

VERY ROUGH DRAFT

Inferring period migration streams from two lifetime migration datasets

Jacques Ledent

Urbanisation, Culture et Société
Institut national de la recherche scientifique (INRS)
and
Centre interuniversitaire d'études démographiques (CIED)
Montreal, Quebec (Canada)

jacques.ledent@inrs-ucs.uquebec.ca

Paper prepared for presentation
at the Colorado Conference on the estimation of migration
Estes Park, Colorado, September 24-26, 2004

FOREWORD

This paper reports on an advance to migration estimation that has been in the making for almost a quarter of a century. Indeed, the basic idea behind it was first proposed in 1980 when the author, then affiliated with the Human Settlements and Services Area of the International Institute for Applied Systems Analysis (IIASA), Laxenburg (Austria), was in the midst of a research activity dealing with the influence of the birthplace on interregional migration patterns. But implementation of this idea, beyond the simple case of a two-region system in absence of mortality/emigration, resisted him at the time.

Such implementation became all but a forgotten issue until the late nineties when Andrei Rogers, who had just embarked on a new line of research dealing with the indirect estimation of migration, resurrected it and exhorted the author to take it up again. Some five years of efforts off and on, though, proved necessary to come up with a full satisfactory solution. First, the 1980 solution for a two-region system (in absence of mortality/emigration) was extended to the case of a N-region system (still in absence of mortality/emigration), in time for presentation at the Colorado Migration Conference held in Estes Park in March 1999. Second, in late 2001-early 2002, this extended solution was explicitly formulated before it was illustrated by means of an application that was prepared for inclusion in a research proposal about a multinational comparative study on the estimation of migration under the leadership of Andrei Rogers. Third and finally, in July 2004, further work undertaken in view of a paper presentation at the Colorado Conference on the estimation of migration to be held in Estes Park in September 2004 led to removing the no mortality/emigration assumption in force until then. As a bonus, it also brought about yet another method for estimating migration in a related context.

From the above account on the genesis of this paper, it is clear that the author is heavily indebted for the continuous encouragement and support extended to him by Andrei Rogers from the time he set forth the initial idea for this paper in 1980 to the time he eventually completed it in the summer of 2004. Without this encouragement and support, this paper would have never materialized.

INTRODUCTION

For a long time, little was known of period migration patterns in the less developed countries, essentially because the migration estimates needed for carrying out the analysis of such patterns were lacking. Indeed, until a decade or two ago, many a census carried out in the less developed countries did not include the question on the place of residence one or five years earlier which, since at least the mid-twentieth century, had been a standard feature of the censuses undertaken in the more developed countries.

At the turn of seventies, however, Rogers and von Rabenau (1971) attempted to make up for the inexistence of direct estimates of period migration streams in Brazil by suggesting an indirect estimation method relying on availability of lifetime migration data from two consecutive censuses. Unfortunately, this indirect method proved to be problematic, for it was known in some instances to lead to undesirable results, such as negative migration streams, which could not be attributed solely to measurement errors in the data used. A decade later, in the wake of an empirical study on multiregional life table construction (Ledent, 1980a), it was realized that the main culprit for the problematic character of the method was an implicit assumption—namely, an assumption of independence of destination-specific outmigration propensities vis-à-vis the place of birth—that did not conform at all with reality (Ledent, 1980b). Soon afterward, this realization led to the idea of amending the Rogers-von Rabenau method by introducing some a priori information on the birthregion differences pertaining to each destination-specific outmigration propensity but, for various reasons, it took many years before this idea could be fully implemented.

Basically, this paper sets forth an improvement to Rogers's and von Rabenau's indirect method for estimating period migration streams from lifetime migration data drawn from two consecutive censuses. Following a careful introduction to the set of parameters allowing one to get around the assumption implicit to the Rogers-von Rabenau method, it addresses the derivation of the improved estimates of the period migration streams that

can be obtained on the basis of such parameter set. More specifically, the paper begins in Section 1 with a reminder on the Rogers-von Rabenau method. Then, Section 2 introduces the set of parameters allowing one to modulate the destination-specific outmigration propensities by region of birth. From there, the exposition goes on with establishing the revision to the Rogers-von Rabenau formula. Initially, such a task is limited, in Section 3, to the simple case of a two-region system before it shifts in Section 4 to the general case of a system consisting of any number of regions. Next, Section 5 offers a demonstration of the general formula thus established by means of an illustration to a four-region system of the United States. Finally, Section 6 proposes another indirect method for estimating migration, applicable though in a different but related context, whose underlying framework readily follows from the one considered in this paper.

1. THE STARTING POINT: THE ROGERS-VON RABENAU METHOD

1.1 Description and formulation

Briefly, the Rogers-von Rabenau method enables one, with reference to a given system of regions, to estimate the migration streams having taken place between two points in time at which the lifetime migration streams are known. In other words, from the knowledge of the numbers ${}_i K_j^{(1)}$ and ${}_i K_j^{(2)}$ of those born in region i who resided in region j ($i, j = 1, \dots, N$), respectively, at times 1 and 2, it provides an estimate of the number O_{ij} of those who resided in region i at time 1 but in region j at time 2.

The method relies on a set of survivorship proportions of which the typical element s_{kj} reflects the proportion of those who among the total population

${}_i K_k^{(1)}$ that resided in region k at time 1 were found in region j at time 2 :

$$s_{kj} = \frac{O_{kj}}{{}_i K_k^{(1)}} \quad (1.1)$$

Now let ${}_i O_{kj}$ be the migration stream from region k to region j pertaining to the members of ${}_i K_k^{(1)}$ that were born in region I only (${}_i K_k^{(1)}$). Then we have that

$${}_i K_j^{(2)} = \sum_k {}_i O_{kj} \quad (1.2)$$

and on assuming that the propensity of migrating from region k to region j is independent of the region of birth—that is, it can be applied to each ${}_i K_k^{(1)}$ group—it comes

$${}_i K_j^{(2)} = \sum_k s_{kj} {}_i K_k^{(1)} \quad \text{for } i \text{ and } j = 1, \dots, N \quad (1.3)$$

This can be rewritten more compactly as :

$$\mathbf{K}^{(2)} = \mathbf{S} \mathbf{K}^{(1)} \quad (1.4)$$

where

$$\mathbf{K}^{(T)} = \begin{bmatrix} {}_1 K_1^{(T)} & {}_2 K_1^{(T)} & \dots & {}_N K_1^{(T)} \\ {}_1 K_2^{(T)} & {}_2 K_2^{(T)} & & \dots \\ \dots & & & \dots \\ {}_1 K_N^{(T)} & \dots & \dots & {}_N K_N^{(T)} \end{bmatrix}$$

and

$$\mathbf{S} = \begin{bmatrix} s_{11} & s_{21} & \dots & s_{N1} \\ s_{12} & s_{22} & & \dots \\ \dots & & & \dots \\ s_{1N} & \dots & \dots & s_{NN} \end{bmatrix}$$

Then postmultiplying both sides of (1.4) by $[\mathbf{K}^{(1)}]^{-1}$ yields:

$$\mathbf{S} = \mathbf{K}^{(2)} [\mathbf{K}^{(1)}]^{-1} \quad (1.5)$$

From there, the estimated migration streams readily follow by postmultiplying by a diagonal matrix of the initial regional populations $[\mathbf{K}^{(1)} \mathbf{i}]_d$ where

$$\mathbf{i} = \begin{bmatrix} 1 \\ 1 \\ \dots \\ 1 \end{bmatrix}$$

Thus

$$\mathbf{O} = \mathbf{S} [\mathbf{K}^{(1)} \mathbf{i}]_d = \mathbf{K}^{(2)} [\mathbf{K}^{(1)}]^{-1} [\mathbf{K}^{(1)} \mathbf{i}]_d \quad (1.6)$$

1.2 An illustration to a four-region system of the U. S.

Next is an illustration of the Rogers-von Rabenau which, with the idea of better assessing the problem inherent to this method, is carried out not in the context for which it is destined but rather in an ideal context such that the actual period streams consistent with the two sets of life migration streams available are known.

This ideal context takes advantage of the four birthplace-specific matrices of migration streams between the four regions (Northeast, Northwest, South and West) of the United States for the period 1985-1990 found in Appendix Table 5 in A. Rogers and J. Raymer (1999) and reproduced in Appendix 1. Thus, the lifetime migration streams pertaining to 1985 are made up of the row sums appearing in the four 1x4 green-framed boxes, whereas the lifetime migration streams pertaining to 1990 are made up of the column sums in the four 4x1 red-framed boxes. Application of the Rogers-von Rabenau method to these two lifetime datasets, gathered in the 4x4 green-and red-framed boxes in Table 1 eventually leads to yields the estimated migration streams in the 4x4 gold-framed box of the same table. Comparison with the actual estimates in the 4x4 tan-framed box yields the percentage errors in the 4x4 black-framed box where it appears that the independence vis-à-vis the place of birth inherent to the Rogers-von Rabenau method causes the migration streams to be underestimated by 39.4 to 86.6% and the stayer flows to be overestimated by 3.0 to 4.4%.

Clearly, the error just observed stems from the application in equation (1.3) of the same survivorship proportions to all individuals in a given region regardless of their place of birth, whereas such is not the case in the real world as exemplified by the birthregion-specific survivorship proportions appearing in the left hand-side of Table 2. As is well known of the students of migration, the propensity to move out of a given region is higher for a non-native—that is, a person born outside that region—than for a native—that is, a person born within the region. Moreover, a non-native is more likely to return to his/her region of birth than move to a third region!

2. IMPROVING ON THE ROGERS-VON RABENAU METHOD :

Clearly, if one is to improve on the Rogers-von Rabenau method, necessity is to remove the assumption underlying equation (1.3) by building into the model some modulation of the destination-specific survivorship proportions according to the region of birth.

2.1 Defining meaningful parameters reflecting birth-specific migration differentials

Basically, what is called for is to qualify the destination-specific survivorship proportions defined in (1.1) for each birthregion-specific population with the help of an i subscript placed before each relevant symbol :

$${}_i S_{kj} = \frac{{}_i O_{kj}}{{}_i K_k^{(1)}} \quad (2.1)$$

Now we may attempt to built into the improved method sought some modulation vis-à-vis the region of birth through the consideration of a set of parameters reflecting the relative importance of destination-specific survivorship proportions pertaining to non-natives vis-à-vis those pertaining to natives such as

$${}_i \alpha_{kj} = \frac{{}_i S_{kj}}{{}_k S_{kj}} \quad (2.2)$$

But the problem with such a parameter is that it is influenced by migrations within the system—which is fine—but also by exits from the system (deaths within the system and migrations out of the system)—which is less fine, if one wants to think of the α parameters as exhibiting some regularities that can be predicted in some ways. Clearly, it would be better to define the α parameters in relation to just migrations within the system—that is, migrations observed conditionally on survival in the system.

Thus, corresponding to the full (or unconditional) survivorship proportion ${}_i S_{kj}$, let us define to the conditional survivorship proportion

$${}_i \bar{S}_{kj} = \frac{{}_i O_{kj}}{{}_i O_k} \quad (2.3)$$

in which the denominator ${}_i O_k = \sum_j {}_i O_{kj}$ is simply the number of those who among the the ${}_i K_k^{(1)}$ subpopulation survive to time 2 within the system. Naturally, (2.3) implies that:

$$\sum_j {}_i \bar{s}_{kj} = 1 \quad (2.4)$$

Under such circumstances, we may attempt to improve on the Rogers-von Rabenau method through a set of α parameters reflecting the relative importance of conditional (rather than unconditional survivorship) proportions among natives and non-natives. Thus, let us use instead of (2.2)

$${}_i \alpha_{kj} = \frac{{}_i \bar{s}_{kj}}{{}_k \bar{s}_{kj}} \quad (2.5)$$

2.2 Principles and Directions

$${}_i O_{kj} = {}_i s_{kj} {}_i K_k^{(1)} = {}_i \bar{s}_{kj} {}_i O_k.$$

$${}_i s_{kj} = {}_i \bar{s}_{kj} \frac{{}_i O_k}{{}_i K_k^{(1)}}$$

$$\frac{{}_i O_k}{{}_i K_k^{(1)}} = {}_i s_k.$$

$${}_i s_{kj} = {}_i \bar{s}_{kj} {}_i s_k.$$

Since Eldridge (1965), it is customary to refer to ${}_i O_{kj}$ as a primary migration stream if it applies to natives ($k = i$) and as a second migration stream if it applies to non-natives ($k \neq i$). Moreover, in the latter case, a distinction is usually made between a return migration—

that is, a migration bound for the region of birth ($j = i$)—and an onward migration—that is, a migration bound for a third region ($j \neq i$).

$i = k$ primary

$${}_k \alpha_{kj} = \frac{{}_k \bar{s}_{kj}}{{}_k \bar{s}_{kj}} = 1$$

and $i = j$ return

$i \neq k$ and j onward

$${}_k \bar{s}_{kj} = \hat{s}_{kj}$$

$${}_i O_{kj} = {}_i s_{kj} {}_i K_k^{(1)} = {}_i \bar{s}_{kj} {}_i s_{k.} {}_i K_k^{(1)} = {}_i \alpha_{kj} \hat{s}_{kj} {}_i s_{k.} {}_i K_k^{(1)} \text{ pour } k \neq j$$

$$\begin{aligned} {}_i O_{jj} &= {}_i O_j - \sum_{k \neq j} {}_i O_{jk} \\ &= {}_i s_{j.} {}_i K_j^{(1)} - \sum_{k \neq j} {}_i s_{jk} {}_i K_j^{(1)} \\ &= {}_i s_{j.} {}_i K_j^{(1)} - \sum_{k \neq j} {}_i \alpha_{jk} \hat{s}_{jk} {}_i s_{j.} {}_i K_j^{(1)} \\ &= (1 - \sum_{k \neq j} {}_i \alpha_{jk} \hat{s}_{jk}) {}_i s_{j.} {}_i K_j^{(1)} \end{aligned}$$

$${}_i K_j^{(2)} = \sum_k {}_i O_{kj} = (1 - \sum_{k \neq j} {}_i \alpha_{jk} \hat{s}_{jk}) {}_i s_{j.} {}_i K_j^{(1)} + \sum_{k \neq j} {}_i \alpha_{kj} \hat{s}_{kj} {}_i s_{k.} {}_i K_k^{(1)}$$

$${}_i K_j^{(2)} - {}_i s_{j.} {}_i K_j^{(1)} = - \sum_{k \neq j} {}_i \alpha_{jk} {}_i s_{j.} {}_i K_j^{(1)} \hat{s}_{jk} + \sum_{k \neq j} {}_i \alpha_{kj} {}_i s_{k.} {}_i K_k^{(1)} \hat{s}_{kj} \quad (\text{A})$$

3. THE TWO-REGION CASE

Let us assume a two-region system ($N = 2$). Then for each of the two birthregions i ($=1, 2$), one can write (A) for residence region $j = 1, 2$.

$${}_i K_1^{(2)} - {}_i s_{1,i} {}_i K_1^{(1)} = - {}_i \alpha_{12} \hat{s}_{12} {}_i s_{1,i} {}_i K_1^{(1)} + {}_i \alpha_{21} \hat{s}_{21} {}_i s_{2,i} {}_i K_2^{(1)} \quad (3.1)$$

$${}_i K_2^{(2)} - {}_i s_{2,i} {}_i K_2^{(1)} = - {}_i \alpha_{21} \hat{s}_{21} {}_i s_{2,i} {}_i K_2^{(1)} + {}_i \alpha_{12} \hat{s}_{12} {}_i s_{1,i} {}_i K_1^{(1)} \quad (3.2)$$

Re-ordering the terms on the right-hand side of the latter equation leads to

$${}_i K_2^{(2)} - {}_i s_{2,i} {}_i K_2^{(1)} = {}_i \alpha_{12} \hat{s}_{12} {}_i s_{1,i} {}_i K_1^{(1)} - {}_i \alpha_{21} \hat{s}_{21} {}_i s_{2,i} {}_i K_2^{(1)} \quad (3.3)$$

that is, an equation whose right-hand side is equal to minus the same right-hand side of the former equation. By summing (3.1) and (3.3), it readily follows that

$${}_i K_1^{(2)} + {}_i K_2^{(2)} = {}_i s_{1,i} {}_i K_1^{(1)} + {}_i s_{2,i} {}_i K_2^{(1)} \quad (3.4)$$

an equation expressing the survivors of the i -born population at time 2 in terms of the corresponding survivors at time 1. (3.1) and (3.2) are linearly related and thus suffices it to keep only one of them, for example the one where the region of residence j is different from the region of birth i . Under such circumstances, we have for $i=1$

$${}_1 K_2^{(2)} - {}_1 s_{2,1} {}_1 K_2^{(1)} = {}_1 \alpha_{12} {}_1 s_{1,1} {}_1 K_1^{(1)} \hat{s}_{12} - {}_1 \alpha_{21} {}_1 s_{2,1} {}_1 K_2^{(1)} \hat{s}_{21} \quad (3.5)$$

and for $i=2$

$${}_2 K_1^{(2)} - {}_2 s_{1,2} {}_2 K_1^{(1)} = - {}_2 \alpha_{12} {}_2 s_{1,2} {}_2 K_1^{(1)} \hat{s}_{12} + {}_2 \alpha_{21} {}_2 s_{2,2} {}_2 K_2^{(1)} \hat{s}_{21} \quad (3.6)$$

Equations (3.5) and (3.6) form a linear system of two equations in two unknowns, the primary outmigration propensities, \hat{s}_{12} and \hat{s}_{21} :

$$a \hat{s}_{12} - b \hat{s}_{21} = e$$

$$-c \hat{s}_{12} + d \hat{s}_{21} = f$$

in which a, b, c and d are positive coefficients. From there, it follows that

$$\hat{s}_{12} = \frac{de + bf}{ad - bc} \quad \text{et} \quad \hat{s}_{21} = \frac{ce + af}{ad - bc}$$

where

$$ad - bc = {}_1\alpha_{12} {}_2\alpha_{21} {}_1s_{1.1} K_1^{(1)} {}_2s_{2.2} K_2^{(1)} - {}_1\alpha_{21} {}_2\alpha_{12} {}_1s_{2.1} K_2^{(1)} {}_2s_{1.2} K_1^{(1)}$$

in which ${}_1\alpha_{12} = {}_2\alpha_{21} = 1$

In general ${}_iK_j^{(T)}$ (for $j \neq i$) increases with T so that e and f are likely to be positive and the two primary outmigration propensities, \hat{s}_{12} and \hat{s}_{21} , to have the sign of $ad - bc$.

They are negative if

$${}_1\alpha_{21} {}_2\alpha_{12} > \frac{{}_1s_{1.1} K_1^{(1)} {}_2s_{2.2} K_2^{(1)}}{{}_1s_{2.1} K_2^{(1)} {}_2s_{1.2} K_1^{(1)}}$$

4. THE N-REGION CASE

For each i , N equations such as

$${}_iK_j^{(2)} - {}_iS_{j.} {}_iK_j^{(1)} = \sum_{k \neq j} {}_i\alpha_{kj} {}_iS_{k.} {}_iK_k^{(1)} \hat{s}_{kj} - \sum_{k \neq j} {}_i\alpha_{jk} {}_iS_{j.} {}_iK_j^{(1)} \hat{s}_{jk} \quad (4.1)$$

but they are linearly related

$$\sum_i {}_iK_j^{(2)} - {}_iS_{j.} {}_iK_j^{(1)} = 0 \quad (4.2)$$

The two scalar equations encountered in the treatment of the two-region case can be written in a matrix-vector format as:

$$\begin{bmatrix} {}_1K_2^{(2)} \\ {}_2K_1^{(2)} \end{bmatrix} - \begin{bmatrix} {}_1s_{2.} & 0 \\ 0 & {}_2s_{1.} \end{bmatrix} \begin{bmatrix} {}_1K_2^{(1)} \\ {}_2K_1^{(1)} \end{bmatrix} = \begin{bmatrix} {}_1\alpha_{12} {}_1s_{1.} {}_1K_1^{(1)} & -{}_1\alpha_{21} {}_1s_{2.1} K_2^{(1)} \\ -{}_2\alpha_{12} {}_2s_{1.2} K_1^{(1)} & {}_2\alpha_{21} {}_2s_{2.2} K_2^{(1)} \end{bmatrix} \begin{bmatrix} \hat{s}_{12} \\ \hat{s}_{21} \end{bmatrix} \quad (4.3)$$

Careful observation of the latter may provide useful hint regarding the possibility of extending the two-region solution to the N-region case. Note that this equation contains:

- two 2x1 vectors, each of which includes the lifetime migration streams observed at either the beginning or the end of the observed periods, which we may call $K^{(1)}$ and $K^{(2)}$, respectively,
- one 2x1 column vector containing the two primary outmigration propensities, which we may call \hat{s}
- one 2x2 diagonal matrix, the diagonal elements of which consist of classical survivorship proportions (in the system) pertaining to each region-specific cohort residing in its birthregion at the beginning of the period, which we may call \tilde{S} and finally
- a more complex 2x2 square matrix containing the α coefficients as well as other inputs.

so that (4.3) may be rewritten in a more compact fashion as:

$$K^{(2)} - \tilde{S}K^{(1)} = Z\hat{s} \quad (4.4)$$

Interestingly, the three vectors distinguished above have (scalar elements) depending on two indices referring to an origin and a destination regions, which readily suggests that generalization to the N-region case is likely to involve not a substitution of matrices for the three vectors in (4.4) but rather an expansion of these vectors to include all pairs of origin and destination regions, except those with identical origin and destination regions which are unneeded in the case of lifetime migration streams (as seen earlier, for each birth cohort, the lifetime stayer streams are linearly related to the corresponding migration streams) and the primary outmigration propensities (1 minus the sum of all primary outmigration propensities). Under such circumstances, the three vectors can be expected to be $R \times 1$ vectors and the two matrices $R \times R$ matrices where $R = N(N-1)$.

so that one can be ignored, for example, the one such that $j = i$.

Let us take up the case of a three-region system and write the three subsets of two equations (for each region of birth)

$${}_1K_2^{(2)} - {}_1s_{2,1}K_2^{(1)} = {}_1\alpha_{12}{}_1s_{1,1}K_1^{(1)}\hat{s}_{12} + {}_1\alpha_{32}{}_1s_{3,1}K_3^{(1)}\hat{s}_{32} - {}_1\alpha_{21}{}_1s_{2,1}K_2^{(1)}\hat{s}_{21} - {}_1\alpha_{23}{}_1s_{2,1}K_2^{(1)}\hat{s}_{23}$$

$${}_1K_3^{(2)} - {}_1s_{3,1}K_3^{(1)} = {}_1\alpha_{13}{}_1s_{1,1}K_1^{(1)}\hat{s}_{13} + {}_1\alpha_{23}{}_1s_{2,1}K_2^{(1)}\hat{s}_{23} - {}_1\alpha_{31}{}_1s_{3,1}K_3^{(1)}\hat{s}_{31} - {}_1\alpha_{32}{}_1s_{3,1}K_3^{(1)}\hat{s}_{32}$$

$${}_2K_1^{(2)} - {}_2s_{2,2}K_1^{(1)} = {}_2\alpha_{21}{}_2s_{2,2}K_2^{(1)}\hat{s}_{21} + {}_2\alpha_{31}{}_2s_{3,2}K_3^{(1)}\hat{s}_{31} - {}_2\alpha_{12}{}_2s_{1,2}K_1^{(1)}\hat{s}_{12} - {}_2\alpha_{13}{}_2s_{1,2}K_1^{(1)}\hat{s}_{13}$$

$${}_2K_3^{(2)} - {}_2s_{3,1}K_3^{(1)} = {}_2\alpha_{13}{}_2s_{1,2}K_1^{(1)}\hat{s}_{13} + {}_2\alpha_{23}{}_2s_{2,2}K_2^{(1)}\hat{s}_{23} - {}_2\alpha_{31}{}_2s_{3,2}K_3^{(1)}\hat{s}_{31} - {}_2\alpha_{32}{}_2s_{3,2}K_3^{(1)}\hat{s}_{32}$$

$${}_3K_1^{(2)} - {}_3s_{2,3}K_1^{(1)} = {}_3\alpha_{21}{}_3s_{2,3}K_2^{(1)}\hat{s}_{21} + {}_3\alpha_{31}{}_3s_{3,3}K_3^{(1)}\hat{s}_{31} - {}_3\alpha_{12}{}_3s_{1,3}K_1^{(1)}\hat{s}_{12} - {}_3\alpha_{13}{}_3s_{1,3}K_1^{(1)}\hat{s}_{13}$$

$${}_3K_2^{(2)} - {}_3s_{3,3}K_2^{(1)} = {}_3\alpha_{12}{}_3s_{1,3}K_1^{(1)}\hat{s}_{12} + {}_3\alpha_{32}{}_3s_{3,3}K_3^{(1)}\hat{s}_{32} - {}_3\alpha_{21}{}_3s_{2,3}K_2^{(1)}\hat{s}_{21} - {}_3\alpha_{23}{}_3s_{2,3}K_2^{(1)}\hat{s}_{23}$$

$$K^{(2)} = \tilde{S}K^{(1)}$$

where

$$K^{(T)} = \begin{bmatrix} {}_1K_2^{(T)} \\ {}_1K_3^{(T)} \\ {}_2K_1^{(T)} \\ {}_2K_3^{(T)} \\ {}_3K_1^{(T)} \\ {}_3K_2^{(T)} \end{bmatrix}$$

$$\hat{\mathbf{s}} = \begin{bmatrix} {}_1s_1. & 0 & 0 & 0 & 0 & 0 \\ 0 & {}_1s_3. & 0 & 0 & 0 & 0 \\ 0 & 0 & {}_2s_1. & 0 & 0 & 0 \\ 0 & 0 & 0 & {}_2s_3. & 0 & 0 \\ 0 & 0 & 0 & 0 & {}_3s_1. & 0 \\ 0 & 0 & 0 & 0 & 0 & {}_3s_2. \end{bmatrix}$$

$\mathbf{Z}\hat{\mathbf{s}}$ where

$$\hat{\mathbf{s}} = \begin{bmatrix} \hat{s}_{12} \\ \hat{s}_{13} \\ \hat{s}_{21} \\ \hat{s}_{23} \\ \hat{s}_{31} \\ \hat{s}_{32} \end{bmatrix}$$

$$\mathbf{Z} = \begin{bmatrix} {}_1\alpha_{12} {}_1s_{1.1} K_1^{(1)} & 0 & -{}_1\alpha_{21} {}_1s_{2.1} K_2^{(1)} & -{}_1\alpha_{23} {}_1s_{2.1} K_2^{(1)} & 0 & {}_1\alpha_{32} {}_1s_{3.1} K_3^{(1)} \\ 0 & {}_1\alpha_{13} {}_1s_{1.1} K_1^{(1)} & 0 & {}_1\alpha_{23} {}_1s_{2.1} K_2^{(1)} & -{}_1\alpha_{31} {}_1s_{3.1} K_3^{(1)} & -{}_1\alpha_{32} {}_1s_{3.1} K_3^{(1)} \\ -{}_2\alpha_{12} {}_2s_{1.2} K_1^{(1)} & -{}_2\alpha_{13} {}_2s_{1.2} K_1^{(1)} & {}_2\alpha_{21} {}_2s_{2.2} K_2^{(1)} & 0 & {}_2\alpha_{31} {}_2s_{3.2} K_3^{(1)} & 0 \\ 0 & {}_2\alpha_{13} {}_2s_{1.2} K_1^{(1)} & 0 & {}_2\alpha_{23} {}_2s_{2.2} K_2^{(1)} & -{}_2\alpha_{31} {}_2s_{3.2} K_3^{(1)} & -{}_2\alpha_{32} {}_2s_{3.2} K_3^{(1)} \\ -{}_3\alpha_{12} {}_3s_{1.3} K_1^{(1)} & -{}_3\alpha_{13} {}_3s_{1.3} K_1^{(1)} & {}_3\alpha_{21} {}_3s_{2.3} K_2^{(1)} & 0 & {}_3\alpha_{31} {}_3s_{3.3} K_3^{(1)} & 0 \\ {}_3\alpha_{12} {}_3s_{1.3} K_1^{(1)} & 0 & -{}_3\alpha_{21} {}_3s_{2.3} K_2^{(1)} & -{}_3\alpha_{23} {}_3s_{2.3} K_2^{(1)} & 0 & {}_3\alpha_{32} {}_3s_{3.3} K_3^{(1)} \end{bmatrix}$$

$$Z_{ij,pq} = \varepsilon {}_i\alpha_{pq} {}_is_{p.i} K_p^{(1)} \quad j \neq i \text{ et } p \neq q$$

$$\begin{aligned} \varepsilon &= -1 \quad \text{if } i=q \text{ and } j=p \text{ (return)} \\ &= 1 \quad \text{if } i=p \text{ and } j=q \text{ (primary)} \\ &= -1 \quad \text{if } i \neq (\text{and } p)q \text{ and } j=p \text{ (onward 1)} \\ &= 1 \quad \text{if } i \neq p \text{ (and } q) \text{ and } j=q \text{ (onward 2)} \\ &= 0 \quad \text{otherwise } (j \neq p \text{ et } q) \end{aligned}$$

$$K^{(2)} - \tilde{S}K^{(1)} = Z\hat{s}$$

$$\hat{s} = Z^{-1} [K^{(2)} - \tilde{S}K^{(1)}]$$

5. USING THE IMPROVED METHOD IN PRACTICE

$${}_i s_k = s_k.$$

$$S = \bar{S} [i'S]_d$$

$$i' = [1 \quad 1 \quad \dots \quad 1 \quad 1]$$

$$i'S = i'K^{(2)} [K^{(1)}]^{-1}$$

$$\bar{S} = S \{ [i'S]_d \}^{-1} = K^{(2)} [K^{(1)}]^{-1} \{ [i'K^{(2)} [K^{(1)}]^{-1}]_d \}^{-1}$$

$${}_i s_k \text{ is given, possibly } {}_i s_k = s_k. \quad i'S = i'K^{(2)} [K^{(1)}]^{-1}$$

5.1

$${}_i s_{kj} = {}_i \alpha_{kj} \hat{s}_{kj} \text{ pour } k \neq j$$

$${}_i O_{kj} = {}_i s_{kj} \quad {}_i K_k^{(1)} = {}_i \bar{s}_{kj} \quad {}_i s_k \quad {}_i K_k^{(1)} = {}_i \alpha_{kj} \hat{s}_{kj} \quad {}_i s_k \quad {}_i K_k^{(1)} \text{ pour } k \neq j$$

(déjà plus haut)

$$O_{kj} = \sum_i O_{kj} = [\sum_i \alpha_{kj} S_{k.} K_k^{(1)}] \hat{s}_{kj}$$

5.2 Straightforward application of the improved formula

An illustration of this improved method was carried out by setting the ${}_k\alpha_{ij}$ proportions equal to their observed values over the 1975-80 period: see the “Test 1985-90” sheet where the predetermined ${}_k\alpha_{ij}$ values appearing in the four blue-framed boxes were calculated in the “1975-80 (probs)” sheet. Again, the estimated migration flows, which appear in the 4x4 gold-framed box, are compared with the corresponding actual flows in the 4x4 tan-framed box. It appears that the thus estimated migration flows are considerably better than estimated previously as they are in error (see 4x4 black-framed box) by a proportion ranging from -25.9 to 23.5% (vs. 39.4 vs. 86.6%) and stayer flows by a proportion ranging from -0.8 to 0.3% (vs. 3.0 to 4.4%).

Correctness of the specification of the linear system of equations that underlies the improved method as well as adequacy of its implementation in the attached Excel file can be verified by applying the model to a situation for which the values of the migration flows to be estimated are known. Two such situations were investigated and they brought out the expected results:

- First, the ${}_k\alpha_{ij}$ values were set to their observed values over the 1985-90:
As shown in the “Test 1985-90 (actual)” sheet, the estimated migration flows thus obtained are identical to the actual migration flows.
- Second, the ${}_k\alpha_{ij}$ values were all set equal to 1 (so that the destination-specific outmigration propensities are independent of the birthplace):
As shown in the “Test 1985-90 (Markov)” sheet, the estimated migration flows thus obtained are identical to the migration flows obtained by a straightforward application of the Rogers-von Rabenau method.

Finally, the migration flows were estimated again by setting the ${}_k\alpha_{ij}$ proportions equal to their observed values over the 1965-70 period (rather than in the 1975-80 period): see the “Test 1985-90 (2)” sheet in which the predetermined ${}_k\alpha_{ij}$ values in the four blue-framed boxes were calculated in the sheet “1965-70 (probs)”. As one would expect (since the ${}_k\alpha_{ij}$ relate to a period removed from twenty rather than ten years), the estimates obtained are much worse than when the ${}_k\alpha_{ij}$ proportions were assigned to their 1975-80 values. The migration flows are in error by a proportion ranging from -77.9 to 51.4%, whereas the stayer flows are in error by a proportion ranging from -1.5 to 0.7% (see 4x4 black-framed box).

5.3 An alternative application based on an iterative technique

Application of the improved formula (?.?) involves inverting a $R \times R$ matrix where $R = N(N-1)$ increases with the square of the number of regions N . More specifically, if N is being set equal to successive values of 2, 3, 4, 5, ..., R goes up more rapidly, taking successive values of 2, 6, 12, 20, Clearly, as N increases, inverting the \mathbf{Z} matrix becomes a more and more complicated matter, thus rendering the straightforward application of the improved formula (?.?) less effective. Were this formula applied in relation to an example for the 50-state system of the U. S., the \mathbf{Z} matrix to be inverted would be a 2450 x 250 matrix!!

Fortunately, there is an alternative to applying the new formula (?.?) that is capable of delivering more efficiently the right result. Such an application involves solving iteratively equation (?.?) from which formula (?.?) is derived, thanks to the linear nature of the system of scalar equations behind (?.?). Basically, suffices it to let the \hat{s} vector appear on both sides of (?.?) so as to obtain an expression of \hat{s} in terms of itself, capable of underlying an iterative derivation of the \hat{s} vector. The idea is to pick an initial (first) estimate of \hat{s} before introducing it on the right-hand side of the latter equation so as to obtain another (second) estimate of \hat{s} and then repeat the above process until convergence defined by an appropriate criterion is obtained.

6. EPILOGUE: ESTIMATING PRIMARY AND SECONDARY MIGRATION FROM A SINGLE CENSUS

Before moving along to the conclusion, it appears useful to this author to digress a little bit and to take up another problem in the estimation of migration that, as it will be soon made clear, may be solved directly on the basis of the equation (6.1) central to this paper. Basically, this problem boils down to estimating the primary (secondary) migration streams that are associated with the total migration streams drawn from the same census.

Equation (6.1) still holds:

$$K^{(1)} = \tilde{S}^{-1} [K^{(2)} - Z \hat{s}]$$

but, further to the knowledge of the total migration streams, we also have that:

$$\hat{s}_{kj} = \frac{O_{kj}}{\sum_i \alpha_{kj} S_{k.} K_k^{(1)}}$$

Then, if O and $K^{(2)}$ are known from a single census, possibility of estimating \hat{s} over the observation period as O requires one to derive $K^{(1)}$ while setting \tilde{S} equal to a unity matrix, while imposing no restriction whatsoever.

In practice, this can be accomplished through an iterative technique that consists of picking an initial estimate of $K^{(1)}$ and then deriving in succession:

- a first estimate of \hat{s} on the basis of

$$\hat{s}_{kj} = \frac{O_{kj}}{\sum_i \alpha_{kj} K_k^{(1)}} \quad (6.2)$$

- a second estimate of $K^{(1)}$ on the basis of

$$K^{(1)} = [K^{(2)} - Z \hat{s}] \quad (6.3)$$

- a second estimate of \hat{s} on the basis of (6.2)

- a second estimate of $K^{(1)}$ on the basis of (6.3)

etc... until convergence of the process.

Once again, the iterative technique just described is illustrated in relation to the four-region system of the U. S. previously considered, using the data on i) lifetime migration streams and ii) total migration streams over the last five years coming from the 1990 census. This illustration is carried out in three steps, mimicking the same approach as the one used in section 5 with regard to the iterative estimation of period migration streams from two lifetime migration datasets.

Step One: The α coefficients are set equal to their observed values:

The successive values of $\mathcal{K}^{(t)}$ are shown in Panel A of Table 6.1, along with the final estimates of the primary outmigration propensities which as expected are identical to the corresponding actual values in Table ??

Step Two: The α coefficients are set equal to 1

The successive values of $\mathcal{K}^{(t)}$ are shown in Panel B of Table 6.1, along with the final estimates of the primary outmigration propensities which as expected are identical to the corresponding total outmigration propensities in Table ??

Step Three : The α coefficients are set equal to their 1975-1980 values

The successive values of $\mathcal{K}^{(t)}$ are shown in Panel C of Table 6.1, along with the final estimates of the primary outmigration propensities as well as the error committed.

CONCLUSION

To summarize, this paper has set forth and demonstrated an alternative to the Rogers-von Rabenau method for estimating period migration streams from two lifetime migration datasets and, as a bonus, has given rise to a related method for estimating primary migration patterns from the information commonly available in a single census. But since the two methods proposed rest upon the prior knowledge of a set of coefficients reflecting differences in destination-specific outmigration propensities according to the birthplace, a final conclusion on their usefulness hinges on the possibility of picking appropriate values for these coefficients. It has been shown here that use of their recently observed values leads to quite acceptable migration streams but, in many instances, such values will not be available; which suggests the necessity for the analyst to pick adequate values from the little information to them.

This suggests that the methods hinge on the possibility of ascertaining spatio-temporal regularities exhibited by the coefficients. Thus addressing and establishing such regularities is thus an item to be placed at the top of research agenda pertaining in migration estimation.

REFERENCES

- Eldridge, H. T., 1965, "Primary, secondary and return migration in the United States, 1955-60," *Demography* 2:444-455.
- Ledent, J., 1980a. *Constructing multiregional life tables using place-of-birth-specific migration data*. WP-80-96, International Institute for Applied Systems Analysis, Laxenburg, Austria.
- Ledent, J., 1980b. *A note on the estimation of interregional migration streams from place-of-residence-by-place-of-birth (PRPB) data*. WP-80-106, International Institute for Applied Systems Analysis, Laxenburg, Austria.
- Rogers, A. and B. von Rabenau, 1971. "Estimation of interregional migration streams from place-of-birth-by-residence data," *Demography* 8(2):185-194.
- Rogers, A. and J. Raymer, 1999. *Origin Dependence, Return Migration, and the Spatial Distribution of the U. S.-Born Population*. WP-99-6. Population Program, Institute of Behavioral Science, University of Colorado, Boulder, Colorado.
- Rogers, A., J. Raymer and L. Jordan, 2003. *Inferring migration flows from birthplace-specific population stocks*. Population Program, Institute of Behavioral Science, University of Colorado, Boulder, Colorado.