FROM THE GROUP TO THE INDIVIDUAL: WHAT CAN BE LEARNED FROM MIGRATORY BEHAVIOUR*

By taking the analysis of population renewal as subject of study, demography has deliberately positioned itself in a context that privileges the analysis of aggregate values: the 'stock' of individuals is modified both in terms of volume and structure by the 'stream' of births, deaths and all the other events which intervene in the movements of populations. Understanding these streams involves dealing with behaviour patterns which enable the individual to emerge from under the mass of statistics: the increasing use made of data from surveys based on biographical reconstitutions reveals more clearly than ever the existence of a microdemography alongside, or in complement to, a macrodemography.

*This dual aspect is not specific to demography and, in fact, affects all of the social sciences. Mark Blaug**, for example, when referring to the coexistence of a microeconomy and a macroeconomy, underlines the fact that this creates "a kind of intellectual schizophrenia in which the techniques of either approach do not entirely cover the domain of the other. This situation is far from satisfactory and economists have been trying to bridge the gap between the consumer and the function of global consumption or between the investor and the factors involved in investment aggregates. The bridging of this gap, however, is only partially completed and the economics student must be prepared to use two different tool boxes".*

In this paper, Daniel COURGEAU*** *brings his contribution to this bridge-building by drawing a parallel between the analysis of individual data and aggregate data concerning migrations.*

Depending on whether the human sciences deal with group or individual behaviour, the assumptions, objectives, formulations and characteristics considered will be very different. In the present paper, we shall

* Translated by MEDIANE (Paris).

The author would like to thank Nico Keilman (Central Statistics Office, Norway) and Nicole Tabard (CNRS, France) for their remarks on a first draft of the present paper.

** Mark BLAUG, «Economics», in *Encyclopedia Universalis*, Macropedia, vol. 6, p. 270, ed. 1973-1974.

*** INED.

Population, An English Selection, 7, 1995, 145-162

be looking at the modelization of those behaviour patterns which help clarify the problems encountered when passing from one level to the other. These problems will first be presented in a general form, and then more fully explained using the example of migratory behaviour.

The first approach takes the *aggregate level* and seeks to define the global characteristics of the group in order to understand its behaviour. It will thus attempt to represent a complex reality through a simplified schema which brings out the interrelations between the group's main characteristics. What is more, this approach will often be based on hypothetical individual behaviour types which are impossible to verify given that the data used are of an aggregate nature. Thus, migration models, for example, will explain migratory streams assuming that the behaviour of migrants is influenced by various characteristics in the departure and arrival areas and by the physical or social distance separating these areas. This approach has been followed over a long period of time [Young, 1924] and automatically operates at a *macrogeographical* level: it is the characteristics of the areas which alone influence the movements of individuals. Those researchers working in the field of structural, contextual, ecological, etc. analysis will here recognise similarities with their own problematics.

The second approach operates at the *individual level* and will use, not global characteristics, but those elements in each person's event history to explain his/her behaviour. Corresponding models have recently been developed in the fields of sociology [Tuma and Hannan, 1984], economy [Lancaster, 1990] and demography [Courgeau and Lelièvre, 1989]. These models will explain, for example, migrants' behaviour rather than migratory streams [Courgeau, 1976; Sandefur and Scott, 1981]: these are at a *microgeographical* level. To do so, they use event history survey data which provide details on a person's professional life-course, migratory life, etc. They will then link the probability of an individual experiencing a given event, such as migration, to his/her past history, characteristics and the information he/she possesses on the different areas of arrival.

Integrating these two approaches is made difficult by the fact that the former predicts collective behaviour using *group characteristics*, whilst the latter predicts individual behaviour using *event history characteristics*. An additional problem arises given that the first approach works on a *cross-sectional basis*, whereas the second works *longitudinally*: taking the case of migrations, the cross-sectional approach predicts a stream of migrants, whilst the longitudinal approach involves the probability to migrate throughout an individual's life-course. As a result, the attempts to integrate these two approaches are still very limited [Sanders, 1992, pp. 51-56].

In the present paper, we shall try and link the group to the individual within the context of migratory models. In order to begin this delicate task, we thought it necessary to give as full an account as possible of the conditions under which the comparison is to be made. We do, however, consider that the results obtained can be generalised to the broader problem which we raised in the introduction.

So as to examine in greater detail how these two levels are intertwined, we shall first define the models which are most often applied to each of them separately. These models will then be simplified, as they will be used initially to analyse out-migrant streams. These models will then be reformulated using the same characteristics for the two levels and a theoretical comparison will be undertaken of the results they produce. We shall apply these formulae to data from the 'Family, occupational and migration histories' (*Triple biographie*) survey, known as the '3B'survey. This makes it possible to obtain a fuller definition of the relationship between the two levels of aggregation.

I. – A classic formulation of the two approaches

We shall first present the two types of models, giving a precise definition of their objectives and their formulation. No attempt will yet be made to interrelate the characteristics used.

Aggregate-level analysis The following analysis will explain the interregional migratory streams on the basis of various characteristics and by multiple regression models. These models may be formalized in different ways: they are most often multiplicative, but may also be additive or of a more complex nature [Stillwell, 1975].

Although the approach is a macrogeographical one, the selection of the characteristics for the model is made according to what individual migratory behaviour patterns are imagined to be [Greenwood, 1975; Puig, 1981]. Three main types are often distinguished.

The first type measures the push factors present in the region of departure or, conversely, the reasons for remaining there: unemployment rate, wage levels, the proportion of individuals of a given educational status, etc. It can thus be supposed that a high unemployment rate and low wage levels encourage out-migration. The second type represents the pull factors of the area of destination or, conversely, the reasons for not going there. The characteristics may be the same as the previous ones, now measured for the area of destination with the opposite effect. Thus, a low unemployment rate and a high wage level may make a region attractive. Finally, a third type of variable measures the interaction between the area of departure and the area of arrival. Various types of distances have been proposed in order to do so: physical distance, distance measured by the intervening opportunities between the two areas [Stouffer, 1940], etc.

The choice of these characteristics and the explanatory content they are given clearly show that there is an underlying model of individual behaviour. Furthermore, the effects of these characteristics are often explained using personal terms: it is said, for example, that the appeal of a higher salary is what pushes an individual to migrate, whereas the model shows only the relation between the total number of out-migrants and the average salary in the area of destination.

The general mathematical formulation of these models can be expressed as:

$$M_{ij} = f(X_i, Y_j, I_{ij}, \varepsilon_{ij}) \qquad [1]$$

where M_{ij} is the stream observed between i and j, f being a function whose form may vary, X_i and Y_j are vectors of the characteristics of the area of departure and arrival, I_{ij} measures the interaction between these areas and ε_{ij} is the residual which depends on all the other characteristics not taken into account.

These models are often estimated on the basis of the migratory streams, measured by a census question on the place of residence at a prior date. The various characteristics can either be provided by a previous census or come from different sources (annual declaration of earnings, etc.). These characteristics should normally be measured at the beginning of the period, but as data for this date are often not available, many authors do not retain this assumption [Puig, 1981].

Individual-level analysis This analysis assumes that an individual, α, experiences a series of migratory moves throughout his/her life between r regions within a given territory. In the equation below, T_k^α is the duration separating the $(k-1)^{\text{th}}$ migratory move from the k^{th} move, I_k^α is the area of residence for this period. These are random variables to which the theory of probabilities can be applied. This makes it possible to modelize the hazard rate for the k^{th} move defined as the probability to migrate from i to j between t and $t + dt$, conditioned by the stay at i until time t, divided by dt :

$$m_{ijk}^\alpha(t) = \lim_{dt \to 0} \frac{P(T_k^\alpha < t + dt, I_{k+1}^\alpha = j \mid T_k^\alpha \geq t, I_k^\alpha = i)}{dt} \qquad [2]$$

It is then possible to introduce the effect that many individual characteristics, measured at time t or some previous point in time, have on the hazard rate. These characteristics are often obtained through retrospective surveys which retrace the various life histories: professional activity, educational status, tenancy status of a dwelling, etc. [Courgeau and Lelièvre, 1989]. These characteristics are often measured using binary variables, with a value of 1 if the individuals possess them and 0 if not. Discrete variables may also be introduced (salary, number of children, etc.) or even more complex variables. Finally, these surveys also make it possible to obtain certain information on the individual as to the possible areas of in-migration: former places of residence, location of relatives or friends, etc. This then means that the migratory hazard rate for the individual, α, can be expressed in the following general form:

$$m^{\alpha}_{ijk}\,(t\,;x^{\alpha}_{i}\,(t)\,,\,y^{\alpha}_{j}\,(t)) \qquad [3]$$

where $x^{\alpha}_{i}(t)$ corresponds to his/her characteristics in the area i and $y^{\alpha}_{j}(t)$ to the information that he/she has on j. A more precise formulation of this rate may be done entirely parametrically [Courgeau, 1985a] or semiparametrically [Courgeau and Lelièvre, 1986]. In the first case, the parameters are estimated using the maximum likelihood method, whereas partial likelihood must be used for the second [Cox, 1972]. The estimations of variances and covariances are simultaneously available.

II. – Establishing a relation between the two formulations

To link these two models, additional assumptions are needed. A common time must first be introduced, then the preceding formulae have to be simplified and, finally, the models used must be specified. It is only when this work has been completed that the interrelations between the two will become evident.

A common time for simplified models The above-presented models use different time bases. The aggregate model is applied cross-sectionally to a period of time, whereas the individual model is applied longitudinally over time periods between successive migratory moves.

We shall here postulate a short time period, such as one year (t_0, t_0 + 1), during which we shall measure the migratory stream M_{ij} and the hazard rate m_{ij}, which is assumed to be constant over this period. It is no longer necessary to introduce here the migration rank k, but this then means that the start date, t_α, for the period under study must be known. The length of stay at time t_0 will be $(t_0 - t_\alpha)$.

We shall then proceed to simplify the models by working on the outmigrant streams from the areas, leaving aside here the detailed streams that exist between the areas. In this case, it is no longer necessary to introduce the characteristics Y and I from formula [1], nor the characteristics $y(t)$ from formula [3], as their effect on the out-migration stream is assumed to be negligible in comparison to the effect of the departure area characteristics.

Let us now define the model that we shall be using for the individual data. We have chosen a semiparametric model, which avoids having to specify the baseline hazard. If we work with discrete time, this model can be expressed in the following way [Kalbfleisch and Prentice, 1980, p. 36], depending on the length of stay $(t_0 - t_\alpha)$:

$$m_{i\cdot}(t_0 - t_\alpha\,;x_i^\alpha) = 1 - [1 - m_{i\cdot}^0(t_0 - t_\alpha)]^{\exp x_i^\alpha \beta} = 1 - [1 - m_{i\cdot}^0(t_0 - t_\alpha)]^{\exp \sum_{k=1}^{n} \beta_k x_{ik}^\alpha} \quad [4]$$

where $m_{i\cdot}^0(\cdot)$ is a baseline hazard identical for all the individuals, but dependent on the length of stay for each of them $(t_0 - t_\alpha)$, x_i^α is a line vector of individual characteristics (x_{ik}^α) and β a column vector of parameters to be estimated (β_k). These parameters are, in this case, presumed to be independent of the areas in which the individuals reside. To simplify this relation, we shall assume that the baseline hazards have a low value, which enables the simplified expression:

$$m_{i\cdot}(t_0 - t_\alpha\,;x_i^\alpha) \approx m_{i\cdot}^0(t_0 - t_\alpha)\exp\sum_{k=1}^{n} \beta_k x_{ik}^\alpha \quad [5]$$

When the variables are binary, $\exp \beta_k$ can be interpreted as the relative risk of an individual with the characteristic k, compared to those individuals who do not have that characteristic. We shall see later how to estimate the parameters β.

The aggregate models normally used are multiplicative [Stillwell, 1975]. For the sake of consistency with the individual model chosen, we shall use an exponential-type multiplicative model:

$$\frac{M_{i\cdot}}{N_i} = \exp(X_i\gamma) = \exp(\gamma_0 + \sum_{k=1}^{n} \gamma_k X_{ik}) \quad [6]$$

where $M_{i\cdot}/N_i$ is the rate of out-migration from the area i, X_i the line vector of the various 'macro' characteristics of this area (X_{ik}) and γ the column vector corresponding to the effect they produce (γ_k). If non-centred variables are used, it is necessary to introduce a constant term, γ_o, into this regression. The different parameters can then be estimated by moving over to logarithms, which makes the model linear.

Theoretical relations between the two models

On the basis of the individual out-migration probabilities, the expected value of the annual number of out-migrants from the area i is estimated:

$$E(M_{i\cdot}) \approx \sum_{\alpha=1}^{N_i} m_{i\cdot}^0(t_0 - t_\alpha)\exp x_i^\alpha \beta = \sum_{\alpha=1}^{N_i} m_{i\cdot}^0(t_0 - t_\alpha)\exp\sum_{k=1}^{n} \beta_k x_{ik}^\alpha \quad [7]$$

It should be recalled that the individual characteristics, x_{ik}^α are measured here using binary variables or discrete variables. The mean values can thus be computed for the individual characteristics observed in the area i:

$$X_{ik} = E(x_{ik}^{\alpha}) = \frac{\sum_{\alpha=1}^{N_i} x_{ik}^{\alpha}}{N_i} \qquad [8]$$

A relation is thereby created between the individual characteristics and the characteristics of the areas in which the individuals reside. The aggregate level model can thus be rewritten in the following form:

$$\frac{M_{i.}}{N_i} = \exp\left(\gamma_0 + \sum_{k=1}^{n} \left[\gamma_k \frac{\sum_{\alpha=1}^{N_i} x_{ik}^{\alpha}}{N_i}\right]\right) = \exp\left(\gamma_0 + \frac{1}{N_i} \sum_{\alpha=1}^{N_i} \sum_{k=1}^{n} \gamma_k x_{ik}^{\alpha}\right) \qquad [9]$$

The above relation has been obtained by inverting the order of summation on the characteristics (k) and the individuals (α).

A comparison between the two total numbers of migrants estimated using formula [7], at individual level, and formula [9], at aggregate level, should enable us to see whether simple relations exist between the parameters. In fact, it can be seen that, if we assume that $\beta_k = \gamma_k$, and a simpler notation is adopted:

$$e_i^{\alpha} = \exp \sum_{k=1}^{n} \beta_k x_{ik}^{\alpha} = \exp \sum_{k=1}^{n} \gamma_k x_{ik}^{\alpha} \qquad [10]$$

then a simpler relation is obtained:

$$\sum_{\alpha=1}^{N_i} m_{i.}^{0}(t_0 - t_{\alpha})\, e_i^{\alpha} = N_i (\exp \gamma_0) \left(\prod_{\alpha=1}^{N_i} e_i^{\alpha} \right)^{\frac{1}{N_i}} \qquad [11]$$

Noting that $m_{i.}^{0}(t_0 - t_{\alpha})$ contains practically no information on the values of e_i^{α} [Cox, 1972], these latter can be considered as two independent random variables. This relation can then be rewritten:

$$\overline{m}_{i.}^{0} \sum_{\alpha=1}^{N_i} e_i^{\alpha} = (\exp \gamma_0) \left(\prod_{\alpha=1}^{N_i} e_i^{\alpha} \right)^{\frac{1}{N_i}} \qquad [12]$$

where $\overline{m}_{i.}^{0}$ is the mean of the values $m_{i.}^{0}(t - t_{\alpha})$ *taken for all the individuals in the area i*. It can thus be seen that, as soon as the population N_i comprises more than one individual, the assumption $\beta_k = \gamma_k$ is no longer valid. It is not, therefore, possible to simply relate the parameters estimated at 'micro' level for the individual characteristics to the parameters estimated at 'macro' level for the characteristics of the areas.

On the other hand, if area characteristics regarding the probability to migrate are introduced at individual level, the situation will be different. It has already been mentioned that, at aggregate level, the behavioural model was assumed to be one in which an individual was sensitive to area characteristics: high unemployment in a region is supposed to push an active or unemployed person to a region with a lower unemployment rate. In the light of this, the relation [7] can be rewritten introducing area, rather than individual, characteristics:

$$\sum_{\alpha=1}^{N_i} m_{i\cdot}^{0}\,(t_0 - t_\alpha)\exp\left(\sum_{k=1}^{n}\beta_k \frac{\sum_{\alpha=1}^{N_i} x_{ik}^{\alpha}}{N_i}\right) = \sum_{\alpha=1}^{N_i} m_{i\cdot}^{0}\,(t_0 - t_\alpha)\exp\left(\frac{1}{N_i}\sum_{\alpha=1}^{N_i}\sum_{k=1}\beta_k x_{ik}^{\alpha}\right) \qquad [13]$$

In this case, what clearly appears are the identity conditions of the formulae [9] and [13], whatever the individuals' characteristics may be:

$$\begin{cases}\beta_k = \gamma_k \\ E\,(m_{i\cdot}^{0}\,(t_0 - t_\alpha)) = \exp\gamma_0\end{cases} \qquad [14]$$

The above relation again assumes that $m_{i\cdot}^{0}\,(t_0 - t_\alpha)$ contains no information on the parameters β.

In a final step, it can be of interest to simultaneously introduce the individuals' characteristics and those of the area in which they reside, into a single individual-level model. A comparison of the parameters obtained for the 'macro' and 'micro' characteristics should make it possible to clarify the links existing between their effects on migration.

III. – Testing of the different models on the '3B' survey data

We now need to use a file containing data from the same survey in order to make an empirical comparison of the results produced by the different models. First of all, the file to be used must be more precisely defined mentioning where its weak points lie for the purpose of the test.

Sample and migrations considered The '3B' survey has collected the family, occupational and migration histories of 4,602 individuals aged from 45 to 69 in 1981. To carry out a precise statistical analysis, it would have been necessary to break this population down into generations of men and women, whose departures from various regions in France are observed during the course of the same year. The sample size, however, makes it impossible to enter into such detail.

We shall here limit ourselves to working on all the generations observed, distinguishing nonetheless separate behaviour patterns for men and women. There are, in fact, considerable behavioural differences between men and women as far as migration is concerned [Courgeau, 1985b]. For the purpose of this test, we shall take the female population, which is proportionally the largest (2,552 individuals), and study the out-migration from French regions. Given the limited total number of individuals observed, it was necessary to group together certain regions: Table 1 shows the composition of the regions and groups of regions chosen.

TABLE 1. – REGIONS AND GROUPS OF REGIONS SELECTED FOR THE STUDY

Regions	Code
Ile-de-France	1
Champagne-Ardennes and Picardie	2
Haute-Normandie and Basse-Normandie	3
Centre	4
Burgundy	5
Nord-Pas-de-Calais	6
Lorraine	7
Alsace and Franche-Comté	8
Pays de la Loire	9
Brittany	10
Poitou-Charentes	11
Aquitaine	12
Midi-Pyrénées	13
Rhône-Alpes	14
Limousin and Auvergne	15
Languedoc-Roussillon	16
Provence-Alpes-Côte d'Azur and Corsica	17

The small sample size also made it impossible for us to work on a short one-year period. We have been obliged to group the migratory moves occurring between 1950 and 1960, taking into account the first changes of region made during this period, from the place of residence in 1950. We have naturally excluded those women who were living abroad in 1950 (228), since the conditions of international migration greatly differ from internal migration.

In addition to this lack of precision concerning the observed cohort and the period of migration, there is also the problem of choosing the characteristics to be taken into account. On the basis of previous analyses of changes of *département* [Courgeau, 1985], we have chosen five characteristics for the women in question as observed in 1950. Again, due to the small sample size observed, we are unable to test the effect of a greater number of variables. The first characteristic chosen is marital status: we assume that married women have a reduced mobility. The second and third characteristics are linked to the tenancy status of a dwelling: living at their parents' home or home-owners. We suppose here that the women living with their parents are less mobile than the tenants of rented accommodation or those women living in free accommodation, taken as a reference group, yet more mobile than women who are home-owners. The fourth characteristic is related to activity: we suppose that economically inactive women are more easily mobile than active women. The final and fifth characteristic is occupational: a woman farmer or agricultural worker should be far less mobile than other active women. We shall simultaneously attempt to see whether the corresponding 'macro' characteristics have a similar effect on the migratory streams by calculating the percentages of women who have these different characteristics in each of the areas.

It must be realised at this point that the regional results we obtain are not to be considered as representative ones, but as totally comparable elements, since they have been produced using a single data source. We are concerned here with comparing the different levels of aggregation and not with giving a precise view of regional out-migration, as the sample size does not permit this.

Estimation of the parameters and reformulation of the models

We shall now take the period spanning 1950 to 1960 and observe those women who have made no migratory move between the date of their last installation, which may be their date of birth, and the year 1950. The sample is said to be 'left censored' as those moves prior to 1950 fall outside our scope of observation. Using such data poses a great many problems and necessitates making certain assumptions [Courgeau and Lelièvre, 1989, pp. 52-56]. If, however, it is assumed that the probability of migrating depends uniquely on the time that has elapsed since the last installation and on the characteristics of each individual in 1950, we are able to estimate the β and the $m_{i\cdot}^{0}\,(t_0 - t_\alpha)$.

Assuming, in effect, that the individual α migrates after a period of time $(t - t_\alpha)$ spent in his/her place of origin, the population for which there is a risk of undertaking this migratory move will include all the individuals whose duration of stay is greater than or equal to $(t - t_\alpha)$ and who were observed as having a duration of stay in 1950 less than or equal to $(t - t_\alpha)$. It can thus be seen how this model differs from classic models of event history analysis, since the population at risk will not decrease with the duration of stay, as in the classic model, but will, on the contrary, begin by increasing with the duration of stay, at least as long as this latter is not too long.

Despite this, the probability of migration for the individual in question can still be computed, conditionally on the population at risk:

$$\frac{m^{0}_{i\cdot}(t_0 - t_\alpha)\exp x^{\alpha}_{i}\beta}{\sum_{l \supset R_i} m^{0}_{i\cdot}(t_0 - t_\alpha)\exp x^{l}_{i}\beta} \qquad [15]$$

where R_i is the total number of individuals at risk after 1950 and having spent $(t - t_\alpha)$ years in the region of origin. This clearly gives rise to a partial likelihood as proposed by Cox [1972] since the numerator and denominator can be simplified by $m^0_{i\cdot}(t_0 - t_\alpha)$. However, given that the population at risk does not decrease uniformly, traditional software applications do not permit the parameters β to be estimated.

In order to do so, we have used a multiple-spell regression model which introduces the time that has elapsed since an initial point in time which is the same for all spells [Hamerle, 1989]. The first spell, which is of no interest to us, ends in 1950. The second spell, which serves here for our estimation, begins in 1950 and continues until the next migratory move. Of course, if this does not occur before 1961, the interval is censored. This method enables the parameters β and the function $m^0_{i\cdot}(t_0 - t_\alpha)$ to be estimated correctly(1).

As before, the parameters of the aggregate-level model are estimated by applying multiple linear regression on the logarithms.

Given that the observation period is ten years, formula [7] needs some modification. The probability of an individual migrating between 1950 and 1960 must be estimated according to the durations of stay. The number of expected migrants may thus be expressed as:

$$E(M_{i\cdot}) = \sum_{\alpha=1}^{N_i}\left[S^0(1950 - t_\alpha)^{\exp x^{\alpha}_{i}\beta} - S^0(1960 - t_\alpha)^{\exp x^{\alpha}_{i}\beta}\right] \qquad [16]$$

where $S^0(t)$ corresponds to the survivor function in a given region.

Formula [9], on the other hand, remains unchanged. This means that the relations between β and γ will be more complicated than those given in formula [14]. Practical application will nonetheless show that they continue to have an approximately linear relationship.

Results

We shall first test the proximity of the results obtained at individual and aggregate level using the 'macro' characteristics of the different areas. These results are reported in Table 2. Four characteristics clearly have a significant effect at an individual level on the probabilities to migrate from the areas, whereas only three have an effect at aggregate level. The proportions of married women and women living at their parents' reduce this probability, which is in line with the assumptions made earlier regarding individual characteristics. The proportion of inactive

(1) To do this, we used the TDA programme written by Gotz Rohwer.

TABLE 2. – ESTIMATION OF THE PARAMETERS FOR THE 'MICRO' 'MACRO' MODELS AND OF THEIR STANDARD DEVIATION, INTRODUCING THE DIFFERENT CHARACTERISTICS MEASURED AT 'MACRO' LEVEL

Characteristics	'Micro' model		'Macro' model	
	Estimated parameter	Standard deviation	Estimated parameter	Standard deviation
Proportion of married women	– 6.987***	1.910	– 5.879***	1.920
Proportion of women living at parents'	– 3.107**	1.434	– 2.404	1.452
Proportion if inactive women	5.472***	1.830	4.706**	1.934
Proportion of women working in agriculture	3.468***	0.788	2.897**	0.980
Constant	–		0.936	1.365
Maximum log-likelihood	– 17 886.12			
R^2	–		0.514	

*** Result significant at 1% level.
** Result significant at 5% level.

women increases this probability. The proportion of women working in agriculture has an effect contrary to that expected. We shall come back to this point later.

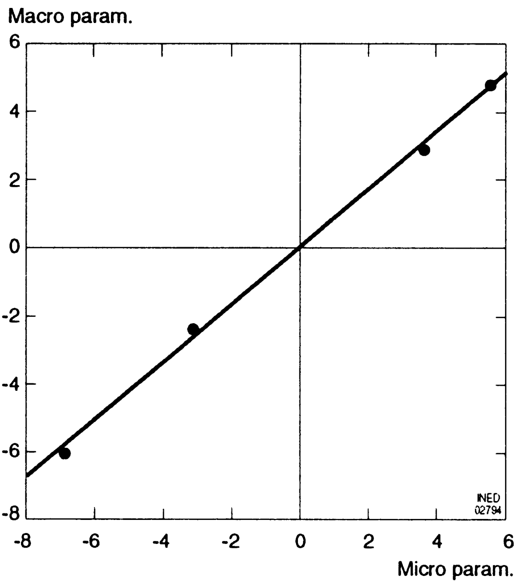

Figure 1. – Relationships between parameters corresponding to the characteristics of the different areas, estimated at individual level (micro) and aggregate level (macro)

Figure 1 shows that the various parameters are linearly linked, but that individual-level estimators are always higher in terms of absolute values than those made at aggregate level. On this basis, the theoretical result advanced earlier stands confirmed: although the parameters are no longer equal when dealing with long periods, they remain acceptably proportional to each other. Moreover, the standard deviations are very close to each other, with the individual level values being consistently lower.

Figures 2 and 3 show respectively, at individual and aggregate level, the comparisons between the numbers of expected out-

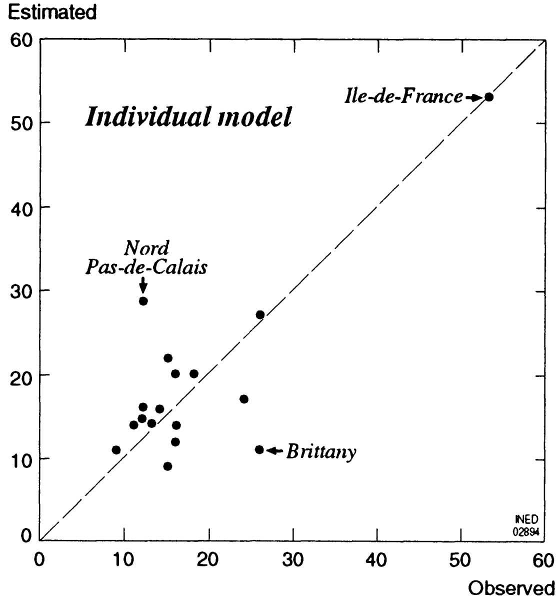

Figure 2. – Out-migrants estimated by the individual model compared to observed out-migrants

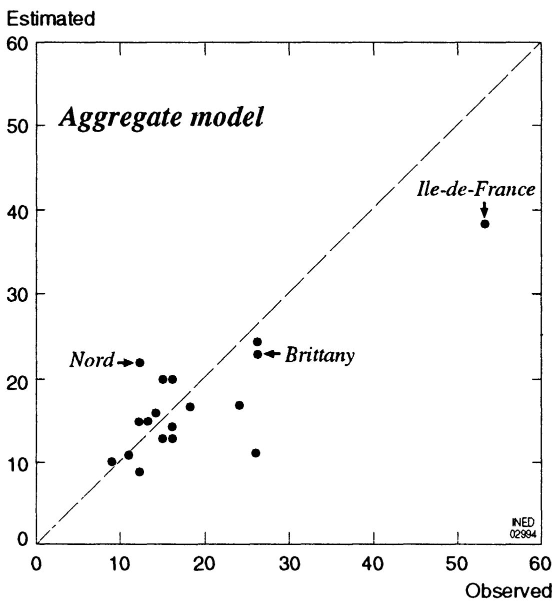

Figure 3. – Out-migrants estimated by the aggregate model compared to observed out-migrants

migrants and the numbers observed in each region. Although the 'micro' model accurately predicts the number of out-migrants from the Ile-de-France region, it grossly overestimates the number for the Nord-Pas-de-Calais region, and largely underestimates that for Brittany. The out-migrants for the other regions are more or less correctly predicted. The 'macro' model, on the other hand, largely underestimates the number of Ile-de-France out-migrants. It overestimates the out-migrants from Brittany, as does the 'micro' model, and makes more or less correct predictions for the other regions, even a little better than the 'micro' model.

Therefore, these two types of models do not give completely identical results, if an attempt is made to calculate the total numbers of out-migrants.

Let us now examine the results obtained at individual level when each respondent's characteristics are introduced. Table 3 provides an estimation of the corresponding parameters.

TABLE 3. – ESTIMATION OF THE PARAMETERS OF THE 'MICRO' MODEL AND OF THEIR STANDARD DEVIATION, INTRODUCING THE INDIVIDUAL CHARACTERISTICS

Characteristics	Estimated parameter	Standard deviation
Married	– 0.568***	0.127
Home-owner	– 1.106***	0.364
Inactive	0.339***	0.130
In agriculture	– 0.188	0.181
Maximum log-likelihood	– 17 874.23	
*** Result significant at 1% level.		

It can be seen that three characteristics have an effect upon the probability of migrating. These results are consistent with what we assumed earlier, even when the estimated parameters are not significantly different from zero: the lower probability of migration for married women and women home-owners is confirmed; the higher probability of inactive women migrating is confirmed; although the results are not significant for women working in agriculture, we find a lower probability of migration, as for those women living at their parents'. No linear relation appears between the parameters of the models involving area characteristics and those involving individual characteristics. It would seem that these two types of characteristics influence the probabilities of migrating in a very different way.

Let us now see whether the simultaneous introduction of both types of variable produces a more informative model than the former one. The results are shown in Table 4.

This model explains mobility much better than the former one: the addition of four aggregate characteristics causes the maximum log-likelihood to increase by 12.56, which corresponds to a χ^2 of four degrees of freedom equal to 25.12. On the other hand, the parameters estimated for

TABLE 4. – ESTIMATION OF THE PARAMETERS FOR THE 'MICRO' MODEL AND OF ITS STANDARD DEVIATION, SIMULTANEOUSLY INTRODUCING INDIVIDUAL AND AGGREGATE CHARACTERISTICS

Characteristics	Estimated parameter	Standard deviation
Proportion of married women	– 6.772***	1.918
Proportion of women living at parents'	– 3.393***	1.430
Proportion of inactive women	5.383***	1.832
Proportion of women in agriculture	3.785***	0.802
Married	– 0.552***	0.128
Home-owner	– 1.125***	0.364
Inactive	0.304**	0.131
In agriculture	– 0.333*	0.185
Maximum log-likelihood	– 17 861.67	

*** Result significant at 1% level.
** Result significant at 5% level.
* Result significant at 10% level.

the two series of variables are very close to those obtained when the aggregate or individual characteristics are introduced separately (Table 3). This goes to confirm the assumption whereby both types of characteristics influence differently the probabilities of migration. Since the variances and covariances between the various estimated parameters are available, their correlations can be computed (Table 5).

TABLE 5. – CORRELATIONS BETWEEN INDIVIDUAL CHARACTERISTIC AND AGGREGATE CHARACTERISTIC PARAMETERS

Characteristics	Correlation
Married* Proportion of married women	0.100
Living at parents'* Proportion of women living at parents'	0.124
Home-owner* Proportion of home-owners	0.107
Inactive* Proportion of inactive women	0.100
In agriculture* Proportion of women in agriculture	0.240

All the correlations are clearly low, generally around 0.10, which indicates a relatively high independence between these 'macro' and 'micro' characteristics. The highest correlation is that estimated between the parameters which measure the effect of being active in agriculture and the effect of the percentage of women in agriculture on the probability of migrating (0.24). These are, in fact, the parameters which undergo the greatest change when the two categories of variables are introduced simultaneously. It is of particular note that the effect of working in agriculture becomes significant at the 10% level and is inverse to that of the percentage of women in agriculture.

This apparent paradox may be explained by subdividing the population at risk into two distinct groups according to the proportion of women in agriculture in the different regions. These results are presented in Figure 4.

A first observation is that women in agriculture consistently have a lower probability to migrate, regardless of what proportion they represent within a given region. This confirms the negative parameter obtained at individual level (–0.333). Yet, at the same time, it can be seen that, for women working in agriculture and those working in other sectors alike, the probability of migrating increases when the proportion of women in agriculture increases. This results in a positive value (+3.785) for the parameter related to this aggregate variable. The danger then clearly arises of inferring certain assumptions about individual behaviour on the basis of the results obtained at aggregate level: a high proportion of women in agriculture leads to a higher probability of migration for all categories of women, partly because of the scarcity of non-agricultural employment within the region. This, however, in no way implies that women in agriculture have more opportunity to migrate than other women: at individual level, quite the contrary is observed. Moreover, regardless of the region of origin, this result is always confirmed.

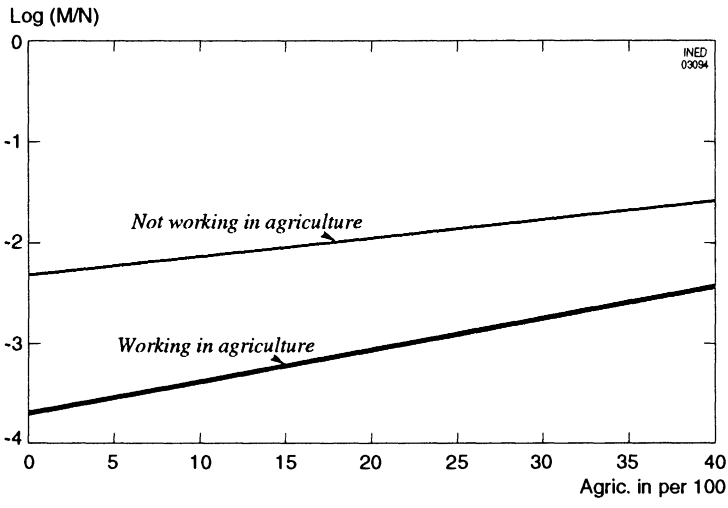

Figure 4.-- Logarithm showing the probability of migration for women working in agriculture and for other categories according to the percentage of people working in agriculture in each area

Conclusions

The combination of theoretical results and concrete results obtained using the '3B' survey, has enabled us to identify a certain number of interrelationships between individual and aggregate levels.

First of all, it was possible to relate the results obtained at the two levels when the characteristics were of the aggregate type. This theoretical result was confirmed by the data from the '3B' survey: the correlation between the parameters estimated for both levels is 0.999; the regression slope is, however, lower than unity. This difference is certainly due to the fact that we are working with a long time period, as only short durations make it possible to verify whether parameters are equal. The question of how these characteristics influence individual probabilities of migrating can also be asked. We believe that this influence is indirect, and fuller surveys are necessary in order to show, for instance, how an individual perceives a high proportion of farmers in his/her region of residence.

When individual characteristics are introduced into the 'micro' model, their effect can be very different from that of the characteristics measured at aggregate level. We have theoretically demonstrated that no simple relation is apparent between these parameters. Moreover, this result has been empirically confirmed, as the correlation between the parameters estimated with region-related characteristics and those estimated with individual characteristics becomes non-significantly different from zero. What is more, when region-related and individual characteristics are introduced simultaneously, the various parameters remain unchanged, whereas the quality of the model greatly improves. We have thus been able to verify that the correlations between region-related and individual characteristics are very low.

It thus seems that both types of characteristics have practically independent effects on individual behaviour patterns. This is contrary to many models where results are underpinned by supposed individual behaviour patterns which, when grouped together, produce aggregate behaviour patterns and characteristics [Weidlich and Haag, 1988, pp. 11-20; Puig, 1981, pp. 49-50]. As has been shown in the present paper, the areas where there is a high proportion of women in agriculture have much greater out-migration than the other areas, even though these women have themselves a much lower probability of migrating. The effect of the aggregate characteristic is, in this case, contrary to the sum of individual behaviour patterns.

In the light of these first encouraging results, it would seem necessary to pursue research in this field. In the following paragraphs, we give an indication of the exploratory paths which hold an interest for future work.

Firstly, the '3B' survey covers too small a sample to enable satisfactory verification of the theoretical results. The use of data from population registers providing information on movements of the population should have allowed more effective verification. These registers, however, only exist in a few countries (France has none) and only take into account a small number of individual information items, which thus sets a limit on their usefulness.

Secondly, in the present case, we were obliged to use a great many restrictive assumptions for the models employed, so as to facilitate comparison. Our theoretical results need to be extended to a greater number of models. It is particularly important to introduce destination area characteristics, along with ones which measure the interaction between the area of departure and the area of arrival. Although these characteristics are relatively easy to define and measure at 'macro' level, it is much more difficult to gather the information which an individual may have on the different possible destinations, at 'micro' level.

As can be seen, the links between individual and aggregate behaviour patterns are complex ones, and a great deal of research is still needed in order to totally elucidate how these two levels interrelate. The results obtained here show that it is nevertheless possible to make headway in a domain which as yet is still largely unexplored.

Daniel COURGEAU

REFERENCES

Courgeau D., 1975.– «L'intensité des changements de catégorie de communes», *Population*, 1, 81-102.

Courgeau D., 1985a.– «Interaction between spatial mobility, family and career life-cycle: a French survey», *European Sociological Review*, 1, 2, 139-62.

Courgeau D., 1985b.– «Changements de logement, changements de département et cycle de vie», *L'Espace géographique*, 4, 289-306.

Courgeau D., Lelièvre E., 1986.– «Nuptialité et agriculture», *Population*, 2, 303-26.

Courgeau D., Lelièvre E., 1992.– *Event History Analysis in Demography*, Clarendon Press, Oxford, 226 pp.

Cox D., 1972.– «Regression models and life tables», *Journal of the Royal Statistical Society*, B, 34, 187-220.

Greenwood M., 1975.–«Research on internal migration in the United States: a survey», *Journal of Economic Literature*, XII, 397-433.

Hamerle A., 1989.– «Multiple-spell regression models duration data», *Applied Statistics*, 38, 1, 127-138.

Kalbfleisch J., Prentice R., 1980.– *The statistical analysis of failure time data*, John Wiley and Sons, New York, 322 pp.

Lancaster T., 1990.– *The Econometric Analysis of Transition Data*, Cambridge University Press, Cambridge.

Puig J.-P., 1981.– «La migration régionale de la population active», *Annales de l'INSEE*, 44, 41-79.

Sandefur G., Scott W., 1981.– «A dynamic analysis of migration: an assessment of the effect of age, family and career variables», *Demography*, 18, 3, 355-368.

Sanders L., 1992.– *Système de villes et synergétique*, Anthropos, Paris, 274 pp.

Stillwell J., 1975.– «Models of interregional migration: a review», Working Paper, Dept. of Geography, Leeds, 60 pp.

Stouffer S., 1940.– «Intervening opportunities: a theory relating mobility and distance», *American Sociological Review*, 5, 845-67.

Tuma N.B., Hannan M.T., 1984.– *Social dynamics. Models and methods*, Academic Press, Orlando, 578 pp.

Weidlich W., Haag G. (eds.), 1989.– *Interregional migration*, Spinger-Verlag, Berlin, 388 pp.

Young E.C., 1924.– «The movement of farm population», *Cornell Agriculture Experiment Station Bulletin*, 426.