METHODS OF LINKING MIGRATION STATISTICS COLLECTED FROM NATIONAL SURVEYS WITH THOSE FROM POPULATION CENSUS

Daniel COURGEAU
INSTITUT NATIONAL d'ETUDES DEMOGRAPHIQUES
27, rue du Commandeur
75675 PARIS CEDEX 14
FRANCE

Prepared for the ESCAP Technical Working Group on Migration and Urbanization, 1 - 4 December 1981 in Bangkok.

December 1981

Final Report

nf

METHODS OF LINKING MIGRATION STATISTICS COLLECTED FROM NATIONAL SURVEYS WITH THOSE FROM POPULATION CENSUS

1. THE PROBLEM

The data collected through national migration surveys among Asian and Pacific countries complement and do not substitute for the data collected from national population censuses. Ideal utilization of the two sources of data is feasible if data from one source can be related to the others that are available. The purpose of this paper is to discuss different approaches for linking these varied sets of data.

We are mainly interested here in two particular sets : retrospective statistics collected from population censuses and those from national surveys. So that we will let aside statistics collected from population registers and those from repeat surveys for which it will be necessary to add some hypotheses about the links between migration and death. We will also let aside some survey data about commuting, circulation ...

The census data on internal migration relate to a little number of questions about previous residences. Cost considerations often make it desirable to keep this number to a minimum. On the other side answers to such questions will be available for the whole population, especially for small areas populations.

These questions are mainly of two kinds. The first one relates to place of residence on a specific date before the census. The length of the interval may vary according to the country considered and even for one country according to the census considered. The time intervals most commonly selected are five years and one year. The major part of the countries have also a question on place of birth, that will give a time interval variable from an individual to another. In any case such a question gives a count of migrants, who were alive at the beginning of the period and survived in the country to the end of it. When the question related to place of birth we can speak of lifetime migrants. Such a census question does not count as migrants those who move away from an area during the interval and returned to it before the end of the inverval, or died before this end. In another side persons may be counted as migrants between two areas, however they had never done a direct migration from one area to the other. For such repeat migrations only the residence at two definite points in time will be given by such a census question.

./.

Another kind of census questions may relate to the latest migration. One of these questions will be about the duration of residence in the place of enumeration, and the other about the place of last previous residence. Some countries use only one of these questions alone as others combine the two questions. We suppose here that we have a combination of them, so that latest migration streams can be studied for each period. It is important here to know the definition of the places of enumeration between which a latest migration is counted. Contrary to migrants, it is no more possible to have a count of latest migration to a consolidation of these places of enumeration (Courgeau, 1980 : 123). Such questions will give for any duration of residence in a place of enumeration the number of latest migrations classified by place of previous residence. Often to have enough numbers it is necessary to consider a longer duration of residence, 0-4 years for example. Such a number will be similar to the numbers of migrants over a five years period : it may be presented in the same matrix form. But the numbers in the two matrices will be different as we shall see further.

To get more complete information on multiple migrations and return migration it is necessary to use survey data because such information will be too expensive and too difficult to obtain through a census questionnaire.

The national migration surveys undertaken in the ESCAP region will give " information on all the movements that a person has made since the age of 15 that involved a stay at destination for 12 months or more for any reason, and any other moves involving shorter stays that were specifically for work, to look for work or to study " (ESCAP, 1980, II). So that all changes in residence or migrations of the surveyed sample will be recorded since the age of 15. Such surveys will also give the timing of each migration, and other informations on the demographic and socio-economic characteristics of the sample.

On the other side such surveys cannot give a detailed spatial view of the migrations undertaken by the whole population. For example they may be inadequate for estimating small areas migration streams. But it is possible to use the relationships among variables derived from the survey data to ameliorate the census data on migrants or latest migrations. These relationships when applied to the census data would allow more reliable estimates to be made of migration parameters for small areas. This imply to reconcile demographic data collected over different periods of time and with different sizes of observed samples.

To enlighten the different methods proposed in this paper we will simultaneously treat a numerical example, following the different parts of the paper. We suppose here to have a census question on place of residence two years ago, a census question about last place of residence and survey data about each migration done during previous time, with a sample of $1/10^{th}$. We will follow the course of the paper, so that the reader can easily refer to the theoretical approach.

2. THEORETICAL RELATIONS BETWEEN NUMBERS OF MIGRATIONS, MIGRANTS OR LATEST MIGRATIONS.

It will be useful to see in more detail how these statistics may be related. Let us first present the notations we will further use.

2.1 Notation

We are working on a period of t years (o, t), with a parcelling of the country into r areas.

$M_{ij}(t)$ = number of migrations from area i to area j during the (o,t) period

$\mathcal{M}_{ij}(t)$ = number of latest migrations from area i to area j during the (o,t) period

$\mathfrak{M}_{ij}(t)$ = number of migrants from area i to area j during the (o,t) period

When area i and j are the same, we define $M_{ii}(t)$ (or $\mathcal{M}_{ii}(t)$, or $\mathfrak{M}_{ii}(t)$) as :

$$M_{ii}(t) = - \sum_{j \neq i} M_{ij}(t) = - M_{i.}(t)$$

a dot indicating a sommation over all other areas.

All these numbers may be represented in a more synthetic matrix form : $M(t)$, $\mathcal{M}(t)$ and $\mathfrak{M}(t)$

Now, without considering the times at which moves take place, we represent here the spatial trajectory of an individual by the different successive areas of residence during the studied period. So that, for example :

$(i, j, i, k)_t$ = number of individuals who lived in area i at the beginning of the period, then undertake a migration to area j, after come back again to area i to undertake a last migration to area k, where they live at the end of the period.

These individuals have done three migrations during the (o,t) period. They will be counted as migrants from area i to area k and also for doing a latest migration from area i to area k, but for different reasons : in the first case because their first place of residence was in area i and their last one in area k, in the second case as their latest migration was from i to k.

The different kind of measurement may then be represented on keeping only the used information :

$$M_{ij}(t) = (\ldots\ i, j\ \ldots)_t$$

$$\mathcal{M}_{ij}(t) = (\ldots\ldots\ i, j)_t$$

$$\mathfrak{M}_{ij}(t) = (i\ \ldots\ldots\ j)_t$$

When completing the dotted lines by different possible places of residence we will know exactly what kind of migration trajectory distinguish one kind of measurement from another.

2.2 One area out - or in-flows

Let this area be i , the remaining areas of the country, that are not to be differentiated, being a, b, ... We will observe here only two migrations during the period, taking into account that its duration is short. However the results may easily be extended to a greater number of migrations.

We will put in correspondence in the same column the identical terms, so that it will be easy to see the kinds of migrations responsible for the difference. Let us first consider out-flows :

$$M_{i.}(t) = (i,a)_t + (i,a,b)_t + (a,i,b)_t + (i,a,i)_t$$

$$\mathcal{M}_{i.}(t) = (i,a)_t \qquad\qquad\qquad + (a,i,b)_t$$

$$\mathfrak{M}_{i.}(t) = (i,a)_t + (i,a,b)_t$$

When there is only one move the three out-flows will be identical. The differences are created by higher order moves. If it is so the number of out-migrations will be greater than the number of latest out-migration or of out-migrants, without any ordinal relation between the two last numbers. An individual making a <u>return migration</u> to area i will not be counted as an out-migrants or as a latest out-migration $(i, a, i)_t$. Individuals making <u>multiple migrations</u> will not be counted as out-migrants if the area i is the second area of residence $(a, i, b)_t$ and will not be counted as latest out migrations if it is the area of departure.

./.

To develop the numerical example we suppose that the country has been decomposed into three parts :

(1) a rural one

(2) an urban one of medium-size cities

(3) an urban one with major-size cities

Let us first decompose rural out-flows, into their different components, that are given by survey data. Using the same presentation than in the paper, we have :

$$M_{1.}(2) = 10{,}190 + 500 + 610 + 700 = 12{,}000$$
$$\mathcal{M}_{1.}(2) = 10{,}190 + 610 = 10{,}800$$
$$\mathfrak{m}_{1.}(2) = 10{,}190 + 500 = 10{,}690$$

From the 12,000 individuals having done an out-migration from rural areas, 700 return ones will not be counted as latest out-migrants or as out-migrants. From the 1,110 individuals having done multiple migrations, only 500 will be registered by a census question about place of residence two years ago. A question about last place of residence will register 610 moves, done after a previous one.

Let us now consider in-flows :

$$M_{.i}(t) = (a,i)_t + (a,b,i)_t + (i,a,i)_t + (a,i,b)_t$$
$$\mathcal{M}_{.i}(t) = (a,i)_t + (a,b,i)_t + (i,a,i)_t$$
$$\mathfrak{m}_{.i}(t) = (a,i)_t + (a,b,i)_t$$

Again the differences are created by moves of order higher than one, but there exists a perfect ordinal relation between the three numbers :

$$\mathfrak{m}_{.i}(t) \leqslant \mathcal{M}_{.i}(t) \leqslant M_{.i}(t)$$

In this case an individual making a return migration to area i will not be counted as an in-migrant but will be as a latest in-migrant. Only individuals making multiple migrations, when the area i is the second area of residence, will not be counted as well as in-migrants as latest in migrants $(a, i, b)_t$.

The numerical example will give rural in-flows :

$$M_{.1}(2) = 3{,}310 + 380 + 700 + 610 = 5{,}000$$
$$\mathcal{M}_{.1}(2) = 3{,}310 + 380 + 700 = 4{,}390$$
$$\mathfrak{m}_{.1}(2) = 3{,}310 + 380 = 3{,}690$$

From 5,000 individuals having done an in-migration to rural areas, 610 are not counted as latest in-migrants or as in-migrants, because, their in-migration has been followed by a new migration. In addition 700 other individuals cannot be counted as in-migrants, because their first migration had been undertaken from a rural area so that they are return migrants.

Let us now consider the net flow of area i :

$$M_{.i}(t) - M_{i.}(t) = (a,i)_t - (i,a)_t + (a,b,i)_t - (i,a,b)_t$$

$$\mathcal{M}_{.i}(t) - \mathcal{M}_{i.}(t) = (a,i)_t - (i,a)_t + (a,b,i)_t \qquad + (i,a,i)_t - (a,i,b)_t$$

$$\mathfrak{M}_{.i}(t) - \mathfrak{M}_{i.}(t) = (a,i)_t - (i,a)_t + (a,b,i)_t - (i,a,b)_t$$

In this case we can verify that net migration and net number of migrants are identical :

$$M_{.i}(t) - M_{i.}(t) = \mathfrak{M}_{.i}(t) - \mathfrak{M}_{i.}(t)$$

It is easy to see that this relation remains always true whatever the number of migrations considered. First an out-migrant from area i may have done more than one out-migration from this area, but each of these will be followed by an in-migration to i , so that only one net out-migration will be counted. The same is true for an in-migrant. Finally a non-migrant may undertake some out-migrations from area i , but each of these out-migrations will be followed by one in-migration so that the net migration will also be null.

On the other hand the net number of latest migrations is different from net migration.

During this first approach we have pointed out the kinds of migrations that give different estimations : return migrations, multiple migrations defined by the precise trajectory followed. It is an estimation of these flows we need from retrospective surveys. They will permit to link the different statistics.

2.3 Moves between two areas

Let these areas be i and j the remaining ones of the country being a, b,... Again we will observe only two migrations during the period but the results may easily be extended to a larger number of moves.

Calculating the moves between i and j , we have :

$$M_{ij}(t) = (i,j)_t + (a,i,j)_t + (j,i,j)_t + (i,j,a)_t + (i,j,i)_t$$

$$\mathcal{M}_{ij}(t) = (i,j)_t + (a,i,j)_t + (j,i,j)_t$$

$$\mathfrak{M}_{ij}(t) = (i,j)_t \qquad + (i,a,j)_t$$

These three streams are different. Only one ordinal relation may be found :

$$\mathcal{M}_{ij}(t) \leq M_{ij}(t)$$

More important is the fact that the number of migrants may be higher than the number of migrations. This is true when the following relation is verified :

$$(a,i,j)_t + (j,i,j)_t + (i,j,a)_t + (i,j,i)_t < (i,a,j)_t$$

In the extreme case, no direct migration from i to j will exist but only step migration. The following relations are then verified :

$$M_{ij}(t) = \mathcal{M}_{ij}(t) = 0 \qquad \mathfrak{M}_{ij}(t) > 0$$

The only number that will be positive is the number of migrants, the others being null.

Let us consider for the numerical example the flows between rural areas and medium size towns. Such flows may be decomposed as :

$$\begin{array}{lclcl}
M_{12}(2) & = & 3180 + 70 + 260 + 150 + 340 & & = 4000 \\
\mathcal{M}_{12}(2) & = & 3180 + 70 + 260 & & = 3510 \\
\mathfrak{M}_{12}(2) & = & 3180 & + 350 & = 3530
\end{array}$$

The 490 individuals making a migration from rural areas to medium size towns will not be registered as latest migrants or as migrants, because this migration has been followed by a new one. In addition, the 330 latest migrants from rural to medium size towns, will not be registered by a census question about place of residence two years ago, because this migration follows a previous one from major size urban areas. But 350 migrants from rural to medium size urban areas, that have an intermediate stay, will not be counted as doing migration or as latest migrants between these two areas.

Finally the net streams between i and j are :

$$M_{ji}(t) - M_{ij}(t) = (j,i) - (i,j) + (a,j,i) - (a,i,j) + (j,i,a) - (i,j,a)$$

$$\mathcal{M}_{ji}(t) - \mathcal{M}_{ij}(t) = (j,i) - (i,j) + (a,j,i) - (a,i,j) \qquad + (i,j,i)_t - (j,i,j)_t$$

$$\mathfrak{M}_{ji}(t) - \mathfrak{M}_{ij}(t) = (j,i) - (i,j) \qquad + (j,a,i)_t - (i,a,j)_t$$

For the net streams the previous identity of net migration with the net number of migrants is no more verified.

Again, if we want to link the different statistics on migration streams, we will have to estimate higher order moves. Such an estimation will be less easy than to estimate in- or out-flows, as more complex streams interfere.

./.

3. THE NEED OF LONGITUDINAL ANALYSIS.

The previous section gave us the kinds of streams that need to be analysed to have valuable estimates of a migration flow when we have statistics on migrant or latest migration flows.

Such an analysis needs to be longitudinal to show more clearly the hypotheses made when developping a model.

As other papers of this conference will treat in more details such an analysis, we will give here the only results that are useful for our purpose. We will consider three aspects of this analysis.

The effect of time on migration behaviour is the most important one. It may appear at different levels. The first one is the age of the individual. Once this age pattern taken into consideration it is useful to analyse the effect of duration-of-stay, as its importance has been emphasized in several studies of mobility. At the third level, analysis of the effect of previous moves should be considered.

The effect of space on migration parameters have then to be taken into account. The previous places of residence may play a part in the migration process and particularly return moves have to be considered.

Finally disagregation of a population into more homogeneous sub-populations will be studied. Different socio-economic groups display very different migration parameters. So that it will be useful to disagregate the population of a small area into such groups to get a more valuable estimate of its migration parameters.

3.1 The effect of time

Let us first remember that this analysis ought to be applicable to the census parcelling into a number, r , of small areas. So that the moves to be taken into account in the survey will be for the same parcelling. But as the sample size is around 14.000 interviews, it is not possible to consider each flow M_{ij} between each couple of areas. So that we may observe here the whole set of M_{ij} moves undertaken by a cohort :

$$M = \sum_{i \neq j} \sum_{j \neq i} M_{ij}$$

But such a set may be disagregated according to the time of the move, its rank ... whatever the areas of origin and destination may be.

Let us develop here some methods to estimate parameters that will be useful in further models.

Let ${}_aS_x^{n-1}$ the number of individuals of age category " a " at the $(n-1)^{th}$ move, who x years after this move have not made a new migration. Between the duration x and (x + 1), ${}_aM_x^n$ of these individuals make a n^{th} move. Rather than to calculate the traditional rate for this n^{th} move, it will be more useful to summarize these rates with a little number of indices. Even if these indices are rough estimates their use during a short period will give good results.

The first of these indices will be the limiting value of the percentage of individuals who make at least an n^{th} move, knowing that they had made the $(n-1)^{th}$ one at age a :

$$ {}_aK^n = \lim_x {}_xK_a^n $$

An interesting case is when such an index is independant of the rank of the move. Then it may be used whatever the past migration history of the individuals was. The migration process may become Markovian if this value is one.

Given a new move, we can then calculate the annual probability of migration :

$$ {}_ak_x^n = \frac{{}_aM_x^n}{{}_aS_x^{n-1} - (1 - {}_aK^n)\,{}_aS_0^{n-1}} $$

In many cases this probability remains fairly constant so that the migration process may be summarized by only two indices ${}_aK^n$ and ${}_ak^n$. Under such an hypothesis it is possible to estimate these indices on writing the model :

$$ {}_aM_x^n = {}_ak^n \left[{}_aS_x^{n-1} - (1 - {}_aK^n)\,{}_aS_0^{n-1} \right] + \sqrt{v}\, z_x $$

If z_x is a normal distributed variable with a mean of zero and a variance of one, we can estimate the coefficients ${}_ak^n$, ${}_aK^n$ and v from the observed ${}_aM_0^n$, ${}_aM_1^n$, ... ${}_aM_x^n$ variables. The least square method will give the same estimate as the maximum likelyhood method. Omitting, to simplify notations, the indication of age at which the previous move had been made and the rank of that move we have (Courgeau, 1979 b : 29) :

$$\hat{k} = \frac{x\sum_x M_x S_x - \left[\sum_x S_x\right]\left[\sum_x M_x\right]}{x\sum_x S_x^2 - \left[\sum_x S_x\right]^2}$$

$$\hat{K} = 1 - \frac{\left[\sum_x S_x^2\right]\left[\sum_x M_x\right] - S_o\left[\sum_x S_x\right]\left[\sum_x M_x\right]}{x S_o \sum_x M_x S_x - S_o\left[\sum_x S_x\right]\left[\sum_x M_x\right]}$$

If these coefficients depend on the age of previous migration, but are independant of the rank of the move, they may be used to link migration statistics collected from national surveys with those for population censuses. This rank independance is important to be stressed here because population censuses generally give no information on migration rank.

3.2 The effect of space

The previous indices have been calculated from the whole set of moves through a census parcelling into a great number of small areas. When applied to one of these areas such indices will give a good estimation of the real ones if each area has a size, population ... very near to the size, population of any other area or if these indices depend only slightly on the size, population of the observed area. Such a size dependance may be studied from survey data when considering different parcelling of the territory : change of residence, of county, of state ... These studies will ensure the use of the same indices for areas of varying size or population.

Another problem relates to return migration. It will be useful to know if the probability of returning back to a place of origin i from an intermediate place j will be the same as the probability of making a first move from j to i. Again such probabilities may be estimated from migration surveys. An interesting point to be studied is to see if these return migrations represent a constant part of migrations of rank higher than one during each year (Courgeau, 1979 a : 24). If this hypothesis is true, the formulation of the links between migrants and migrations will be easier.

3.3 Disaggregation of the population

The composition of the population of a small area may differ for an important part from the national one. So that when applying the national migration indices to this area we shall have very bad estimates of its migration parameters. For example, an entirely rural area may behave quite differently from the national one.

./.

The most obvious method of dealing with these differences is to disagregate the population into more homogeneous subgroups. It is then possible to estimate parameters separately for each subgroup on which there are enough data.

Such subgroups may be defined from different ways. Mobility may vary with occupation - income class, household type, race ... But it will be difficult to take into account a too important number of subgroups because we are always working on a survey sample. The occupation - income classes seems the more important to consider here.

As for the age variable we will try to have estimates of K and k for each of these group and also an estimate of return moves.

For the numerical example, such an analysis can be undertaken at different levels. For a short period of observation, here for two years, we can first suppose that the major part of the population undertake only one migration, or if they undertake more migrations that each move is independant of every other move and of the duration of stay.

However analysing the whole set of migrations between the three previous areas, we get some estimates of the following parameters :

$\hat{K} = 0.6 \qquad \hat{k} = 0.2 \qquad \hat{l} = 0.6$

This indicates that for 100 moves undertaken from rural to medium-size towns, for example, 60 new moves will be registered in the future, with an annual probability of moving of 0.2. For these 60 new moves, 36 will be return ones and 24 will go to a new destination.

A more detailed analysis of these moves will give us different sets of parameters, when we decompose these moves into migrations to larger urban areas and other ones. When a move to a major-town is undertaken the individual will have a greater probability to remain into this area. So that the parameters for a new migration for these areas are :

$\hat{K}_3 = 0.5 \qquad \hat{k}_3 = 0.2 \qquad \hat{l}_3 = 0.6$

In the other case, between rural areas and medium-size towns, the moves are more generally followed by a new one, so that the parameters for these new migrations are :

$\hat{K}_1 = 0.7 \qquad \hat{k}_1 = 0.2 \qquad \hat{l}_1 = 0.7$

We can see that for these moves the return ones are more frequently done by the population.

4. EVALUATION AND DISCUSSION OF DIFFERENT MATRIX METHODS.

The methods developed here introduce a discrete time. But their results can be generalized on using a continuous time model (Singer and Spilerman, 1976, Ginsberg, 1971 : 251, Keyfitz, 1980).

Let us first introduce the notations.

4.1 Notation

We will use here matrices of probabilities. The first one may be easily derived from the M (1) (or $\mathcal{M}(1)$, or $\mathfrak{m}(1)$) matrix : it gives the probabilities that a person in region i at the beginning of the period, will be living in a region j one year after. Such a matrix can be written :

$$P(1) = \begin{bmatrix} -\sum_{j\neq 1} p_{1j} & p_{21} & \cdots & p_{r1} \\ p_{12} & -\sum_{j\neq 2} p_{2j} & \cdots & p_{r2} \\ \vdots & \vdots & & \vdots \\ p_{1r} & p_{2r} & \cdots & -\sum_{j\neq r} p_{rj} \end{bmatrix}$$

where p_{ij} is the previous migration probability.

The initial populations of each area may be represented by a diagonal matrix :

$$N(0) = \begin{bmatrix} N_1(0) & 0 & \cdots & 0 \\ 0 & N_2(0) & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & N_r(0) \end{bmatrix}$$

where $N_i(0)$ is the number of observed persons living in area i at time 0. Such populations may be only a part of the total area i population: it may be an age group , a social group ...

We can then write the matrix of migration flows at time 1 :

$$M(1) = P(1)\, N(0)$$

With these notations the problem we have to solve will be : how to calculate $P(1)$ (or $M(1)$) from a census information giving $\mathcal{M}(t)$ or $\mathfrak{M}(t)$?
To solve this problem we may use other informations taken from the survey data.

We suppose in our numerical example that each area has an initial population of 100.000.
The observed matrix of migrants given by census data is :

$$\mathfrak{M}(2) = \begin{bmatrix} -3690 & 3527 & 7162 \\ 2440 & -5617 & 5178 \\ 1250 & 2090 & -12340 \end{bmatrix}$$

The observed matrix of latest migration given by census data :

$$\mathcal{M}(2) = \begin{bmatrix} -4385 & 3507 & 7297 \\ 2631 & -6138 & 5474 \\ 1754 & 2631 & -12771 \end{bmatrix}$$

The survey gives us an estimate of the annual matrix, we want to calculate from census data :

$$M(1) = \begin{bmatrix} -2150 & 1870 & 3780 \\ 1350 & -3130 & 2780 \\ 800 & 1260 & -6560 \end{bmatrix} \text{ or } P(1) \begin{bmatrix} -0.0215 & 0.0187 & 0.0378 \\ 0.0135 & -0.0313 & 0.0278 \\ 0.0080 & 0.0126 & -0.0656 \end{bmatrix}$$

4.2 Migration uniformly distributed

Let us first suppose that there are no multiple moves during the period of observation and that the migration flows are uniformly distributed during the t years period. In that case we may have $K^n = 0$ for each $n \geqslant 2$ so that the annual flow $p_{ij}\, N_i(0)$ may be replaced by a flow t times larger $t\, p_{ij}\, N_i(0)$. This will lead to the following expression :

$$P(t) = t\, P(1)$$

So that we may have

$$M(t) = \mathcal{M}(t) = \mathfrak{M}(t) = t\, P(1)\, N(0)$$

and the solution of the problem will be

$$P(1) = \frac{1}{t}\, \mathcal{M}(t)\, N(0)^{-1} = \frac{1}{t}\, \mathfrak{M}(t)\, N(0)^{-1} \qquad \textbf{(1)}$$

As the matrix $N(0)$ is diagonal $N(0)^{-1}$ will be very easy to calculate, each diagonal term being $\frac{1}{N_i(0)}$, the others remaining null.

In fact, such an hypothesis may be rarely verified as multiple migration is the usual case (Kitsul and Philipov, 1981 : 2). It will be approximatively true for a very short period but is quickly unusable as the duration increases.

In the numerical example, the $\underline{P}$ (1) estimate will be very easy to calculate on dividing by two each non diagonal term. So that this estimate will be :

$$\hat{P}_{ud}(1) = \begin{bmatrix} -0.01845 & 0.01763 & 0.03581 \\ 0.01220 & -0.02838 & 0.02589 \\ 0.00625 & 0.01045 & -0.06170 \end{bmatrix}$$

with a matrix of errors :

$$\hat{P}_{ud}(1) - \underline{P}(1) = \begin{bmatrix} +0.00305 & -0.00107 & -0.00199 \\ -0.00130 & +0.00292 & -0.00191 \\ -0.00175 & -0.00215 & +0.00390 \end{bmatrix}$$

Such a method under-estimates all the migration flows and gives an over-estimate for sedentary individuals. Even for a two years period, these error terms are not neglectible.

For the number of latest-migrants we have :

$$\hat{\underline{P}}'_{ud}(1) = \begin{bmatrix} -0.02193 & 0.01753 & 0.03649 \\ 0.01316 & -0.03069 & 0.02737 \\ 0.00877 & 0.01316 & -0.06386 \end{bmatrix}$$

with a matrix of errors :

$$\hat{\underline{P}}'_{ud}(1) - P(1) = \begin{bmatrix} +0.00043 & -0.00117 & -0.00131 \\ -0.00034 & +0.00061 & -0.00043 \\ -0.00077 & -0.00056 & +0.00174 \end{bmatrix}$$

Such a method also underestimate migration flows, but less than the previous one.

Another simple hypothesis is to suggest that an individual's migratory behaviour may be represented as a stochastic process. In the Markovian assumptions, each move is independant of every other move and the probability of a move is independant of the duration of residence in the origin area. It depends only on the two considered areas.

Such assumptions may be verified from the longitudinal analysis. In this case we may have $K^n = 1$ for each $n \geqslant 2$ and the annual probability of a move, h, will be independant of the duration of stay. Return migrations are also allowed with these assumptions : they are produced with the same probability as direct moves. For example for a two year period we have for different kind of migrants :

$$(i,j,i)_2 = p_{ij}\, p_{ji}\, N_i(0)$$

$$(i,j)_2 = \left[p_{ij} \left(1 - \sum_{a \neq j} p_{ja}\right) + \left(1 - \sum_{a \neq i} p_{ia}\right) p_{ij} \right] N_i(0)$$

$$(i,j,k)_2 = p_{ij}\, p_{jk}\, N_i(0)$$

Such an example shows how to calculate after a t years period the number of migrants of each kind and further how to estimate the numbers of migrations, latest migrations and migrants during this period. Keeping the two years example, we can write :

$$M_{ij}(2) = p_{ij} N_i(0) + \left(1 - \sum_{a \neq i} p_{ia}\right) p_{ij} N_i(0) + \left[\sum_{a \neq i} p_{ai} N_a(0) \right] p_{ij}$$

so that :

$$M_{ij}(2) = p_{ij} \left[\left(2 - \sum_{a \neq i} p_{ia}\right) N_i(0) + \sum_{a \neq i} p_{ai} N_a(0) \right]$$

The number of latest migrations will be :

$$\mathcal{M}_{ij}(2) = p_{ij}\left(1 - \sum_{a \neq j} p_{ja}\right) N_i(0) + \left(1 - \sum_{a \neq i} p_{ia}\right) p_{ij} N_i(0) + \left(\sum_{a \neq i} p_{ai} N_a(0) \right) p_{ij}$$

so that :

$$\mathcal{M}_{ij}(2) = p_{ij} \left[\left(2 - \sum_{a \neq j} p_{ja} - \sum_{a \neq i} p_{ia}\right) N_i(0) + \sum_{a \neq i} p_{ai} N_a(0) \right]$$

$$= M_{ij}(2) - p_{ij} N_i(0) \sum_{a \neq j} p_{ja}$$

Finally the number of migrants will be :

$$\mathfrak{M}_{ij}(2) = \left[p_{ij} \left(1 - \sum_{a \neq j} p_{ja}\right) + \left(1 - \sum_{a \neq i} p_{ia}\right) p_{ij} + \sum_{\substack{a \neq i \\ a \neq j}} p_{ia} p_{aj} \right] N_i(0)$$

In matrix form we can easily generalize this last equation :

$$N(0) + \mathcal{M}(t) = \left[I + P(1) \right]^t N(0)$$

So that the solution of the problem, when knowing the number of migrants over a t years period is

$$P(1) = \left(I + \mathcal{M}(t) N(0)^{-1} \right)^{\frac{1}{t}} - I \qquad (2)$$

For the number of latest migrations the solution is more complex and will need an algorithm to find the P (1) matrix.

Such a model that is better than the previous one, because it allows multiple migration and return migration to be considered, does not always give satisfactory results (Rees 1977, p. 262). It seems that generally a more complex process than the Markovian one is involved in multiple migration.

For the numerical example using the Markovian assumptions we can estimate from the number of migrants, the annual migration matrix. We have to calculate the square root of the following matrix :

$$I + \mathcal{M}(2) N(0)^{-1} = \begin{bmatrix} 0.96310 & 0.03527 & 0.07162 \\ 0.02440 & 0.94383 & 0.05178 \\ 0.01250 & 0.02090 & 0.87660 \end{bmatrix}$$

Such a root can be estimated by successive approximation :

$$\left(I + \mathcal{M}(2) N(0)^{-1} \right)^{\frac{1}{2}} = \begin{bmatrix} 0.98114 & 0.01786 & 0.03710 \\ 0.01241 & 0.97125 & 0.02691 \\ 0.00645 & 0.01089 & 0.93599 \end{bmatrix}$$

So that we will have an estimate

$$\hat{P}_m(1) = \begin{bmatrix} -0.01886 & 0.01786 & 0.03710 \\ 0.01241 & -0.02875 & 0.02691 \\ 0.00645 & 0.01089 & -0.06404 \end{bmatrix}$$

with a matrix of errors :

$$\hat{P}_m(1) - P(1) = \begin{bmatrix} +0.00264 & -0.00084 & -0.00070 \\ -0.00109 & +0.00255 & -0.00089 \\ -0.00155 & -0.00171 & +0.00159 \end{bmatrix}$$

Such a method always underestimate migration flows, but improves all the estimates.

4.4 Mover - stayer model and its extension

One way to improve the Markov model is to divide the population into different groups having a different migration matrix.

The simplest of these models is the "mover-stayer" one, where a certain part of the population, K, has a non-zero probability of migration (movers) while the rest of the population, (1 - K), has a zero probability of migration (stayers).

Such assumptions may be easily related to the previous longitudinal analysis. If we suppose that the whole population had just migrated before time 0, then the K coefficient will be exactly the same as previously defined (section 3.1). As such an hypothesis is not usually true the coefficient may be however related to the longitudinal index.

If these assumptions are verified we will have two independant Markovian processes whose mixture is not itself Markovian.

Let us again calculate the numbers of migrations, latest migrations and migrants we will have after a two years period. These numbers are only calculated on the $K\,N(0)$ part of the population but with a migration matrix that will be $\frac{1}{K}\,P(1)$. The number of migrations will be :

$$M_{ij}(2) = p_{ij}\,N_i(0) + \left(1 - \frac{1}{K}\sum_{a\neq i} p_{ia}\right) p_{ij}\,N_i(0) + \left[\sum_{a\neq i} \frac{1}{K}\,p_{ai}\,N_a(0)\right] p_{ij}$$

So that :

$$M_{ij}(2) = p_{ij}\left[\left(2 - \frac{1}{K}\sum_{a\neq i} p_{ia}\right) N_i(0) + \frac{1}{K}\sum_{a\neq i} p_{ai}\,N_a(0)\right]$$

The number of latest migrations will be :

$$\mathcal{M}_{ij}(2) = p_{ij}\left(1 - \frac{1}{K}\sum_{a\neq j} p_{ja}\right) N_i(0) + \left(1 - \frac{1}{K}\sum_{a\neq i} p_{ia}\right) p_{ij}\,N_i(0) + \left[\sum_{a\neq i} \frac{1}{K}\,p_{ai}\,N_a(0)\right] p_{ij}$$

So that :

$$\mathcal{M}_{ij}(2) = p_{ij}\left[\left(2 - \frac{1}{K}\sum_{a\neq j} p_{ja} - \frac{1}{K}\sum_{a\neq i} p_{ia}\right) N_i(0) + \frac{1}{K}\sum_{a\neq i} p_{ai}\,N_a(0)\right]$$

Finally the number of migrants will be :

$$\mathfrak{m}_{ij}(2) = \left[p_{ij}\left(1 - \frac{1}{K}\sum_{a\neq j} p_{ja}\right) + \left(1 - \frac{1}{K}\sum_{a\neq i} p_{ia}\right) p_{ij} + \sum_{\substack{a\neq i\\ a\neq j}} \frac{1}{K}\,p_{ia}\,p_{aj}\right] N_i(0)$$

$$= \left[p_{ij}\left(2 - \frac{1}{K}\sum_{a\neq j} p_{ja} - \frac{1}{K}\sum_{a\neq i} p_{ia}\right) + \frac{1}{K}\sum_{\substack{a\neq i\\ a\neq j}} p_{ia}\,p_{aj}\right] N_i(0)$$

Again the last relation can easily be generalized in matrix form :

$$N(0) + \mathcal{M}(t) = K\left[I + \frac{1}{K}P(1)\right]^{t} N(0) + (1-K)N(0)$$

So that the solution of the problem, when knowing the number of migrants over a t years period, is :

$$P(1) = K\left\{\left[I + \frac{1}{K}\mathcal{M}(t)N(0)^{-1}\right]^{\frac{1}{t}} - I\right\} \qquad \textbf{(3)}$$

Such a model will generally allow better estimates than the previous ones of the annual migration matrix. It is possible to generalize it on introducing two subpopulations, one with a high intensity of migration, the other with a low intensity of migration (Kitsul and Philipov, 1981). Such a model may be written :

$$N(0) + \mathcal{M}(t) = K\,R(1)^{t}\,N(0) + (1-K)\,S(1)^{t}\,N(0)$$

with the condition

$$K\,R(1) + (1-K)\,S(1) = I + P(1)$$

Again this mixture of two Markovian processes is generally not itself a Markovian process. But as to estimate the P (1) matrix we need two sets of data - for example one-year and five-year observations - and also the hypothesis that these matrices can be diagonalized by equal matrices (Kitsul and Philipov, 1981), we should not further develop this method, refering the reader to this paper. It may also be generalized to a greater number of subpopulations (Ginsberg, 1973 : 115-118).

For the numerical example, we begin to use information from survey data : only sixty percent of the individuals undertaking a previous migration will undertake a new one. We can then use a mover-stayer model.

We have first to calculate the square root of the following matrix :

$$I + \frac{5}{3}\mathcal{M}(t)N(0)^{-1} = \begin{bmatrix} 0.93850 & 0.05878 & 0.11937 \\ 0.04067 & 0.90639 & 0.08630 \\ 0.02083 & 0.03483 & 0.79433 \end{bmatrix}$$

that is :

$$\left(I + \frac{5}{3}\,\mathcal{M}(t)\,N(0)^{-1}\right)^{\frac{1}{2}} = \begin{bmatrix} 0.96808 & 0.03001 & 0.06348 \\ 0.02082 & 0.95126 & 0.04614 \\ 0.01100 & 0.01873 & 0.89038 \end{bmatrix}$$

that will give the following estimate :

$$\hat{P}_{ms}(1) = \begin{bmatrix} -0.01909 & 0.01801 & 0.03809 \\ 0.01249 & -0.02925 & 0.02768 \\ 0.00660 & 0.01124 & -0.06577 \end{bmatrix}$$

with a matrix of errors :

$$\hat{P}_{ms}(1) - P(1) = \begin{bmatrix} +0.00241 & -0.00069 & +0.00029 \\ -0.00101 & +0.00205 & -0.00012 \\ -0.00140 & -0.00136 & -0.00017 \end{bmatrix}$$

As we can see this method improves all the estimates, particularly those of migration to large towns, and we have no more an underestimation of all migration flows.

4.5 Migration as a semi-Markov process

Such a process is another generalization of the mover-stayer model. It introduces the existence of " cumulative inertia " that had been emphasized in many migration studies. With such an hypothesis the probability of going from an area i to another area j depends not only on i and j but also on how long the individual has been in i . As this duration of residence increases this probability will be less important. Such a process is known as a Semi-Markov or Markov-Renewal Process (Ginsberg, 1971).

We will present here such a process in a discrete time model, but again it may be easily converted into a continuous time model.

To develop this model we need the series of probabilities of going from i to j for each duration of stay. Such a series may be derived from retrospective surveys as the probabilities of making a migration through the same parcelling for each duration of stay. Again we have to suppose that the whole population had just migrated before time 0.

So that we can write :

$$p_{ij}(1) = p_{ij} \qquad p_{ij}(2) = p_{ij}\, f(2) \quad \ldots$$

Knowing this distribution (Ginsberg, 1979) it is possible to construct the whole Semi-Markov process. We will give here the estimations of the number of migrations, latest migrations and migrants that may be observed after a two-years period.

The number of migrations will be :

$$M_{ij}(2) = N_i(0)\left[p_{ij} + \left(1 - \sum_{a \neq i} p_{ia}\right) p_{ij}\, f(2) \right] + \left[\sum_{a \neq i} p_{ai}\, N_a(0) \right] p_{ij}$$

The number of latest migrations will be :

$$\mathcal{M}_{ij}(2) = N_i(0)\left[p_{ij}\left(1 - \sum_{a \neq j} p_{ja}\right) + \left(1 - \sum_{a \neq i} p_{ia}\right) p_{ij}\, f(2) \right] + \left[\sum_{a \neq i} p_{ai}\, N_a(0) \right] p_{ij}$$

The number of migrants will be :

$$\mathfrak{M}_{ij}(2) = N_i(0)\left[p_{ij}\left(1 - \sum_{a \neq j} p_{ja}\right) + \left(1 - \sum_{a \neq i} p_{ia}\right) p_{ij}\, f(2) + \sum_{\substack{a \neq i \\ a \neq j}} p_{ia}\, p_{aj} \right]$$

Knowing $\mathfrak{M}_{ij}(t)$ and the series $(1, f(2), f(3) \ldots f(t))$ it may be possible to estimate $\mathcal{P}(1)$

Such a process gives a more general view of the migration behaviour than the previous ones. However it will generally be impossible to tell the difference between the Semi-Markov model and the mover-stayer model, looking only at the distribution of intervals between events (Ginsberg, 1971 : 254). So that for our purpose the Semi-Markov process will not greatly improve the results, giving only a more realistic model. We will refer the reader to Ginsberg papers for further developments of such a model (Ginsberg, 1971, 1979 and others).

5. TOWARD A MORE GENERAL MODEL

In all the previous models the process is assumed to start from scratch at the instant a given area is entered. So that the times between entering and leaving a given state are assumed to be independent, identically distributed random variables. Such an hypothesis may be too restrictive. Several authors had stressed the importance of return move : the probability of returning back to a place of origin i from an intermediate place j will not be the same as the probability of making a first move from j to i. It is now necessary to take into account such moves, to give a more general model.

./.

5.1 A simple model of out-migration and in-migration.

We are dealing here with a small area, i , for which we suppose the migration flows to be even during the whole census period. So that during a very short interval of time, dx, the in-migration is $m_{.i}\,dx$ and the out-migration $m_{i.}\,dx$ for the sub-population on which we are working. Such a model introduces a continuous time, but it may easily be transformed into a discrete time model.

From the census we have a number of latest in-migration $\mathcal{U}_{.i}(t)$ (or a number of in-migrants $\mathcal{M}_{.i}(t)$) and a number of latest out-migration $\mathcal{U}_{i.}(t)$ (or a number of out-migrants $\mathcal{M}_{i.}(t)$), for the period (o, t).

Let us consider the $m_{.i}\,dx$ individuals undertaking an in-migration to i . Some of them will later undertake an out-migration from i , so that they will not be counted as latest in-migrants or in-migrants during the (o, t) period. Let us see how these new migrations may occur. First we will suppose that only a part K of these migrants will undertake a new migration, and that for this sub-group the annual probability of migration will remain constant, k . The new migrations occuring during a short interval of time (t, t + dt), dm (t), will then be proportional to the population at risk $(K m_{.i}\,dx - m(t))$, so that we can write :

$$dm(t) = k\left[K m_{.i}\,dx - m(t)\right]dt$$

The integration of this relation gives :

$$K m_{.i}\,dx - m(t) = C\,e^{-kt}$$

The conditions at the initial time $t = x$ gives

$$C = K m_{.i}\,e^{kx}\,dx$$

So that :

$$m(t) = K m_{.i}\,dx\left(1 - e^{-k(t-x)}\right)$$

where $m(t)$ are the new migrations undertaken by individuals having done a previous in-migration toward i.

When x varies from 0 to t we may enumerate all the new migrations out of area i , by calculating the integral :

$$\int_{x=0}^{x=t} K m_{.i}\, dx \left(1 - e^{-k(t-x)}\right) = K m_{.i}\left[t - \frac{1}{k}\left(1 - e^{-kt}\right)\right] = (a,i,b)_t$$

This will be the number of previous migrants to i that have done an out-migration from i afterwards $(a, i, b)_t$

So that the number of in-migrants remaining in the area i , that is the number of latest out-migrations to i , will be :

$$\mathcal{M}_{.i}(t) = m_{.i}\left[(1-K)t + \frac{K}{k}\left(1 - e^{-kt}\right)\right]$$

To estimate the number of in-migrants we have to go further. Some of the individuals undertaking a last in-migration to area i may not be counted as in-migrants if they are return migrants. We will suppose here that the number of return migrants is proportional to the population at risk, that is those who undertake a new migration from the area where they were. If this proportion is l we can estimate the number of returning migrants to area i as :

$$\int_{x=0}^{x=t} lK m_{i.}\, dx \left(1 - e^{-k(t-x)}\right) = lK m_{i.}\left[t - \frac{1}{k}\left(1 - e^{-kt}\right)\right] = (i,a,i)_t$$

We suppose here that these new migrations are well described with the same parameter K as the area i migrations. Such an assumption may be changed. Under this condition the number of in-migrants to area i will be :

$$\mathcal{M}_{.i}(t) = m_{.i}\left[(1-K)t + \frac{K}{k}\left(1 - e^{-kt}\right)\right] - lK m_{i.}\left[t - \frac{1}{k}\left(1 - e^{-kt}\right)\right]$$

Let us consider now the out-migrations from area i , $m_{i.}\,dx$. We have yet calculated the number of returning migrants to area i that will not be counted as out-migrants. But also the individuals having previously moved to area i and undertaking later an out-migration from this area will not be counted as out-migrants. So that the number of out-migrants to area i will be :

$$\mathcal{M}_{i.}(t) = m_{i.}\,t - lK m_{i.}\left[t - \frac{1}{k}\left(1 - e^{-kt}\right)\right] - K m_{.i}\left[t - \frac{1}{k}\left(1 - e^{-kt}\right)\right]$$

or :

$$\mathcal{M}_{i.}(t) = m_{i.}\left[(1 - lK)t + \frac{lK}{k}\left(1 - e^{-kt}\right)\right] - K m_{.i}\left[t - \frac{1}{k}\left(1 - e^{-kt}\right)\right]$$

We can easily verify that the net migration to area i is equal to its net number of migrants :

$$\mathfrak{M}_{.i}(t) - \mathfrak{M}_{i.}(t) = (m_{.i} - m_{i.})\, t$$

as calculated with these values.

To estimate the number of latest out-migration we have to go further. Tô get this number we may calculate the out-migrations from area i that are followed by a new migration to any other area, i , included. These will be :

$$K m_{i.} \left[t - \frac{1}{k} \left(1 - e^{-kt} \right) \right] = (i, a, b)_t$$

recalling the previous hypothesis about new migrations from outside areas. So that the number of latest out-migrations from area i will be :

$$\mathcal{M}_{i.}(t) = m_{i.} \left[(1-K)\, t + \frac{K}{k} \left(1 - e^{-kt} \right) \right]$$

Resolving the two systems of equations giving $\mathcal{M}_{i.}(t)$ and $\mathcal{M}_{.i}(t)$ (or $\mathfrak{M}_{i.}(t)$ and $\mathfrak{M}_{.i}(t)$ we can get estimates of the in-migration and the out-migration probabilities, from census data :

$$\hat{m}_{.i} = \frac{\mathcal{M}_{.i}(t)}{(1-K)\, t + \frac{K}{k} \left(1 - e^{-kt} \right)} \qquad (4)$$

$$\hat{m}_{i.} = \frac{\mathcal{M}_{i.}(t)}{(1-K)\, t + \frac{K}{k} \left(1 - e^{-kt} \right)} \qquad (5)$$

or :

$$\hat{m}_{.i} = \frac{\mathfrak{M}_{.i}(t) - \left(\mathfrak{M}_{.i}(t) - \mathfrak{M}_{i.}(t) \right) \ell K \left[1 - \frac{1}{kt} \left(1 - e^{-kt} \right) \right]}{\left[1 - K(1+\ell) \right] t + \frac{K(1+\ell)}{k} \left(1 - e^{-kt} \right)} \qquad (6)$$

$$\hat{m}_{i.} = \frac{\mathfrak{M}_{i.}(t) + \left(\mathfrak{M}_{.i}(t) - \mathfrak{M}_{i.}(t) \right) K \left[1 - \frac{1}{kt} \left(1 - e^{-kt} \right) \right]}{\left[1 - K(1+\ell) \right] t + \frac{K(1+\ell)}{k} \left(1 - e^{-kt} \right)} \qquad (7)$$

To have such estimates it is necessary to know the migration parameters from retrospective survey data. Such parameters may be estimated as indicated in section 3.1 and 3.2.

./.

5.2 Extension of this model to migration flows

We are now considering two areas i and j . During the observed period the migration flow remains even , $m_{ij}\,dx$

We have given in section 2.3 the decomposition of the flows for which we need an estimation. It is necessary to give additional assumptions.

Let us first suppose the (i, j ,a) + (i , j , i) flow to be proportional to the previous (i, a, b) flow. So that we can write :

$$(i,j,a)_t + (i,j,i)_t = m_{ij}\,K\left[t - \frac{1}{k}(1 - e^{-kt})\right]$$

This relation verify the condition, previously stated :

$$\sum_{j\neq i}\left[(i,j,a)_t + (i,j,i)_t\right] = \hat{m}_{i.}\,K\left[t - \frac{1}{k}(1 - e^{-kt})\right]$$

as : $\sum_{j\neq i} m_{ij} = \hat{m}_{i.}$

Under this condition we can estimate the number of latest migration from i to j as :

$$\mathcal{M}_{ij}(t) = m_{ij}\left[(1-K)t + \frac{K}{k}(1 - e^{-kt})\right]$$

So that

$$\hat{m}_{ij} = \frac{\mathcal{M}_{ij}(t)}{(1-K)t + \frac{K}{k}(1 - e^{-kt})} \qquad (8)$$

To estimate the number of migrants we need other estimates. First the (a, i, j) + (j, i, j) flow may be decomposed. As a (j, i, j) flow is a return one, we will have :

$$(j,i,j)_t = K\ell\, m_{ji}\left[t - \frac{1}{k}(1 - e^{-kt})\right]$$

For the other moves that are not return ones we will have :

$$(a,i,j)_t = K(1-\ell)\, m_{ij}\left(\sum_{\substack{a\neq i\\ a\neq j}} \frac{m_{ai}}{m_{i.} - m_{ia}}\right)\left[t - \frac{1}{k}(1 - e^{-kt})\right]$$

./.

We need further an estimate of the (i, a, j) flow. As such a flow is not a return one we have :

$$(i,a,j)_t = K(1-\ell)\left(\sum_{\substack{a\neq i\\ a\neq j}} \frac{m_{ia}\, m_{aj}}{m_{a.} - m_{ai}}\right)\left[t - \frac{1}{k}(1-e^{-kt})\right]$$

Under these conditions the number of migrants from area i to area j will be :

$$\mathcal{M}_{ij}(t) = m_{ij}\,t - K\left[m_{ij} + \ell m_{ji} + (1-\ell)\, m_{ij}\left(\sum_{\substack{a\neq i\\ a\neq j}} \frac{m_{ai}}{m_{i.} - m_{ia}}\right) - (1-\ell)\left(\sum_{\substack{a\neq i\\ a\neq j}} \frac{m_{ia}\, m_{aj}}{m_{a.} - m_{ai}}\right)\right]\left[t - \frac{1}{k}(1-e^{-kt})\right] \qquad (9)$$

Such a system of equations for each m_{ij} flow verifies the conditions :

$$\sum_j \mathcal{M}_{ij}(t) = \mathcal{M}_{i.}(t) = m_{i.}\,t - K(\ell m_{i.} + m_{.i})\left[t - \frac{1}{k}(1-e^{-kt})\right]$$

$$\sum_i \mathcal{M}_{ij}(t) = \mathcal{M}_{.j}(t) = m_{.j}\,t - K(\ell m_{j.} + m_{.j})\left[t - \frac{1}{k}(1-e^{-kt})\right]$$

As it is not linear it may be solved by an iterative procedure. We have previously calculated estimates of $m_{i.}$ and $m_{.j}$. From these estimates we can write a first series of m_{ij} :

$$\hat{m}^1_{ij} = \mathcal{M}_{ij}(t)\,\frac{m_{i.}}{\mathcal{M}_{i.}}$$

that will verify :

$$\sum_{j\neq i} \hat{m}^1_{ij} = m_{i.}$$

but no more :

$$\sum_{i\neq j} \hat{m}^1_{ij} = m_{.j}$$

Applying these estimates to equation (9) , we will have a value of $\hat{\mathcal{M}}^1_{ij}(t)$ that may be different from the observed one. Writing the differential equation from (9) we will have :

$$\begin{aligned}
d\left(\mathcal{M}_{ij}(t)\right) = t\, dm_{ij} - K\Bigg[& dm_{ij} + \ell\, dm_{ji} + (1-\ell)\, dm_{ij}\left(\sum_{\substack{a\neq i\\ a\neq j}} \frac{\hat{m}^1_{ai}}{\hat{m}_{i.} - \hat{m}^1_{ia}}\right) \\
& + (1-\ell)\,\hat{m}^1_{ij}\left(\sum_{\substack{a\neq i\\ a\neq j}} \frac{dm_{ai}}{\hat{m}_{i.} - \hat{m}^1_{ia}}\right) - (1-\ell)\,\hat{m}^1_{ij}\left(\sum_{\substack{a\neq i\\ a\neq j}} \frac{\hat{m}^1_{ai}\, dm_{ia}}{(\hat{m}_{i.} - \hat{m}^1_{ia})^2}\right) \\
& - (1-\ell)\sum_{\substack{a\neq i\\ a\neq j}} \frac{\hat{m}^1_{aj}\, dm_{ia}}{\hat{m}_{a.} - \hat{m}^1_{ai}} - (1-\ell)\sum_{\substack{a\neq i\\ a\neq j}} \frac{\hat{m}^1_{ia}\, dm_{aj}}{\hat{m}_{a} - \hat{m}^1_{ai}} \\
& + (1-\ell)\sum_{\substack{a\neq i\\ a\neq j}} \frac{\hat{m}^1_{ia}\, \hat{m}^1_{aj}\, dm_{ai}}{(\hat{m}_{a.} - \hat{m}^1_{ai})^2}\Bigg]\left[t - \frac{1}{k}(1-e^{-kt})\right]
\end{aligned}$$

So we have to solve a system of $r(r-1)$ linear equations, the variables being the dm_{ij} ones. The solutions of this system will give better estimates :

$$\hat{m}^2_{ij} = \hat{m}^1_{ij} + dm_{ij}$$

The same procedure may be applied until we have the needed approximation.

Another procedure will be to find a first set of m_{ij} estimates verifying the following conditions:

$$\hat{m}^1_{ij} = r_i s_j$$

$$\hat{m}^1_{i.} = m_{i.}$$

$$\hat{m}^1_{.j} = m_{.j}$$

Such a set may be estimated on using an algorithm (Courgeau, 1980, 153-154 ; Tugault, 1970, 65-66). We will summarize it here. The first step will be to write :

$$\hat{m}^{11}_{ij} = \frac{m_{i.}}{m_{..} - m_{j.}} \, m_{.j}$$

Such an estimate will verify

$$\hat{m}^{11}_{.j} = m_{.j} \quad \text{but} \quad \hat{m}^{11}_{i.} \neq m_{i.}$$

So that we will introduce a second step on writing :

$$\hat{m}^{12}_{ij} = \hat{m}^{11}_{ij} \frac{m_{i.}}{\hat{m}^{11}_{i.}}$$

Such an estimate will verify

$$\hat{m}^{12}_{i.} = m_{i.} \quad \textit{but} \quad \hat{m}^{12}_{.j} \neq m_{.j}$$

A third step may be to write :

$$\hat{m}^{13}_{ij} = \hat{m}^{12}_{ij} \frac{m_{.j}}{\hat{m}^{12}_{.j}}$$

and so on until we have the needed approximation. Then the previous procedure may be applied to find the solution of the problem : it will converge more quickly than in the previous case because the two marginal conditions are verified. Other ways to get such estimates may be found in Willekens, Por and Raquillet, 1981.

Such a model may be greatly improved if we have estimates of parameters K and ℓ for different kind of moves : rural to urban, intra-rural, intra-urban ... The solution of such a system with different K, k and ℓ parameters will not be more difficult than the case we took here (an unique set of K, k and ℓ parameters).

./.

For the numerical model, the instantaneous migration matrix to be estimated is :

$$M = \begin{bmatrix} -2500 & 2000 & 4000 \\ 1500 & -3500 & 3000 \\ 1000 & 1500 & -7000 \end{bmatrix} \quad \text{or} \quad P = \begin{bmatrix} -0.025 & 0.020 & 0.040 \\ 0.015 & -0.035 & 0.030 \\ 0.010 & 0.015 & -0.070 \end{bmatrix}$$

We will first suppose that the different migrants have an homogeneous behaviour, so that we can use the preceding parameters, estimated from migration surveys :

$$\hat{K} = 0.6 \qquad \hat{k} = 0.2 \qquad \hat{l} = 0.6$$

For the three areas the model can be written, as six linear equations :

$$3527 = 2\,m_{12} - (m_{12} + 0{,}6\,m_{21} + 0.4\,m_{31} - 0.4\,m_{13})\,0.211$$
$$7162 = 2\,m_{13} - (m_{13} + 0.6\,m_{31} + 0.4\,m_{21} - 0.4\,m_{12})\,0.211$$
$$2440 = 2\,m_{21} - (m_{21} + 0.6\,m_{12} + 0.4\,m_{32} - 0.4\,m_{23})\,0.211$$
$$5178 = 2\,m_{23} - (m_{23} + 0.6\,m_{32} + 0.4\,m_{12} - 0.4\,m_{21})\,0.211$$
$$1250 = 2\,m_{31} - (m_{31} + 0.6\,m_{13} + 0.4\,m_{23} - 0.4\,m_{32})\,0.211$$
$$2090 = 2\,m_{32} - (m_{32} + 0.6\,m_{23} + 0.4\,m_{13} - 0.4\,m_{31})\,0.211$$

Such a system can be simplified to give the following ones :

$$3231.6 = 1.7046\,m_{12} - 0.0422\,m_{21}$$
$$2355.6 = 1.7046\,m_{21} - 0.0422\,m_{12}$$
$$6866.6 = 1.7046\,m_{13} - 0.0422\,m_{31}$$
$$1629.8 = 1.7046\,m_{31} - 0.0422\,m_{13}$$
$$5093.6 = 1.7046\,m_{23} - 0.0422\,m_{32}$$
$$2469.8 = 1.7046\,m_{32} - 0.0422\,m_{23}$$

./.

that will give the solution matrix :

$$\hat{P}_{m_1} = \begin{bmatrix} -0.02486 & 0.01931 & 0.04056 \\ 0.01430 & -0.03455 & 0.03027 \\ 0.01056 & 0.01524 & -0.07083 \end{bmatrix}$$

with a matrix of errors :

$$\hat{P}_{m_1} - P = \begin{bmatrix} +0.00014 & -0.00069 & +0.00056 \\ -0.00070 & +0.00045 & +0.00027 \\ +0.00056 & +0.00024 & -0.00083 \end{bmatrix}$$

This method gives a new improvement of migration estimates between small towns and rural areas, but not for migration to major urban areas.

It is necessary to go further and to use the whole information from survey data. For all moves except those to major urban areas, the new migration can be summarized with the set of parameters :

$K_1 = 0{,}7 \qquad k_1 = 0{,}2 \qquad l_1 = 0{,}7$

The new migrations from major urban areas are summarized by an other set of parameters :

$K_3 = 0{,}5 \qquad k_3 = 0{,}2 \qquad l_3 = 0{,}5$

So that we can write the following system of linear equations :

$$\begin{aligned}
3527 &= 2\,m_{12} - 0.246\,(m_{12} + 0.7\,m_{21} + 0.3\,m_{31}) + 0.5 \times 0.17575\,m_{13} \\
7162 &= 2\,m_{13} - 0.17575\,m_{13} - 0.246\,(0.7\,m_{31} + 0.3\,m_{21} - 0.3\,m_{12}) \\
2440 &= 2\,m_{21} - 0.246\,(m_{21} + 0.7\,m_{12} + 0.3\,m_{32}) + 0.5 \times 0.17575\,m_{23} \\
5178 &= 2\,m_{23} - 0.17575\,m_{23} - 0.246\,(0.7\,m_{32} + 0.3\,m_{12} - 0.3\,m_{21}) \\
1250 &= 2\,m_{31} - 0.246\,(m_{31} - 0.3\,m_{32}) - 0.5 \times 0.17575\,(m_{13} + m_{23}) \\
2090 &= 2\,m_{32} - 0.246\,(m_{32} - 0.3\,m_{31}) - 0.5 \times 0.17575\,(m_{23} + m_{13})
\end{aligned}$$

The sollution of this system leads us to a matrix very near from P.

CONCLUSIONS

We have developed in this paper a general approach to analyse the effect of multiple and return moves. These moves give different mobility estimates when working on migrations, latest-migrations or migrants.

Such an approach is based on the existence of a partial estimation of these multiple and return moves from retrospective surveys. Although such surveys did not give valuable migration estimates for small areas, they will give more valuable estimates for some migration parameters. The idea is then to use a parcelling of the national territory into these small areas, but not to consider separately each flow between them. So that we will have an adequate number of migrations to estimate the different useful migration parameters : probability of a new migration, annual migration rates for different migration ranks, probability of a return migration ...

These parameters can be estimated for different age-groups, socio-economic groups ... Such a decomposition will be linked to the size of the sample and cannot be very detailed. However it will give more valuable estimates when splitting each area into these groups.

Then refering to Census data on latest migration or migrants for each small area, we will suppose that these flows may be analysed in the same way as survey data. First we will have an instantaneous migration flow even during the observed period, that will further establish new multiple or return flows with the same probability than estimated from the survey. If such an hypothesis is verified we have a good description of migration flows for each area with a set of parameters estimated from the survey and another set of latest migration or migrants measures from the census.

We have given in this paper different estimates of instantaneous (or annual) migration flows with different hypothesis on the used migration parameters. It is however necessary to undertake the longitudinal analysis of migration survey to be able to take the best suited model for each country and for each parcelling of this country.

The last model we presented allows the greatest variability of these parameters. It releases from the hypothesis made in other models that the process is starting from scratch at the instant a given area is entered. It allows memory effect that had been shown in many studies. Some of the hypothesis taken to develop it may be replaced by more general ones if the longitudinal analysis gives different informations on multiple or return moves.

./.

The problem handled here is an important one, because it gives the possibility to compare small areas migration parameters, cancelling out the effect of multiple or return migration. Such an effect is quite complex and needs a detailed analysis from retrospective surveys. Further improvements on this methodology may be related to a more thorough longitudinal analysis of multiple migration in relation with different characteristics of the individual.

ACKNOWLEDGMENTS

The author would like to thank Philippe Collomb, Sally Findley, Sidney Goldstein, Andrew Struik, Franz Willekens and other participants of the technical working group on Migration and Urbanization, for useful suggestions and criticisms.

REFERENCES

BLUMEN, I., F. KOGAN, and P. McCARTHY (1955) The Industrial Mobility of Labour as a Probability Process. Cornell Studies of Industry and Labour Relations, 4. Ithaca, N.Y.

COURGEAU, D. (1979 a) Migrants and Migrations (Translated from Population 1 - 1973). Population, Selected Papers, 3.

COURGEAU, D. (1979 b) Migration and Demographic Phenomena in France. In The Urban Impact of Internal Migration, edited by J. White, Institute for Research in Social Science, University of North Carolina at Chapel Hill : 1-32.

COURGEAU, D. (1980) Analyse Quantitative des Migrations Humaines. Paris, New York, Barcelone, Milan : Masson.

ESCAP (1980) National Migration Surveys. II. The Core Questionnaire. New York : United Nations.

GINSBERG, R. (1971) Semi-Markov Processes and Mobility. Journal of Mathematical Sociology 1 : 233-262.

GINSBERG, R. (1973) Stochastic models of residential and geographic mobility for heterogeneous populations. Environment and Planning, 5 : 113-124.

GINSBERG, R. (1979) Timing and Duration Effects in Residential Histories and Other Longitudinal Data, I : Stochastic and Statistical Models. Regional Science and Urban Economies 9.

GOODMAN, L. (1961) Statistical Methods for the Mover-Stayer Model. Journal of the American Statistical Association 56 : 841-868.

KEYFITZ, N. (1980) Multidimentionality in Population Analysis, Sociological Methodology, 1980 : 191-218.

KITSUL, P. and D. PHILIPOV (1981) The One-Year / Five-Year Migration Problem. In Advances in Multiregional Demography, edited by A. Rogers, IIASA, Laxenburg, Austria : 1-33

REES, P. (1977) The Measurement of Migration, from Census Data and Other Sources. Environment and Planning A 9 : 247-272.

ROGERS, A. (1975) Introduction to Multiregional Mathematical Demography. New York : Wiley.

ROGERS, A., and J. LEDENT (1976) Increment - Decrement Life Tables : a Comment. Demography 13 : 287-290

SINGER, B. and S. SPILERMAN (1976) The Representation of Social Processes by Markov Models. American Journal of Sociology 82 : 1-54

SPILERMAN, S. (1972) Extensions of the Mover-Stayer Model. American Journal of Sociology 78 : 599-626.

SPILERMAN, S. (1972) The Analysis of Mobility Processes by the Introduction of Independant Variables into a Markov Chain. American Sociological Review 37 : 277-294.

TUGAULT, Y. (1970) Méthode d'analyse d'un tableau "origine destination" de migrations. Population 25 (1) : 59-68.

WILLEKENS, F., A. POR and R. RAQUILLET (1981) Entropy, Multiproportional, and Quadratic Techniques for Inferring Patterns of Migration from Aggregate Data. In Advances in Multiregional Demography, edited by A. Rogers, III.A.S.A., Laxenburg, Austria : 83-124.