duration of residence in the same commune. We begin with a non-parametric method which will provide a first approximation to the distribution of durations of residence, and which we shall use again later in a simple parametric model, which will provide an overall view.

Non-Parametric Models

We begin with the most general case, when the date of the kth move occurs at a random point in time, T_k, and where we can observe the time that elapses between

7 'Local independence', a concept introduced by T. Schweder in 'Composable Markov Processes', *Journal of Applied Probability*, 7 (1970), and used also by O. Aalen, O. Borgan, N. Keiding, and J. Thormann in 'Interaction between Life History Events for Prospective and Retrospective Data in the Presence of Censoring', *Scandinavian Journal of Statistics*, 7 (1980).

this move and the next $(T = T_{k+1} - T_k)$. To simplify matters, we assume that T_k and T are independent discrete random variables measured in years. We wish to estimate the probability that the kth move will occur in year $t_h (m_h = P(T_k = t_h)$ and $v_i = P(T = t_i)$, the probability that the next move will occur t_h years later $(1 \leq h \leq r, 1 \leq i \leq s)$.

Given the nature of our data, neither t_h nor t_i can be observed directly, but we can place them in relation to certain vital events (marriage, births of children, or death). Our observations, therefore, consist of four dates: $t^j, t^{j+1}, t^{j'}$ and $t^{j'+1}$, such that $t^j \leq T_k \leq t^{j+1}$ and $t^{j'} \leq T_{k+1} \leq t^{j'+1}$, where j, $j+1$, j', and $j'+1$ are the orders of the vital events between which the move occurs. The observations may, of course, be censored on the right, so that all we know is that the $(k+1)$th move could not have occurred before the individual's death, for example when the kth move had occurred before his death $(j' = m)$.

A similar problem has been solved by De Gruttola and Lagakos.[8] I shall briefly outline their solution when the probabilities of occurrence of the vital events do not depend on a move having occurred during the interval between them, and refer the reader to the original paper for more detail.

We establish a series of indicators $\alpha_{h,i}$ which are equal to unity if $t^j \leq t_h \leq t^{j+1}$ and $t^{j'} \leq t_h + t_i \leq t^{j'+1}$, and to zero otherwise. The likelihood for a population of individuals who had moved k times will, therefore be

$$L = \prod_{l=1}^{N} \left(\sum_{h=1}^{r} \sum_{i=1}^{s} \alpha^l_{h,i} m_h v_i \right), \quad (3)$$

where N is the size of this sub-population. The values of m_h and v_i that maximize this likelihood can be estimated by an expectation-maximization algorithm, which begins with an initial value of the parameter and converges rapidly to a value that maximizes the likelihood.[9] It is possible to estimate the variance–covariance matrix as the inverse of the matrix of the negative second derivatives of log L.

Parametric Model

This method, which presupposes a more formal recording of moves, and which has been tested on moves that occurred during the twentieth century, can, however, also be applied to moves that occurred during the past.

There are many studies in which it has been shown that mobility has been very regular over time.[10] The results may be summarized as follows:

8 V. De Gruttola and S. Lagakos, 'Analysis of Double-Censored Survival Data, with Application to AIDS', *Biometrics*, 45 (1989).

9 A. Dempster, N. Laird, and D. Rubin, 'Maximum Likelihood Estimation from Incomplete Data via the EM Algorithm', *Journal of the Royal Statistical Society*, B39 (1977).

10 R. McGinnis, 'A Stochastic Model of Social Mobility', *American Sociological Review*, 83 (1968); D. Courgeau, 'Migrants et migrations', *Population*, 28(1) (also republished as 'Migrants and Migration', *Population*, Selected Papers (1979, 3); R. Ginsberg, 'Timing and Duration Effects in Residence

1. Not all individuals who had experienced a move of a given order would move again; some would remain in their place of residence.
2. The probability of making a new move can in most cases be described by a 'mover-stayer' model or a Gompertz model, both of which give very similar results.

In this study we consider the first move after marriage, but it would be just as easy to apply the method to moves of any order.

The initial date is the date of marriage (or, in some cases, the date of birth of the first child). Accurate information on this date is available from the survey. But the first move can be recorded only from a knowledge of two vital events which place it between the dates t^j and t^{j+1}. If an individual has not moved during his lifetime, his duration of stay will be truncated by death, the date of which, t^m, is known. The likelihood of the observations can, therefore, be written as

$$L=\prod_{l=1}^{p}\left(\int_{t^{j,l}}^{t^{j+1,l}} f_1(t)\,\mathrm{d}t\right)\prod_{i=1}^{q} S_1(t^{m,i}),\qquad (4)$$

where $f_1(t)$ is the probability density function of the dates of the first move after marriage and $S_1(t)$ the survivor function; p is the number of individuals who have moved once, q that of individuals who have died without having moved. The formula assumes that the probabilities of occurrence of vital events are independent of mobility, and conversely.

I shall describe the estimation of this likelihood for the case of a 'mover–stayer' model in detail. We can write[11]

$$f_1(t)=\rho(k\exp(-\rho t)\qquad\text{and}\qquad S_1(t)=1-k[1-\exp(-\rho t)],\qquad (5)$$

where $k=1-S_1(\infty)$ is the proportion of the population who have moved. For the lth individual, we can calculate

$$\int_{t^{j,l}}^{t^{j+1,l}} f_1(t)\,\mathrm{d}t=k\,[-\exp(-\rho t)]_{t^{j+1,l}}^{t^{j,l}}=k\,[\exp(-\rho t^{j,l})-\exp(-\rho t^{j+1,l})],\quad (6)$$

so that the logarithm of the likelihood becomes

$$\log L=\sum_{l=1}^{p}\{\log k+\log[\exp(-\rho t^{j,l})-\exp(-\rho t^{j+1,l})]\}$$
$$+\sum_{i=1}^{q}\{1-k[1-\exp(-\rho t^{m,i})]\}.\qquad (7)$$

Histories and Other Longitudinal Data: II. Studies of Duration Effects in Norway, 1965–1971', *Regional Science and Urban Economics*, 9(4) (1979).

[11] Courgeau and Lelièvre op. cit. in fn. 6.

To estimate the values of k and p that maximize this likelihood, we calculate its derivatives with respect to the two parameters and set them equal to zero:

$$U_1(k,\rho)=\frac{\partial\log L}{\partial k}=\frac{p}{k}-\sum_{i=1}^{q}\frac{1-\exp(-\rho t^{m,i})}{1-k[1-\exp(-\rho t^{m,i})]}=0$$

$$U_2(k,\rho)=\frac{\partial\log L}{\partial\rho}=-\sum_{l=1}^{p}\frac{t^{j,l}\exp(-\rho t^{j,l})-t^{j+1,l}\exp(-\rho t^{j+1,l})}{\exp(-\rho t^{j,l})-\exp(-\rho t^{j+1,l})}\qquad (8)$$
$$-\sum_{i=1}^{q}\frac{kt^{m,i}\exp(-\rho t^{m,i})}{1-k[1-\exp(-\rho t^{m,i})]}=0.$$

This system may be solved by the Newton–Raphson method,[12] which requires calculation of the negative matrix of the second derivatives of $U_1(k,\rho)$ and $U_2(k,\rho)$ with respect to the two parameters. These terms form Fisher's 'information matrix' $I(k,\rho)$. Beginning with an initial vector of parameters $\binom{k_0}{\rho_0}$, we calculate new values ρ_1 and k_1 which satisfy the relation

$$\begin{Bmatrix}k_1\\ \rho_1\end{Bmatrix}=\begin{Bmatrix}k_0\\ \rho_0\end{Bmatrix}+I(k_0,\rho_0)^{-1}\begin{Bmatrix}U_1(k_0,\rho_0)\\ U_2(k_0,\rho_0)\end{Bmatrix}.\qquad (9)$$

This process is continued until the values converge to an acceptable solution of equation (8). The inverse of the matrix $I(k,\rho)$ will yield an estimate of the variance–covariance matrix of the parameters.

These results may be extended to the case where the initial and final values of the period of residence are bounded by two intervals, and where the effect of different individual characteristics can be used as acting multiplicatively on the hazard function.

We obtain the probability of the first move on the assumption that it follows a Gompertz distribution:

$$h_1(t)=\lambda\rho\exp(-\rho t),\qquad (10)$$

where $S(\infty)=\exp(-\lambda)$ stands for the proportion of the population that never moves. The two parameters of the distribution may be estimated in a similar fashion by the Newton–Raphson method.

A Test on Twentieth-Century Data

Before applying the method to historical data, we shall test it on a set of migration histories for which we possess complete information. We use data collected in population registers in Belgium, which contain not only the dates of different vital events, but also information on moves, and on the different places of

[12] Details of the calculations and results are available from the author.

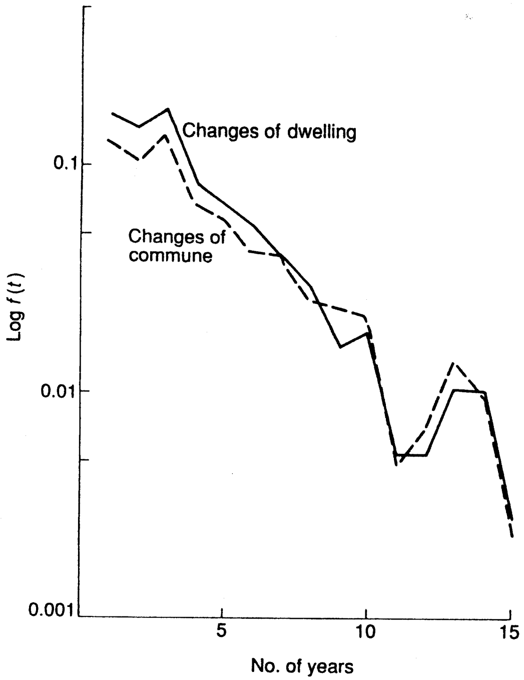

Fig. 12.1 Probability density for changes of dwelling and changes of commune (Belgian registration data)

residence of individuals throughout their lives. These data were collected to test the reliability of individuals' memories when they were questioned about events that had occurred in the past.[13] The sample consisted of 445 individuals born between 1933 and 1942.

I have shown elsewhere[14] that the probabilities of a vital event occurring are affected by change of residence, but that this effect becomes considerably weaker if we confine the analysis to moves between different *départements*.[15] Those who move in connection with a vital event generally move for a short distance only, whereas moves associated with economic or political factors generally involve a longer distance.

[13] M. Poulain, B. Riandey and J. M. Firdion, 'Une expérimentation Franco-Belge sur la fiabilité des enquêtes rétrospectives: l'enquête 3B bis', *Population*, 46(1) (1991); D. Courgeau, 'Analyse des données biographiques erronées', *Population*, 46(1) (1991).

[14] D. Courgeau, 'Interaction between Spatial Mobility, Family, and Career Life-Cycle: A French Survey', *European Sociological Review*, 1 (2) (1985).

[15] D. Courgeau, 'Changements de logement, changements de département et cycle de la vie', *L'Espace Géographique*, 4 (1985).

As our historical study is concerned with moves between communes (or *départements*), we shall compare the results obtained when changes of residence and movement between communes are determined first by direct observation, and secondly by observation of two vital events.

We shall first use a non-parametric method to ascertain which of these models can best be applied to moves that have been observed directly. On Fig. 12.1 the logarithm of the density function of moving is plotted as a function of duration of stay. The relation is approximately linear, and the 'mover–stayer' model can be used to study both changes of residence and movements between communes. In this case log $f(t)$ is in fact a linear function of duration of stay. It would also have been possible to use the logarithm of the instantaneous rate as a function of duration of stay to show that Gompertz's model can be used to yield an empirical adjustment that, though not as good as that given by the previous model, is still acceptable.

We shall next compare the cumulative distribution of the random variate T estimated non-parametrically, again depending on the method of observation and the type of movement. Changes of residence and moves between communes have been plotted in Fig. 12.2. It will be seen that the greatest discrepancy between the results yielded by the two methods is found in respect of changes of residence for durations of stay of between 4 and 12 years. We shall see later that this result is in accordance with expectation. For moves between communes the results are

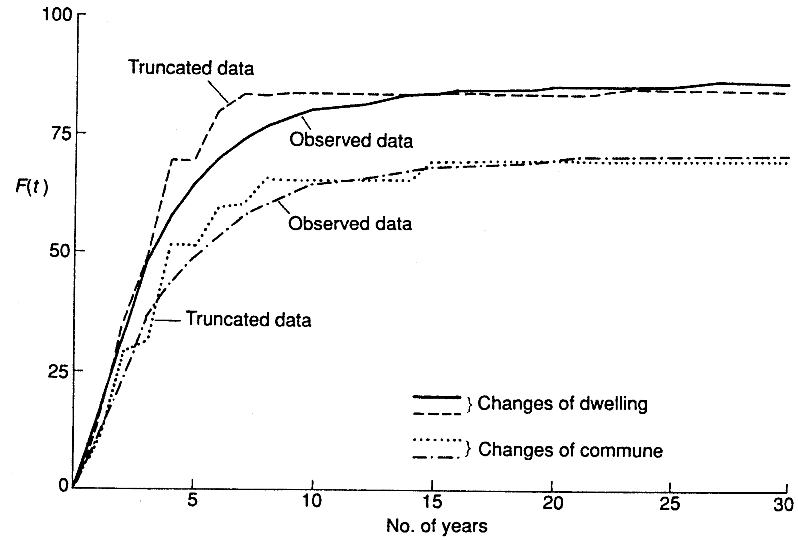

Fig. 12.2 Distribution of first move after marriage between dwellings and communes, depending on whether they are based on observed or truncated data (Belgian registration data)

TABLE 12.1 *Estimates of different parameters and their confidence intervals at the 5 per cent level for the 'mover–stayer' and Gompertz models–proportion of stayers, logarithm of the likelihood of moving between lodgings, and communes following marriage–by whether the data were observed completely or truncated by births or exit from observation*

Data	'Mover–stayer' model				Gompertz model			
	k	ρ	Proportion of stayers $(1-k)$	Log L	λ	ρ	Proportion of stayers $(\exp(-\lambda))$	Log L
Change of dwelling								
Observed completely	0.860 (±0.035)	0.261 (±0.029)	0.140	−912.08	2.020 (±0.290)	0.133 (±0.025)	0.133	−915.28
Truncated	0.858 (±0.035)	0.317 (±0.043)	0.142	−448.52	1.980 (±0.401)	0.159 (±0.047)	0.138	−454.53
Change of commune								
Observed completely	0.713 (±0.043)	0.222 (±0.027)	0.287	−1,014.53	1.264 (±0.159)	0.147 (±0.024)	0.283	−1,015.39
Truncated	0.712 (±0.043)	0.243 (±0.035)	0.288	−569.08	1.258 (±0.188)	0.158 (±0.035)	0.284	−570.23

Note: Figures in parentheses are confidence intervals.
Source: Belgian registration data for 445 individuals, born between 1933 and 1942.

much less discrepant, and the small differences observed relate to a much shorter period. By contrast, both for changes of residence and for moves between different communes, the probabilities of not moving at all are practically unchanged by truncation. (0.15 for changes of residence, 0.30 for movements between communes).

What are the interactions between mobility and fertility when changes of residence are considered? It is possible to calculate the probability of occurrence of a first birth, depending on whether the individual had or had not moved before its occurrence. If the probabilities were the same, this would support the view that the occurrence of a birth is independent of the phenomenon studied—in this case the first move. To test this hypothesis, I show in Fig. 12.3 the integrated hazard of a first birth, separately for cases in which the individual had, and those in which he had not, moved before its occurrence. For durations of three years or less, fertility is independent of mobility; for longer periods, the probability of a first birth is larger for individuals who have moved. In this case we could rewrite equation (1), where $t = 0$, because we are considering the interval between marriage and first birth:

$$P(T_1 \leq t' | T^1 = t') = P(T_1 \leq t') \frac{P(T^1 = t' | T_1 \leq t')}{P(T^1 = t')}. \tag{11}$$

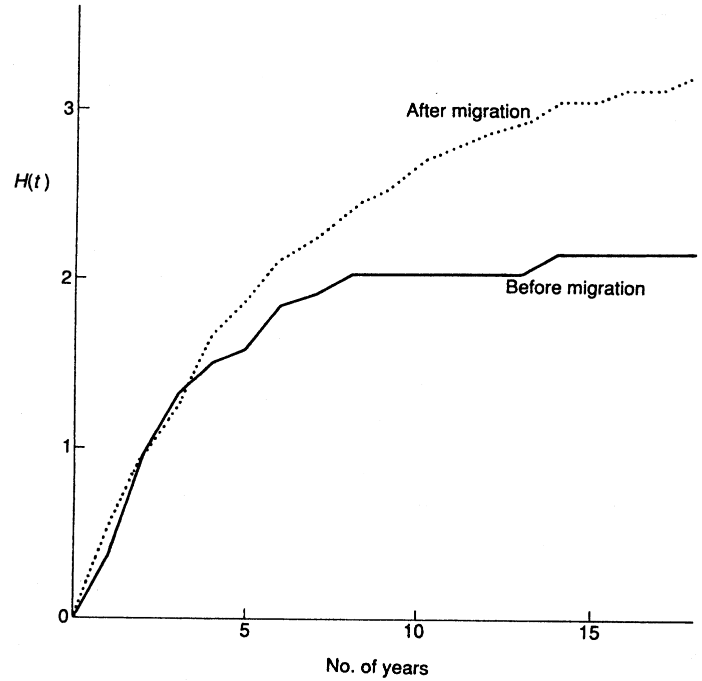

Fig. 12.3 Cumulative first birth rates before and after migration (Belgian registration data)

It is clear that in this case the cumulative distribution function obtained for durations exceeding three years lies above that obtained from observation, for $P(T'=t'|T_1 \leq t') > P(T^1 = t')$. To complete the demonstration, we would need to examine births of different orders, but for the present our result will suffice.

We next consider estimation with parametric models. In Table 12.1 are shown values of parameters estimated from direct observation and from information about vital events by using both a 'mover–stayer' and a Gompertz model. Results are given both for changes of residence and for moves between communes.

It can be seen that, in estimating the proportion of stayers, neither the method of observation nor the model used makes a great deal of difference, either for changes of residence or for moves between communes; and the estimates are very close to those that had been obtained by non-parametric methods.

In contrast, the parameter ρ, which in the 'mover–stayer' model is a hazard function independent of duration of stay for those members of the population who, according to the model, would be expected to move in the future, does depend on the method of observation when changes of residence are studied. It is much larger for the indirectly observed data, and confirms what we have already learned from the non-parametric analysis. However, if we are interested only in moves between communes, no significant difference is found between results from data that had been directly and indirectly observed.

This confirms that, at present, moves linked with family events normally take place within the same commune. It is likely that this result would have held even more strongly in the past, when the commune was clearly the area for familial interaction for the majority of French people.[16]

Application to Moves in the Past

As mentioned earlier, we do not yet have a complete list of family files for the survey of social, geographical, and property mobility in France during the nineteenth and twentieth centuries. For this study I have been able to use data for 225 families, born mainly in Normandy, northern and eastern France, and the Lyons region. The data are not representative of the French population as a whole, and no substantive conclusions should be drawn from this first application of the method.

I have limited this study to the first three generations of each family, as data for more recent cohorts are still incomplete. Members of the first of these cohorts will have been born approximately at the end of the eighteenth century, those of the second during the first quarter of the nineteenth century, and those of the third cohort during the second quarter of the nineteenth century. These data make it possible to obtain results that relate to the mobility of the population during the

[16] P. Ogden, 'Migration, Marriage and the Collapse of Traditional Peasant Society in France', in P. White and R. Woods (eds.), *The Geographical Impact of Migration* (London and New York, 1980).

nineteenth century, a period which, in France, coincided with the Industrial Revolution.

I present data only for first moves between communes after marriage, for the sample is too small to study moves of higher orders, or moves between *départements*.

Figure 12.4 shows the cumulative distributions of first moves after marriage estimated by a non-parametric method for the three cohorts studied. The probability of a first move is fairly low for the first cohort (0.294), but increases considerably for the second (0.484) and more slowly for the third (0.527). The curves are irregular, because the sample is small and the data are truncated. This suggests that it would be useful to adjust the data by means of a parametric function.

In Table 12.2 we show the estimated parameters using both a 'mover–stayer' and a Gompertz model. Again, the proportion of non-movers is very close to that obtained with the non-parametric model, and the difference between the results obtained from the two parametric models that we have tried is small. Again, the parameter ρ, which in the 'mover–stayer' model is a probability independent of duration of stay for the population (which, according to the model, would be expected to move at some time in the future), increases in successive cohorts.

During the nineteenth century, two aspects of movement between communes can be seen. First, we note a strong increase in the proportion of individuals in the second cohort born during the first quarter of the century who will migrate,

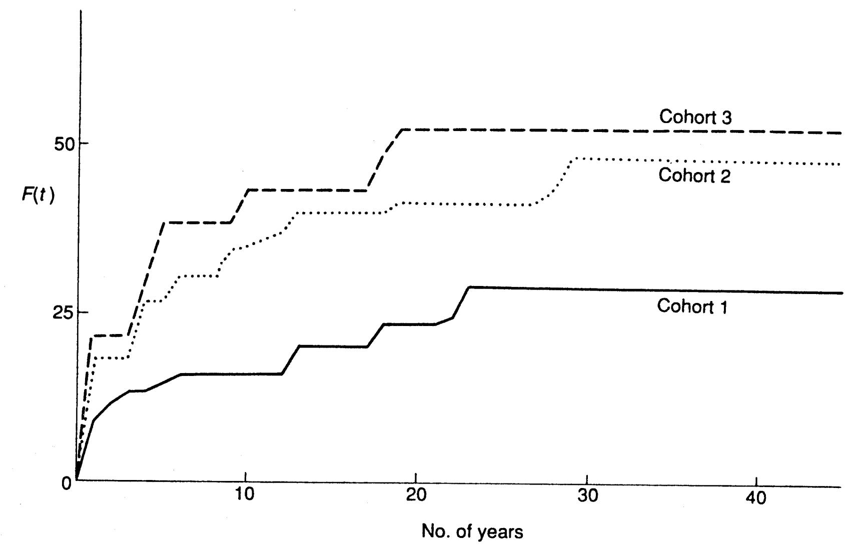

Fig. 12.4 Distribution of first moves between communes after marriage in nineteenth-century France: truncated data for the first three cohorts

TABLE 12.2 *Estimates of parameters and their confidence intervals at the 5 per cent level, 'mover–stayer' and Gompertz models: proportion of stayers and logarithm of the likelihood of moving between communes after marriage for the first three observed cohorts*

Cohort (No. of observations)	'Mover–stayer' model: k	ρ	Proportion of stayers $(1-k)$	Log L	Gompertz model: λ	ρ	Proportion of stayers $(\exp(-\lambda))$	Log L
1 (225)	0.312 (±0.068)	0.093 (±0.034)	0.688	−235.65	0.380 (±0.103)	0.083 (±0.034)	0.684	−234.84
2 (283)	0.481 (±0.063)	0.143 (±0.034)	0.519	−360.54	0.669 (±0.129)	0.119 (±0.034)	0.512	−358.02
3 (263)	0.530 (±0.068)	0.170 (±0.041)	0.470	−338.72	0.778 (±0.153)	0.135 (±0.039)	0.459	−335.25

Note: Figures in parentheses are confidence intervals.

compared with the first cohort. Secondly, the probability of moving increases from the first cohort to succeeding ones for those individuals who will actually move.

Discussion and Conclusion

The object of the present exercise is primarily methodological. I wished to test the feasibility of estimating hazard functions of moving from information about the commune of residence of individuals at a time when they experienced different vital events. In the absence of population registers, this is the only information available for nineteenth-century France. I have been successful in devising methods that, under certain conditions, make it possible to estimate such hazard functions.

The chances of a first move after marriage could be underestimated by this method because of return migration that occurred between two vital events. Because the probability of a birth after marriage was high in nineteenth-century France, such omissions are likely to have been rare during the 15 years following marriage. At higher durations, however, they may become relatively more important, but as they only apply to a very small number of moves, the hazard function of these events is unlikely to be unduly affected. For moves after the first, these errors become more important depending on the extent to which such moves are concentrated at ages when the probability of a birth is lower. The risk of short periods of residence being missed increases, and there may be confusion about the order of the move. We shall see later how to deal with this important risk of bias.

I have also shown that an estimate of mobility from data relating to residence at the time when particular vital events occur risks giving biased results, if the chance of occurrence of the vital event were to depend on previous moves by the individual concerned. I have shown that this is the case for twentieth-century data that relate to moves between different places of residence, but that changes of commune or *département* have a much weaker effect on vital events. In contemporary France, moves caused by family growth are normally confined to within the same commune, and there are strong reasons for believing that this was even more true of nineteenth-century France.

We can see two improvements to methods of analysis. The first is to use data from retrospective French surveys and to extend the comparisons that I have begun here. Thus, in the 'three life histories' survey,[17] information about the family and migration history of a sample of individuals makes it possible to compare the results obtained from direct information on mobility with those obtained by indirect methods, by artificially truncating the data. This will make it possible to estimate the risk of omitting moves after the first. Unfortunately, these surveys relate to cohorts born during the twentieth century, and it would be

[17] Courgeau, op. cit. in fn. 13.

necessary to assume some degree of continuity and similarity with the situation relating to mobility during the nineteenth century to make these data useful.

The second possibility is to look at data for other European countries in which population registers have been kept throughout the nineteenth century. I am thinking particularly of Sweden, where work of this kind has already been carried out.[18] With the help of these registers, the migration and reproductive behaviour of members of cohorts born during the nineteenth century can be described, and tests could be devised which would artificially truncate the data. Comparisons between the results from directly and indirectly observed data could then be applied to data for France.

In spite of these non-negligible risks of bias, the French data throw a completely new light on mobility during the nineteenth century. The study that I have begun should be continued by taking account in the survey of different characteristics of individuals that could affect their mobility, e.g. mobility of parents in childhood, age at marriage, occupation, number of siblings, birth order, number of children born before moving, etc. Different types of mobility may also be distinguished: moves between *départements* or regions, moves to large towns (Paris, Lyon, Marseille), migration between countryside and small towns, return migrations, etc. These data may be linked to those obtained by INED on urbanization in France, and should make it possible to place individuals within the urban environment of their period, and to use individual data for a better understanding of the process of urbanization in France. This paper shows that such a study would be feasible provided certain conditions are fulfilled.

[18] Wendel, op..cit. in fn. 2.

13 Some Applications of Recent Developments in Event History Analysis for Historical Demography

IAN DIAMOND, RHODRI DAVIES, PETER EGGER
Department of Social Statistics, University of Southampton

Introduction

Demographers have always been interested in change, and the study of the social, economic, and biological determinants of the timing of various demographic phenomena or 'events' has rightly been accorded great importance in the demographic literature. There have been two main focuses: the age at which a certain event or 'milestone' occurs, and the length of time between events—or what has become known as event-history analysis. Examples of the former include ages at weaning, first intercourse, marriage, first birth, migration, and death, while the latter include birth intervals and the time that elapses between different episodes of illness. In the past, these processes were largely studied by univariate techniques—Hajnal's development of the Singulate Mean Age at Marriage is an example[1]—but during recent years three factors have led to a huge industry in event-history analysis.

First, substantial changes in the timing and pattern of fertility throughout most of the world have led to a demand for substantive research on the factors that influence these changes. Examples include work on the influence of the growth in women's education and labour force participation on fertility in developed countries, or of the factors that influence duration of breastfeeding, or later marriage in less developed countries. Secondly, these substantive questions have led to the development of techniques for the collection of retrospective event histories which have resulted both in major improvements in the accuracy of event-history data and in their availability; for example, in most countries there now exists some large data set which contains intricate details of the dates of marriage, childbearing, labour force participation, and contraception by women of reproductive age. Thirdly, there have been great improvements in the range of statistical techniques available for the analysis of event-history data. Traditional methods of survival analysis, such as the life table, have been used for some time, but during recent years major developments on the statistical and computational

[1] J. Hajnal, 'Age at Marriage and Proportions Marrying', *Population Studies*, 7 (1953), pp. 111–36.