Inferring period migration streams from two lifetime migration datasets

Jacques Ledent
INRS-UCS and CIED
Montreal QC (Canada)

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

Objectives

- Main :

  Estimation of migration streams between two points in time from lifetime migration (PRPB) datasets pertaining to each point

- Other :

  Estimation of primary (and secondary) migration streams from a single census

The starting point

- If period migration streams are not available, can one estimate them from two lifetime migration data sets: ${}_iK_j^{(1}$ and ${}_iK_j^{(2}$?
- The Rogers-von Rabenau method (1971):

$$ {}_iK_j^{(2)} = {}_k\Sigma\, s_{kj}\, {}_iK_k^{(1)} $$

where s_{kj} is the proportion of those born in k surviving in j or, in matrix format,

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

- Such a method is problematic because in the real world s_{kj} is not independent of the birthplace i (see “cooked” example with error-free measurement)

1985 PRPB flows (row : 1985 res - column : birthplace)				
	Northeast	Midwest	South	West
Northeast	37621064	1196866	2391730	458857
Midwest	1529050	46300033	4480479	1035637
South	5278888	5911281	58911619	1891081
West	2781341	6315393	4096061	26224345

1990 PRPB flows (row : 1990 res - column : birthplace)				
	Northeast	Midwest	South	West
Northeast	36517344	1267642	2420715	503894
Midwest	1581200	45296036	4567451	1105357
South	6148733	6491785	58591285	2027615
West	2963066	6668110	4300438	25973054

Estimated Migrations (row : origin - column : destination)				
	Northeast	Midwest	South	West
Northeast	40428301	68826	972450	198940
Midwest	97922	52142347	653379	451550
South	102036	208341	71441771	240721
West	81336	130529	191817	39013457

Actual migrations				
	Northeast	Midwest	South	West
Northeast	39265635	329847	1603751	469284
Midwest	343920	50381870	1672414	946995
South	760298	1180046	68930484	1122041
West	339742	658281	1052769	37366348

Relative errors (%)				
	Northeast	Midwest	South	West
Northeast	3,0	-79,1	-39,4	-57,6
Midwest	-71,5	3,5	-60,9	-52,3
South	-86,6	-82,3	3,6	-78,5
West	-76,1	-80,2	-81,8	4,4

Improving on the Rogers-von Rabenau method (I)

- Make s_{kj} dependent on the birthplace i
- We know that :

  $${}_{i}s_{kj} > {}_{k}s_{kj}$$

  Non-native (secondary) prop.> Native (primary) prop.

  and for non-natives

  $${}_{i}s_{ki} > {}_{i}s_{kj}$$

  Return (j = i) prop. >Onward (j ≠ i) prop.
- Then the idea is to assume that ${}_{i}s_{kj}$ is linked to the corresponding primary propensity ${}_{k}s_{kj}$

Improving on the Rogers-von Rabenau method (II)

- For example, we may define ${}_i\alpha_{jk}$ be the ratio of ${}_is_{jk}$ to the corresponding primary migration propensity ${}_ks_{jk}$

$$ {}_i\alpha_{kj} = \frac{{}_is_{kj}}{{}_ks_{kj}} $$

- The problem though is that ${}_is_{jk}$ and thus ${}_i\alpha_{jk}$are affected by mortality / emigration => Instead of working with unconditional migration propensities, we will work with conditional migration propensities—that is, propensities that are conditional on survival within the system :

$$ {}_is_{kj} = \frac{{}_iO_{kj}}{{}_iK_k^{(1)}} \qquad {}_i\bar{s}_{kj} = \frac{{}_iO_{kj}}{{}_iO_{k.}} $$

Improving on the Rogers-von Rabenau method (III)

- Let us define ${}_i\alpha_{jk}$ be the ratio of ${}_i\bar{s}_{kj}$ to the corresponding primary migration propensity ${}_k\bar{s}_{kj}$

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

- Given knowledge of :

  1) the full set of α coefficients and

  2) the total survivorship proportions within the system ${}_is_{k.}$

  we may estimate the various primary migration proportions (conditional on survival within the system)

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

  and from there the primary migration streams and finally the total migration streams

Northeastern-born				
	Northeast	Midwest	South	West
Northeast		1	1	1
Midwest	16,13456		2,60717	1,896528
South	14,26346	1,656227		2,159798
West	13,68037	1,472481	2,699238	

Southern-born				
	Northeast	Midwest	South	West
Northeast		2,139558	3,389963	1,758957
Midwest	1,633935		3,336322	1,043726
South	1	1		1
West	1,82942	1,825314	6,481271	

Midwestern-born				
	Northeast	Midwest	South	West
Northeast		15,15975	2,343032	4,056978
Midwest	1		1	1
South	2,256801	12,26886		3,092806
West	1,384012	7,340066	1,847982	

Western-born				
	Northeast	Midwest	South	West
Northeast		4,702857	2,347743	15,25112
Midwest	2,932777		2,18283	8,285538
South	3,074073	3,4343		13,91487
West	1	1	1	

The two-region case

- For each birthplace i, one can write one equation reflecting change in the group of those residing in each region but, given that the total survivorship proportions are known, the two equations are linearly related => One equation is enough [Take out the one with place of residence = birthplace]
- Two regions => Linear system of two equations in two variables, $\hat{s}_{12}$ and $\hat{s}_{21}$

The two-region case (cont'd)

$$\begin{bmatrix} {}_1K_2^{(2)} \\ {}_2K_1^{(2)} \end{bmatrix} - \begin{bmatrix} {}_1s_{2.} & 0 \\ 0 & {}_2s_{1.} \end{bmatrix} \begin{bmatrix} {}_1K_2^{(2)} \\ {}_2K_1^{(2)} \end{bmatrix} = \begin{bmatrix} {}_1\alpha_{12}\,{}_1s_{1.}\,{}_1K_1^{(1)} & -{}_1\alpha_{21}\,{}_1s_{2.}\,{}_1K_2^{(1)} \\ -{}_2\alpha_{12}\,{}_2s_{1.}\,{}_2K_1^{(1)} & {}_2\alpha_{21}\,{}_2s_{2}\,{}_2K_2^{(1)} \end{bmatrix} \begin{bmatrix} \hat{s}_{12} \\ \hat{s}_{21} \end{bmatrix}$$

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

- Can we generalize it to the N-region case?

The N-region case

- For each birthplace i, one can write an equation reflecting change in the group of those residing in each region but, given that the total survivorship proportions are known, the N equations are linearly related => N-1 equations are enough [Take out the one with place of residence = birthplace]
- N regions => Linear system of N(N-1) equations in N(N-1) variables $\hat{s}_{kj}$ where $j \neq k$

The N-region case (cont'd)

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

$$K^{(T)} = \begin{bmatrix} {}_1K_2^{(T)} \\ {}_1K_3^{(T)} \\ {}_2K_1^{(T)} \\ {}_2K_3^{(T)} \\ {}_3K_1^{(T)} \\ {}_3K_2^{(T)} \end{bmatrix} \quad \widetilde{\mathbf{S}} = \begin{bmatrix} {}_1s_{1.} & 0 & 0 & 0 & 0 & 0 \\ 0 & {}_1s_{3.} & 0 & 0 & 0 & 0 \\ 0 & 0 & {}_2s_{1.} & 0 & 0 & 0 \\ 0 & 0 & 0 & {}_2s_{3.} & 0 & 0 \\ 0 & 0 & 0 & 0 & {}_3s_{1.} & 0 \\ 0 & 0 & 0 & 0 & 0 & {}_3s_{2.} \end{bmatrix} \quad \hat{s} = \begin{bmatrix} \hat{s}_{12} \\ \hat{s}_{13} \\ \hat{s}_{21} \\ \hat{s}_{23} \\ \hat{s}_{31} \\ \hat{s}_{32} \end{bmatrix}$$

Z is an N(n-1) x N(N-1) matrix

$$
\mathbf{Z}=\begin{bmatrix}
{}_1\alpha_{121}s_{1.1}K_1^{(1)} & 0 & -{}_1\alpha_{211}s_{2.1}K_2^{(1)} & -{}_1\alpha_{231}s_{2.1}K_2^{(1)} & 0 & {}_1\alpha_{321}s_{3.1}K_3^{(1)} \\
0 & {}_1\alpha_{131}s_{1.1}K_1^{(1)} & 0 & {}_1\alpha_{231}s_{2.1}K_2^{(1)} & -{}_1\alpha_{311}s_{3.1}K_3^{(1)} & -{}_1\alpha_{321}s_{3.1}K_3^{(1)} \\
-{}_2\alpha_{122}s_{1.2}K_1^{(1)} & -{}_2\alpha_{132}s_{1.2}K_1^{(1)} & {}_2\alpha_{212}s_{2.2}K_2^{(1)} & 0 & {}_2\alpha_{312}s_{3.2}K_3^{(1)} & 0 \\
0 & {}_2\alpha_{132}s_{1.2}K_1^{(1)} & 0 & {}_2\alpha_{232}s_{2.2}K_2^{(1)} & -{}_2\alpha_{312}s_{3.2}K_3^{(1)} & -{}_2\alpha_{322}s_{3.2}K_3^{(1)} \\
-{}_3\alpha_{123}s_{1.3}K_1^{(1)} & -{}_3\alpha_{133}s_{1.3}K_1^{(1)} & {}_3\alpha_{213}s_{2.3}K_2^{(1)} & 0 & {}_3\alpha_{313}s_{3.3}K_3^{(1)} & 0 \\
{}_3\alpha_{123}s_{1.3}K_1^{(1)} & 0 & -{}_3\alpha_{213}s_{2.3}K_2^{(1)} & -{}_3\alpha_{233}s_{2.3}K_2^{(1)} & 0 & {}_3\alpha_{323}s_{3.3}K_3^{(1)}
\end{bmatrix}
$$

in which

$$Z_{ij,pq} = \varepsilon_i \alpha_{pq} s_{p.i} K_p^{(1)} \quad j \neq i \; et \; p \neq q$$

where

$$\varepsilon = -1 \quad if \; i = q \; and \, j = p \, (return)$$

$$= \; 1 \quad if \; i = p \; and \, j = q \, (primary)$$

$$= -1 \quad if \; i \neq q \; (and \, p) \; and \, j = p \, (onward1)$$

$$= \; 1 \quad if \; i \neq p \; (and \, q) \, and \, j = q \, (onward2)$$

$$= \; 0 \quad otherwise \, (j \neq p \, et \, q)$$

Estimation in practice

- Method I (matrix inversion)

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

- Method II (iterative) using

$$K^{(2)} - \widetilde{\mathbf{S}} K^{(1)} = [\mathbf{Z}_d + \mathbf{Z}_{od}]\hat{s}$$

in which Z has been broken down into two separate matrices (one with diagonal elements and the other with off-diagonal elements)

Actual values

Relative errors (%)				
	Northeast	Midwest	South	West
Northeast	0,0	0,0	0,0	0,0
Midwest	0,0	0,0	0,0	0,0
South	0,0	0,0	0,0	0,0
West	0,0	0,0	0,0	0,0

All equal to 1

Relative errors (%)				
	Northeast	Midwest	South	West
Northeast	3,0	-79,1	-39,4	-57,6
Midwest	-71,5	3,5	-60,9	-52,3
South	-86,6	-82,3	3,6	-78,5
West	-76,1	-80,2	-81,8	4,4

Equal to 75-80

Relative errors (%)				
	Northeast	Midwest	South	West
Northeast	-0,5	-3,4	16,3	-7,8
Midwest	6,0	0,0	13,4	-25,9
South	23,5	15,0	-0,8	15,7
West	4,2	-25,1	4,3	0,3

Then we have that

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

- Pick a first estimate of $\hat{s}$ (= 0)
- Obtain a second estimate of $\hat{s}$ using the above equation
- And so on until convergence

Epilogue : Estimating primary and secondary migration from a single census

- Given knowledge of 1) the set of ${}_iK_j^{(2)}$ and the set of period migration streams ${}_{.}O_{kj}$, one may estimate in a single procedure:
  - the set of ${}_iK_j^{(1)}$ and
  - the set of $\hat{s}_{kj}$
- For this, use two equations:
  - 1) $K^{(2)} - \tilde{\mathbf{S}}\ K^{(1)} = \mathbf{Z}\ \hat{\mathbf{s}}$ still holds but $\tilde{\mathbf{S}} = \mathbf{I}$
  - 2) ${}_{.}O_{kj} = \hat{s}_{kj}\,[\sum_i {}_i\alpha_{kj}\ {}_is_{k.}\ {}_iK_k^{(1)}\ ]$

The (iterative) estimation method

- Pick a first estimate of the ${}_iK_k^{(1)}$ $(= {}_iK_k^{(2)})$ and derive a first estimate of the primary migration propensities from

$$\hat{s}_{kj} = \frac{{}_{.}O_{kj}}{\sum_i {}_i\alpha_{kj}\ {}_is_{k.}\ {}_iK_k^{(1)}}$$

- Obtain a new estimates of the ${}_iK_k^{(1)}$ from

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

- and so on until convergence

Actual values

Pourcentage error				
	Northeast	Midwest	South	West
Northeast		0,0	0,0	0,0
Midwest	0,0		0,0	-0,1
South	0,0	0,0		0,0
West	0,0	0,0	0,0	

All equal to 1

Pourcentage error				
	Northeast	Midwest	South	West
Northeast		51,3	19,1	28,8
Midwest	52,4		26,5	17,1
South	113,0	103,7		59,6
West	104,3	113,5	82,6	

Equal to 75-80

Pourcentage error				
	Northeast	Midwest	South	West
Northeast		-25,3	-1,1	-31,8
Midwest	7,0		-4,3	-21,2
South	-1,8	-4,4		-5,5
West	18,0	-1,6	-2,6	

Conclusion

Utility of all this :

- Hinges on prior knowledge of the α-coefficients (α-coefficients from previous time period seem to do OK job)
- Requires a better understanding of the REGULARITIES observed in the α-coefficients at the AGE-SEX level
- Calls for a method allowing one to pick good values of the α-coefficients in cross-sectional/temporal contexts