Working Paper

A NOTE ON THE ESTIMATION OF
INTERREGIONAL MIGRATION STREAMS
FROM PLACE-OF-RESIDENCE-BY-
PLACE-OF-BIRTH (PRPB) DATA

Jacques Ledent

June 1980
WP-80-106

International Institute for Applied Systems Analysis
A-2361 Laxenburg, Austria

A NOTE ON THE ESTIMATION OF INTERREGIONAL MIGRATION STREAMS FROM PLACE-OF-RESIDENCE-BY-PLACE-OF-BIRTH (PRPB) DATA

Jacques Ledent

June 1980
WP-80-106

INTERNATIONAL INSTITUTE FOR APPLIED SYSTEMS ANALYSIS
A-2361 Laxenburg, Austria

FOREWORD

Interest in human settlement systems and policies has been a central part of urban-related work at the International Institute for Applied Systems Analysis (IIASA) from the outset. From 1975 through 1978 this interest was manifested in the work of the Migration and Settlement Task, which was formally concluded in November 1978. Since then, attention has turned to dissemination of the Task's results, to the conclusion of its comparative study, and to the exploration of possible future work that might apply the newly-developed mathematical methodology to other research topics.

This paper considers the impact of the Markovian assumption frequently made regarding interregional migration patterns and shows that such an assumption creates inaccuracies in procedures for inferring migration streams from place of birth data.

Papers summarizing previous work on migration and settlement at IIASA are listed at the back of this paper.

Andrei Rogers
Chairman
Human Settlements
and Services Area

ABSTRACT

This note attempts to assess the impact of the assumption that interregional migration patterns are independent of the birthplace on the accuracy of the method suggested by Rogers and von Rabenau (1971) to infer interregional migration streams from place-of-residence-by-place-of-birth (PRPB) data.

This assumption is shown to be the main element responsible for the relative inaccuracies of the method, a result which suggests that the use of an additional subscript to identify the birthplace, in matrix generalizations of some standard mathematical models of population growth and distribution, should be made with caution.

TABLE OF CONTENTS

A NOTE ON THE ESTIMATION OF INTERREGIONAL MIGRATION STREAMS FROM PLACE-OF-RESIDENCE-BY-PLACE-OF-BIRTH (PRPB) DATA

INTRODUCTION

Traditionally, the analysis of internal migration focuses on its role as a component of population change in a given geographical unit (region, administrative area, urban area, and so forth). Most migration studies are in effect devoted to the discussion of net migration, i.e., the balance of entries and exits. The data used in those studies are generally measured by using census information, since it is usually the only source of available data.

Generally this data is obtained with a direct census question about migration, but frequently it has to be estimated using information from two consecutive censuses. In the latter case, the method consists of estimating population growth expected in the absence of migration and comparing it with the observed intercensal migration (Lee et al. 1975). When reliable vital statistics are not available, the expected growth in the absence of migration has to be estimated through a survival rate method in which the survival rates are assumed to be specific to the place of birth (Eldridge and Kim 1968).

The traditional single-region focus in analyzing migration, however, reveals only the "tip of the iceberg". For example, when examining the pattern of migration within a system of regions, it forces one to examine the regional net migration flows independently of each other. In other words, this approach does not allow an investigation of the interdependence between regions. Hence, motivated by planning purposes to search for a better understanding of internal migration, demographers are gradually shifting their interest from regional net migration flows to interregional migration streams. This remark applies to developed countries--in which the interaction between regions tends to increase with economic growth--as well as to developing countries--in which the transfer of population between rural and urban areas is no longer primarily unidirectional.

The transition from the net migration flow approach to the interregional migration stream approach has, in a large part, been facilitated by the development of a formal demography (Rogers 1968, 1975 and 1979) but it has been hampered by a data problem that is most severe in the developing countries. Most censuses in developed countries contain a direct question on migration which allows one to produce tables of interregional migration streams observed over a fixed period preceding the census. By contrast, censuses in developing countries, if they include any question concerning migration at all, generally include a question regarding the place of birth which by comparison with the place of residence at time of the census allows one to produce tables of interregional streams of lifetime migration. Thus, for such countries, the problem is one of estimating interregional migration streams for a recent period.

For this purpose, a method has been proposed by Rogers and von Rabenau (1971) which generalizes the method used in the traditional net migration analysis mentioned above (Eldridge and Kim 1968). It estimates intercensal migration streams from place-of-residence-by-place-of-birth (PRPB) data in two consecutive censuses. Unfortunately, the method often leads to inaccuracies, such as the presence of negative flows, especially in the case of older age groups. This raises questions regarding the reliability of the method, something that we attempt to answer in this note.

This essay consists of three sections. Section One, intended as a background section, provides a brief description of the Rogers-von Rabenau PRPB method and discusses its weaknesses. Then, with the help of place-of-birth-specific migration stream data for the female population of the United States taken from the 1970 Census (US Bureau of the Census 1973), Section Two presents a numerical assessment of this method to test the adequacy of the main underlying assumption. This leads to the formulation of some conjectures about the accuracy of the method which, in Section Three, are analytically justified with regard to the case of a two-region system.

1. THE ROGERS-VON RABENAU PRPB METHOD: AN OVERVIEW

The PRPB method is based on the multiregional model of population growth and distribution developed by Rogers (1968, 1975). This model can be expressed in compact form as

$$\{w^{(t+n)}\} = \underset{\sim}{G}\{w^{(t)}\} \tag{1}$$

where $\{w^{(t)}\}$ is a vector consisting of K subvectors (one for every age group) each of which contains R elements (one for each region) representing the size of the population at time t in the corresponding age group, $\underset{\sim}{G}$ is a growth matrix operator similar to the classic Leslie matrix but with R × R submatrices $\underset{\sim}{S}_x$ substituted for the usual survivorship proportions s_x (similarly, submatrices $\underset{\sim}{B}_x$ replace the usual fertility elements b_x). Typically, the (i,j)-th element ${}^{ji}S_x$ of $\underset{\sim}{S}_x$ represents the proportion of those aged x to x+n (n being the length of the time and age interval) in region i who survive in region j at the end of an n-year time interval. Thus

$$\{w_{x+n}^{(t+n)}\} = \underset{\sim}{S}_x \{w_x^{(t)}\} \qquad (2)$$

where $\{w_x^{(t)}\}$ is the R × 1 subvector of $\{w^{(t)}\}$ relating to those aged x to x+n at time t.

Equation (2) suggests that the availability of age-specific population data by region in two consecutive censuses could lead to the estimation of $\underset{\sim}{S}_x$--i.e., the estimation of intercensal interregional migration streams--especially if the data available would be consistent with a matrix extension of the equation. Thus, Rogers and von Rabenau (1971) make the assumption that the migration-mortality pattern of the various regional populations contained in $\{w_x^{(t)}\}$ is independent of their place of birth, which enables them to generalize (2) into

$$\underset{\sim}{W}_{x+n}^{(t+n)} = \underset{\sim}{S}_x \, \underset{\sim}{W}_x^{(t)} \qquad (3)$$

where $\underset{\sim}{W}_x^{(t)}$ is a S × S matrix in which the i-th column represents the regional distribution of those born in the i-th region, who at time t are aged x to x+n: typically, the (i,j)-th element ${}^{ji}W_x^{(t)}$ represents the number of those born in region j who at time t are aged x to x+n and living in region i.

Observing that

$$\underset{\sim}{S}_x = \underset{\sim}{W}_{x+n}^{(t+n)} \left[\underset{\sim}{W}_x^{(t)}\right]^{-1} \qquad (4)$$

it follows that the availability of place-of-residence-by-place-of-birth (PRPB) data in two consecutive censuses apparently makes it possible to estimate intercensal interregional migration streams.

Rogers and von Rabenau (1971) applied this method to the case of PRPB data for white females in the two-region system of the East North Central Division and the Rest of the United States in 1950 and 1960. They found a few obvious inaccuracies such as the presence of negative elements in some of the submatrices $\underset{\sim}{s}_x$ of survivorship proportions, which indeed casts some doubts about the soundness of the method.

As Rogers and von Rabenau (1971) themselves indicate, there are several factors which are acting to produce such a result. First, there are errors related to the production of the data, such as (1) errors in age reporting; (2) errors in enumeration; (3) errors in reporting the place-of-birth. Second, the method is based on a rather questionable assumption: it is a well-known fact that survivorship proportions <u>are not independent</u> of the place of birth of the individuals concerned.

It is likely that the former factor--the data reliability problem--is of smaller importance in accounting for the inaccuracies observed by Rogers and von Rabenau. It appears that even in a situation relatively free of any data reliability problem, such inaccuracies would remain in existence. This is illustrated in the next section, which also attempts to assess the impact of the aforementioned methodological assumption on the results obtained.

2. THE IMPACT OF THE PRINCIPAL UNDERLYING ASSUMPTION: AN EMPIRICAL TEST

The US Bureau of the Census (1973) reports place-of-birth specific migration stream data within the United States between 1965 and 1970 (see Table 11, pages 91-430). The data for the female population of each of the 10 age groups considered: 0-4, 5-9, 10-14, 14-19, 20-24, 25-29, 30-39, 40-49, 50-50 and 60+ in 1965 have been aggregated to yield, for each of the age groups concerned, two matrices of lifetime migration streams between the four US Census Regions: one denoted by $\underset{\sim}{K}_x^{(1965)}$, relates to the transitions observed between the time of birth and 1965, while the other, denoted by $\underset{\sim}{K}_x^{(1970)}$, relates to the transitions observed between the time of birth and 1970 (the age subscript x relates to the actual age in 1965). Typically, the (i,j)-th element ${}^{ji}K_x^{(t)}$ of $\underset{\sim}{K}_x^{(t)}$ represents the number of those born in region i belonging to age group x and living in region j at time t.

Let $\underset{\sim}{S}_x$ denote the transition probability matrix between 1965 and 1970 corresponding to those in age group x in 1965. Then, in accordance with the method proposed by Rogers and von Rabenau, we can obtain an estimate $\hat{S}_x$ of this matrix from

$$\hat{\underset{\sim}{S}}_x = \underset{\sim}{K}_x^{(1970)} \left[\underset{\sim}{K}_x^{(1965)}\right]^{-1} \quad , \qquad (5)$$

or from

$$\hat{\underset{\sim}{S}}_x = \underset{\sim}{T}_x^{(1970)} \left[\underset{\sim}{T}_x^{(1965)}\right]^{-1} \qquad (6)$$

where $\underset{\sim}{T}_x^{(t)}$ is a stochastic matrix obtained by dividing each element of $\underset{\sim}{K}_x^{(t)}$ by the sum of the elements in the column to which it belongs.

Note that the above application of the Rogers-von Rabenau method is performed under circumstances which significantly differ from the conditions underlying the normal utilization of the Rogers and von Rabenau method. First, the population of reference is the same at both ends of the time interval, hence the possibility of errors due to the implicit treatment of mortality is ruled out. Second, the two sets of lifetime migration streams required by the application of the method are known from a single census, which considerably diminishes the possibility of errors due to census enumeration.

The interest of such an application of the Rogers-von Rabenau method is to reduce to a minimum the impact of errors concerning the production of data in order to produce a more meaningful test of the impact of the principal underlying assumption on the results obtained.

In Table 1, we show, as exhibits (a) and (b), the two sets of transition probabilities $\underset{\sim}{T}_x^{(1965)}$ and $\underset{\sim}{T}_x^{(1970)}$ relative to the group of females aged 20 to 24 in 1964. The estimate $\hat{S}_{20}$ following from the application of (5) to these two sets appears as exhibit (c) in the same table: it does not indicate any obvious inaccuracy. However, as in Rogers and Rabenau's illustration, some of the matrix estimates have elements which are negative or greater than one: see for example, exhibit (c) in Table 2 which shows the estimate $\hat{\underset{\sim}{S}}_{60+}$.

The above finding; namely, the result that $\hat{\underset{\sim}{S}}_x$ is not necessarily a stochastic matrix, is hardly surprising in view of the fact that the product of a stochastic matrix by the inverse of a stochastic matrix yields a matrix which has column sums of unity but may have negative elements (see Harary, Lipstein, and Styan 1970, p. 1172).

Actually, in the case of our application, such inaccuracies appear with the sixth age group (one negative element) and tend to increase in number with age (three negative elements in the case of the seventh age group; six negative elements and two elements greater than one for the eighth age group; two additional negative elements in the case of the ninth age group).*

Is there any rational explanation to account for such a result? The answer to this seems to be positive as we note that the off-diagonal elements of $\hat{\underset{\sim}{S}}_x$ are of the same magnitude as the difference between the corresponding elements of $\underset{\sim}{T}_x^{(1970)}$ and $\underset{\sim}{T}_x^{(1965)}$. Such a result which can be easily checked from the numerical illustrations provided in Tables 1 and 2 is rigorously demonstrated in Section Three for the two-region case.

---

*Note that the similarity of this result with the one reported in Ledent (1978), concerning the implementation of the so-called Option 2 method developed by Rogers (1975) to calculate a multiregional life table from observed data on region-specific survivorship proportions. The same problem is at work work in both cases.

Table 1. Transition probability matrices in the US four-region system for females aged 20-24 in 1965.

To	From Northeast	North Central	South	West
	(a) $\underset{\sim}{T}_{20}^{(1965)}$			
Northeast	0.8544	0.0210	0.0493	0.0177
North Central	0.0365	0.8213	0.0943	0.0493
South	0.0647	0.0590	0.7902	0.0514
West	0.0443	0.0988	0.0663	0.8817
	(b) $\underset{\sim}{T}_{20}^{(1970)}$			
Northeast	0.8276	0.0262	0.0536	0.0210
North Central	0.0420	0.7933	0.1006	0.0519
South	0.0741	0.0673	0.7738	0.0575
West	0.0563	0.1132	0.0720	0.8697
	(c) $\hat{\underset{\sim}{S}}_{20}$ (estimated)			
Northeast	0.9676	0.0063	0.0064	0.0037
North Central	0.0069	0.9644	0.0115	0.0041
South	0.0119	0.0106	0.9767	0.0074
West	0.0136	0.0187	0.0055	0.9848
	(d) $\underset{\sim}{S}_{20}$ (actual)			
Northeast	0.9118	0.0184	0.0265	0.0207
North Central	0.0226	0.8983	0.0392	0.0423
South	0.0470	0.0439	0.9052	0.0488
West	0.0251	0.0394	0.0291	0.8882

Source: Calculated using data reported in US Bureau of the Census (1973).

Table 2. Transition probability matrices in the US four-region system for females aged 60+ in 1965.

To	From Northeast	North Central	South	West
	(a) $\underset{\sim}{T}_{60+}^{(1965)}$			
Northeast	0.8387	0.0181	0.0367	0.0127
North Central	0.0403	0.7380	0.0763	0.0304
South	0.0733	0.0726	0.8238	0.0252
West	0.0476	0.1712	0.0631	0.9317
	(b) $\underset{\sim}{T}_{60+}^{(1970)}$			
Northeast	0.8241	0.0171	0.0358	0.0114
North Central	0.0390	0.7274	0.0751	0.0286
South	0.0861	0.0796	0.8253	0.0258
West	0.0508	0.1758	0.0638	0.9341
	(c) $\hat{\underset{\sim}{S}}_{60+}$ (estimated)			
Northeast	0.9827	-0.0009	-0.0003	-0.0011
North Central	-0.0009	0.9860	0.0000	-0.0014
South	0.0148	0.0090	1.0003	0.0002
West	0.0034	0.0057	-0.0001	1.0024
	(d) $\underset{\sim}{S}_{60+}$ (actual)			
Northeast	0.9702	0.0021	0.0041	0.0024
North Central	0.0035	0.9701	0.0068	0.0096
South	0.0213	0.0167	0.9848	0.0098
West	0.0050	0.0110	0.0043	0.9783

Source: Calculated using data reported in US Bureau of the Census (1973).

For the young age groups, the off-diagonal elements of $\underset{\sim}{T}_x^{(1970)}$ are generally higher than the corresponding elements of $\underset{\sim}{T}_x^{(1965)}$ and thus $\hat{\underset{\sim}{S}}_x$ has positive diagonal elements. In the case of the older age groups, the relationship between the off-diagonal elements of $\underset{\sim}{T}_x^{(1965)}$ and $\underset{\sim}{T}_x^{(1970)}$ tends to reverse itself under a double influence (there remain fewer people likely to move out of their region of birth, while returns to the region of birth, although diminishing rapidly with age, persist) which accounts for the negative off-diagonal elements of $\hat{\underset{\sim}{S}}_x$.

The interest of the data set used in this paper is that it can also be aggregated to obtain the age-specific matrices $\underset{\sim}{M}_x$ of the actual interregional streams observed between 1965 and 1970. By dividing each element of this matrix by the sum of the elements in the relevant column, we immediately obtain the observed transition probability matrices $\underset{\sim}{S}_x$, of which we show those relating to age groups 20-24 and 60+ in exhibits (d) of Tables 1 and 2 respectively. The comparison of those matrices with the corresponding estimated matrices $\hat{\underset{\sim}{S}}_x$ [exhibits (c) in Tables 1 and 2] allows for a rough assessment of the Rogers-von Rabenau PRPB method.

Such a comparison reveals that:

a) Off-diagonal elements are consistently underestimated by a considerable amount: in the case of the 20-24 age group, the estimated off-diagonal elements are about two to ten times smaller than the observed ones.
b) The smaller the value of the observed off-diagonal element, the higher the relative overestimation.

A slight modification of the method perhaps could provide more acceptable estimates. In effect, we could easily avoid obtaining estimates of the survivorship proportions which are not situated between 0 and 1. Observing that the estimated matrix of transition probabilities for the whole population is likely to be stochastic in virtually all situations, we could estimate a consistent set of age-specific stochastic transition probability matrices using the entropy-maximizing method described in Willekens (1977). Unfortunately, although this addition to the Rogers-von Rabenau method yields more credible estimates of the transition probability matrices--especially in the case of the older age groups--it does not improve the accuracy of such estimates by very much. The underestimation of the off-diagonal elements remains, simply because the central matrix used in the implementation of Willekens's method--i.e., the estimated probability matrix for the whole population--also presents, by construction, underestimated off-diagonal elements.

An alternative perspective on the performance of the PRPB method can be obtained by comparing the predicted values of the total migration flows,implied by the survivorship proportions as estimated above,with their corresponding actual values. It turns out (Table 3) that, for the whole of the age groups, the number of predicted migrations (1,495,810) represents only 37.0 percent of the number of migrations actually observed (4,037,11). In fact, the ratio of predicted to actual migrations exhibits sharp variations with age. It takes on its highest value in the case of the 15 to 19 age group (61.0 percent) and its lowest value (13.9 percent) in the case of the 30 to 39 age group. Note that this ratio increases with age beyond age 40 in spite of the increasing number of obvious inaccuracies.

Table 3. Total migration flows in the US four-region system, 1965-70, females: actual versus predicted.

Age Group	Actual	Predicted	Predicted / Actual
0-4	511263	249916	0.489
5-9	408415	122881	0.301
10-14	380591	161955	0.426
15-19	694541	423527	0.610
20-24	575577	161628	0.281
25-29	343713	53458	0.156
30-39	440664	61202	0.139
40-49	269235	46982	0.175
50-59	196822	59721	0.303
60+	216292	69367	0.321
All Age Groups	4037113	1495810	0.370

The conclusion here is that the Rogers-von Rabenau PRPB method is relatively inaccurate even in the most favorable case, i.e., in the absence of any data reliability problem. We are thus left with the conclusion that the principal assumption underlying the Rogers-von Rabenau method; namely, the independence of the survivorship proportions from the place of birth of the individuals concerned, is too crude an assumption to yield estimated interregional streams in agreement with the actual ones. Evidence of this is provided in Table 4 which shows the actual transition probability matrices--cross-classified by place of birth--in our US four-region system for females aged 20 to 24 in 1965. What the figures shown in this table mainly suggest is that a woman living outside her region of birth has a very high probability of returning to her region of birth within a five-year span. In the case of the 20-24 age group, the smallest probability of returning to the region of birth which was observed concerns South-born females living in the Northeast: it is as high as 0.1342. Clearly, the main assumption underlying the Regions-von Rabenau PRPB method is invalid in the real world.

Furthermore, the comparison of the estimated transition probability matrix $\underset{\sim}{S}_{20}$ [see exhibit (d) in Table 1] with the actual place-of-birth-specific transition probability matrices $\underset{\sim}{S}_{20}^{r}$ (r = 1,2,3,4) shown in Table 4 suggests that the off-diagonal elements of $\hat{\underset{\sim}{S}}_{x}$ are smaller than all of the corresponding elements in $\underset{\sim}{S}_{20}^{r}$ (r = 1,...,R where R is the number of regions). Note that, if this is true, then the above presumptive evidence that the off-diagonal elements of $\hat{\underset{\sim}{S}}_{x}$ are smaller than the corresponding elements of $\underset{\sim}{S}_{x}$ automatically holds.

Now summarizing, the above numerical test leads to conjecture that the main assumption of the Rogers-von Rabenau PRPB method, i.e., the assumption that survivorship proportions are independent of the place of birth of the individuals concerned

1) leads to underestimating the off-diagonal elements of the actual transition probability matrices and

Table 4. 1965-70 transition probability matrices cross-classified by place of birth in the US four-region system for females aged 20-24 in 1965.

To	From Northeast	North Central	South	West
Born anywhere in the US ($\underset{\sim}{S}_{20}$)				
Northeast	0.9118	0.0184	0.0265	0.0207
North Central	0.0226	0.8983	0.0392	0.0423
South	0.0470	0.0439	0.9052	0.0488
West	0.0251	0.0394	0.0291	0.8882
Born in the Northeast ($\underset{\sim}{S}^1_{20}$)				
Northeast	0.9335	0.2064	0.2404	0.1549
North Central	0.0157	0.6588	0.0449	0.0378
South	0.0304	0.0708	0.6613	0.0615
West	0.0204	0.0641	0.0533	0.7457
Born in the North Central ($\underset{\sim}{S}^2_{20}$)				
Northeast	0.6029	0.0119	0.0356	0.0170
North Central	0.2226	0.9253	0.2491	0.1412
South	0.0930	0.0281	0.6440	0.0436
West	0.0815	0.0347	0.0714	0.7982
Born in the South ($\underset{\sim}{S}^3_{20}$)				
Northeast	0.8074	0.0148	0.0143	0.0160
North Central	0.0302	0.8084	0.0260	0.0352
South	0.1342	0.1453	0.9396	0.1664
West	0.0282	0.0315	0.0200	0.7824
Born in the West ($\underset{\sim}{S}^4_{20}$)				
Northeast	0.5794	0.0245	0.0275	0.0092
North Central	0.0539	0.6717	0.0495	0.0173
South	0.0728	0.0627	0.6539	0.0221
West	0.2939	0.2414	0.2691	0.9514

Source: Ledent (1980).

2) yields estimates of these off-diagonal elements which are less than the smallest of the corresponding elements in the actual place-of-birth specific transition probability matrices.

Unfortunately, we cannot demonstrate the validity of this conjecture in the general case, i.e., for any value of R. Only in the two-region case are we able to justify it, as we will show later in the course of the next section.

3. THE TWO-REGION CASE: A RIGOROUS ASSESSMENT

In this section, we present a theoretical assessment of the principal assumption of the Rogers-von Rabenau method in the case of a two-region system.

First, let us take up the following question: when does the estimated transition probability matrix $\hat{\tilde{S}}_x$ cease to be stochastic? That is, what values must $\tilde{T}_x^{(t+T)}$ and $\tilde{T}_x^{(t)}$ take on so that the elements of $\hat{\tilde{S}}_x$ are not located between zero and one?

Let

$$\tilde{T}_x^{(t)} = \begin{pmatrix} 1-a & b \\ a & 1-b \end{pmatrix} \tag{7}$$

and

$$\tilde{T}_x^{(t+T)} = \begin{pmatrix} 1-c & d \\ c & 1-d \end{pmatrix} \tag{8}$$

Thus, by application of (6), we obtain a matrix $\hat{\tilde{S}}_x$ whose column sums are equal to one, i.e.,

$$\hat{\tilde{S}}_x = \begin{pmatrix} 1-u & v \\ u & 1-v \end{pmatrix}, \quad , \tag{9}$$

where

$$u = \frac{c-a + ad-bc}{1 - (a+b)} \tag{10}$$

and

$$v = \frac{d-b - ad+bc}{1 - (a+b)} \tag{11}$$

From these two formulas, it is readily established that u and v are not necessarily located between zero and one for all possible values of a, b, c, and d (themselves located between zero and one). For example, the element u of $\hat{\underset{\sim}{S}}_x$ is

(i) negative if

$$a + b < 1 \quad \text{and} \quad ad - bc < a - c \tag{12}$$

or

$$a + b > 1 \quad \text{and} \quad ad - bc > a - c \tag{13}$$

(ii) greater than one if

$$ad - bc < 1 - (b + c) \tag{14}$$

In practice, the off-diagonal elements of $\underset{\sim}{T}_x^{(t)}$ and $\underset{\sim}{T}_x^{(t+n)}$ are small so that (14) holds. Thus u is generally less than one but can be easily negative. Since a + b is normally less than 1, it follows that u is negative if the value of the c-element is less than the value of the a-element plus a positive or negative quantity of small magnitude, bc - ad. Thus, in the two-region case, an off-diagonal element of $\hat{\underset{\sim}{S}}_x$ is negative if the corresponding element of $\underset{\sim}{T}_x^{(t+n)}$ is less than the corresponding element of $\underset{\sim}{T}_x^{(t)}$ up to a small additive term.

Moreover, noting that for small values of a, b, c, and d, we have from (10)

$$u \simeq c-a + a[c+d - (a+b)] \quad , \tag{15}$$

we conclude that an off-diagonal element of $\hat{\underset{\sim}{S}}_x$ is approximately equal to the difference of the corresponding elements in $\underset{\sim}{T}_x^{(t+T)}$ and $\underset{\sim}{T}_x^{(t)}$ because the quantity a[c+d - (a+b)] is of the second order.

Let us now turn to the justification, still in the two-region case, of the conjecture made in Section Two.

For this purpose, let us introduce the following notation. Let the estimated transition probability matrix be

$$\hat{\underset{\sim}{S}}_x = \begin{pmatrix} 1 - \hat{M}_x & \hat{N}_x \\ \hat{M}_x & 1 - \hat{N}_x \end{pmatrix} \tag{16}$$

($\hat{M}_x$ and $\hat{N}_x$ are substituted for u and v) and the actual place-of-birth-specific transition probability matrices be

$$\underset{\sim}{S}_x^1 = \begin{pmatrix} 1 - m_x & n'_x \\ m_x & 1 - n'_x \end{pmatrix} \tag{17}$$

and

$$\underset{\sim}{S}_x^2 = \begin{pmatrix} 1 - m'_x & n_x \\ m'_x & 1 - n_x \end{pmatrix} \tag{18}$$

where the coefficients without a prime sign relate to migration out of the region of birth and those with a prime sign relate to migration toward the region of birth. Of course, in view of the evidence shown in Table 3, we have that

$$m_x < m'_x \quad \text{and} \quad n_x < n'_x \tag{19}$$

Analytically, the principal assumption of the Rogers-von Rabenau PRPB method, i.e., the independence of the transition probability matrices vis-a-vis place of birth, implies the following relationships in vector format:

$$\hat{\underset{\sim}{S}}_x \{{}^1K_x^{(t)}\} = \underset{\sim}{S}_x^1 \{{}^1K_x^{(t)}\} \tag{20}$$

and

$$\hat{\underset{\sim}{S}}_x \{{}^2K_x^{(t)}\} = \underset{\sim}{S}_x^2 \{{}^2K_x^{(t)}\} \tag{21}$$

where $\{{}^iK_x^{(t)}\}$ is the i-th column of $\underset{\sim}{K}_x^{(t)}$.

The matrices $\underset{\sim}{S}_x^1$ and $\underset{\sim}{S}_x^2$ being stochastic matrices, both (20) and (21) consist of two scalar equations which are not independent of each other. Thus we are left with a system of two scalar equations--one being a scalar equation of (20), the other being a scalar equation of (21)--in two variables $\hat{M}_x$ and $\hat{N}_x$, which allows one to derive an expression of $\hat{M}_x$ and $\hat{N}$ in terms of the elements of $\underset{\sim}{S}_x^1$, $\underset{\sim}{S}_x^2$ and $\underset{\sim}{K}_x^{(t)}$.

Let us, for example, choose the first scalar equation in (20) and the second scalar equation in (21). We thus have, after dropping the subscript x and the superscript (t), the following system of equations

$$(1 - \hat{M})\,{}^{11}K + \hat{N}\,{}^{12}K = (1 - m)\,{}^{11}K + n'\,{}^{12}K \tag{22}$$

and

$$\hat{M}\,{}^{21}K + (1 - \hat{N})\,{}^{22}K = m'\,{}^{21}K + (1 - n)\,{}^{22}K \tag{23}$$

which can be rewritten as

$$-\hat{M}\ {}^{11}K + \hat{N}\ {}^{12}K = -m\ {}^{11}K + n'\,{}^{12}K \tag{24}$$

and

$$\hat{M}\ {}^{21}K - \hat{N}\ {}^{22}K = m'\,{}^{21}K - n\ {}^{22}K \tag{25}$$

Multiplying equation (24) by ${}^{22}K$ and adding to the result equation (25) multiplied by ${}^{12}K$ leads to the following value of $\hat{M}$

$$\hat{M} = \frac{m\ {}^{11}K\ {}^{22}K - m'\,{}^{12}K\ {}^{21}K + (n - n')\,{}^{12}K\ {}^{22}K}{{}^{11}K\ {}^{22}K - {}^{12}K\ {}^{21}K}\ , \tag{26}$$

an expression which shows that $\hat{M}$ depends not only on m and m′ but also on n and n′. Similarly, we have

$$\hat{N} = \frac{n\ {}^{11}K\ {}^{22}K - n'\,{}^{12}K\ {}^{21}K + (m - m')\,{}^{21}K\ {}^{11}K}{{}^{11}K\ {}^{22}K - {}^{12}K\ {}^{21}K} \tag{27}$$

In Section Two, we conjectured that the off-diagonal elements of $\hat{S}$ are less than

a) their observed counterparts
b) the smaller of the corresponding elements in $\underset{\sim}{S}^1$ and $\underset{\sim}{S}^2$, i.e., $\hat{M} < m$ and $\hat{N} < n$.

Letting

$$M = \frac{m\ {}^{11}K + m'\,{}^{21}K}{{}^{11}K + {}^{21}K} \tag{28}$$

be the observed counterpart of $\hat{M}$, we can readily establish from (26) and (28) that

$$M - \hat{M} = \frac{(m' - m)\,{}^{11}K\,{}^{21}K({}^{12}K + {}^{22}K) + (n' - n)\,{}^{12}K\,{}^{22}K({}^{11}K + {}^{21}K)}{({}^{11}K + {}^{21}K)({}^{11}K\,{}^{22}K - {}^{12}K\,{}^{21}K)} \qquad (29)$$

Since, in virtually all cases, people born at the same time in a given region are more likely to be found, at any subsequent age, in their region of birth than outside, we have

$${}^{11}K > {}^{12}K \qquad (30)$$

and

$${}^{22}K > {}^{21}K \qquad (31)$$

so that the denominator of the right-hand side of (29) is positive. Then, owing to the inequalities contained in (19), the difference $M - \hat{M}$ is positive.

In addition, subtracting m on both sides of (26) leads, after several manipulations, to

$$m - \hat{M} = {}^{12}K\,\frac{(m' - m)\,{}^{21}K + (n' - n)\,{}^{22}K}{{}^{11}K\,{}^{22}K - {}^{12}K\,{}^{21}K} \quad , \qquad (32)$$

an equation which shows that the difference $m - \hat{M}$ is positive. Thus, we have

$$\hat{M} < m < M \quad , \qquad (33)$$

which justifies our conjecture of Section Two in the two-region case.

In addition, (32) suggests that the independence of inter-regional migration patterns vis-a-vis the place of birth is not only a sufficient but also a necessary condition for a perfect estimation of the migration propensities.

To illustrate the above discussion, the US population system used in our illustration was aggregated into a two-region system consisting of the Southern Region and the Rest of the US. The values of $\hat{M}$, m and M, relating to migration out of the Southern Region, are reported for all concerned age groups in Figure 1.

Clearly, the reliability of the Rogers-von Rabenau PRPB method depends on how much $\hat{M}$ underestimates M. Figures shown in Table 3 as well as in Figure 1 indicate a considerable underestimation which we would like to assess in precise terms. For this purpose, let $\alpha(\beta)$ be the part of the population present in region 1(2) at time t which was born in that same region, i.e.,

$$\begin{aligned} {}^{11}K &= \alpha \cdot {}^{1}K \quad ; \quad {}^{21}K = (1-\alpha)\cdot {}^{1}K \\ {}^{12}K &= (1-\beta)\cdot {}^{2}K \quad ; \quad {}^{22}K = \beta \cdot {}^{2}K \end{aligned} \tag{34}$$

Also, instead of studying directly the difference $M - \hat{M}$, let us break it down into two parts $M - m$ and $m - \hat{M}$ which, owing to (33), are both positive. Let us first examine the difference $M - m$ which we can easily express as

$$M - m = \frac{(m' - m)\,{}^{21}K}{{}^{11}K + {}^{21}K} \tag{35}$$

or, after substitution of (34),

$$M - m = (1-\alpha)(m' - m) \tag{36}$$

As for the difference $m - \hat{M}$, substituting (34) into (32) yields

$$m - \hat{M} = (1-\beta)\frac{(m' - m)(1-\alpha) + (n' - n)\beta\, x}{\alpha + \beta - 1} \tag{37}$$

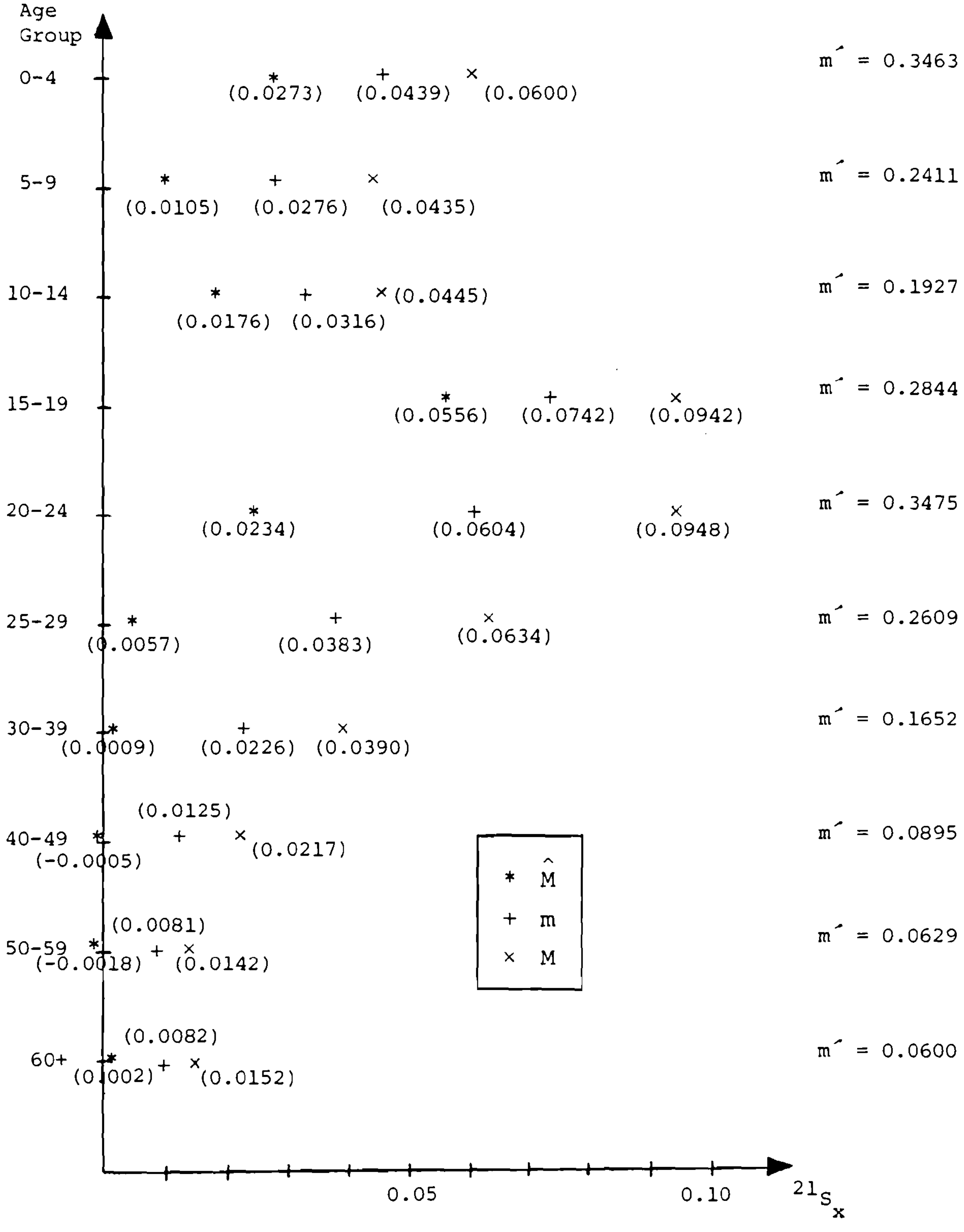

Figure 1. Migration from the South to the rest of the US, 1965-70, females: values of $\hat{M}$, m, M and m´ contrasted.

where

$$x = \frac{.^{2}K}{.^{1}K} \tag{38}$$

From (36) and (37), it is clear that the discrepancy between the predicted and actual values of the migration propensities out of one region is larger

(a) the larger the difference in the migration propensities of the natives and non-natives out of that region
(b) the larger the difference in the migration propensities out of the other region by natives and non-natives of that region
(c) the higher the ratio of the population in the other region to that of the region considered.

In fact, the difference $M - \hat{M}$ consists of two parts, one of which (the discrepancy between M and m) depends only on characteristics of the region of outmigration [see formula (36)] while the other (the discrepancy between m and $\hat{M}$) is primarily affected by characteristics of the other region [see formula (37)]. As suggested by the results shown in Figure 1 both these parts which appear to have similar magnitudes can be quite large.

Let γ denote the larger of the two values α and β and let Z denote the smaller of the two differences $m' - m$ and $n' - n$. From (36), it follows that

$$M - m > (1 - \gamma)Z \tag{39}$$

while, from (37), it can be established (see Appendix) that

$$m - \hat{M} > (1 - \gamma)Z\ x \tag{40}$$

Then we have that

$$M - \hat{M} > (1 - \gamma)Z(1 + x) \tag{41}$$

In general, the discrepancy between the outmigration propensities of natives and non-natives is quite large: for example, in the case of the Southern Region, it varies from 0.05 to 0.30 according to age groups (see Figure 1). Consequently, the lower bound of $M - \hat{M}$ which represents a few percentage units of such a discrepancy is usually high: in the case of the Southern Region (see Figure 1) it varies from 0.015 (in the 60 years and over age group) to 0.071 (in the 20 to 24 age group).

CONCLUSION

In this note, we have demonstrated the relative inaccuracy of the method suggested by Rogers and von Rabenau (1971) to infer interregional migration streams from place-of-residence-by-place-of-birth (PRPB) data. The main reason accounting for this unfortunate result was shown to be the assumption that interregional migration patterns are independent of the birthplace of the individuals concerned.

Clearly, the message of this paper is that the use of a double subscript, relating to the birthplace, to obtain matrix generalizations of some mathematical models of population growth and distribution is not problem-free. The price to be paid for the mathematical convenience allowed by this double subscripting is the introduction, in those models, of a crude assumption, namely the independence of migration patterns vis-a-vis the birthplace, which leads to numerical applications producing inaccurate results.

In the case of the implementation of the PRPB method, the results obtained include some obvious inaccuracies, such as the presence of negative gross migration flows. But, in other cases, such as the construction of a multiregional life table, there are no obvious inaccuracies although the results can be shown to be largely incorrect. Ledent (1980) shows that the current methodology used for the construction of such life tables leads to a considerable overestimation of migration propensities.

How then can one infer recent migration streams (over a short fixed period of time) from lifetime migration stream data? To our knowledge, there does not exist any alternative to the Rogers-von Rabenau PRPB method. This is rather unfortunate in view of the greater interest accorded to the analysis of recent migration patterns by demographers and planners, especially in developing countries which, in most cases, are still experiencing an accelerated transfer of population from rural to urban areas. Thus, censuses should in all circumstances allow one to measure current migration streams, i.e., over a short period before the census year.

Second, the above evidence concerning the impact of the place of birth on migration decisions indicates that an analysis of interregional migration patterns is more meaningful if the data available on current interregional migration streams is cross-classified by place of birth.

It follows from the above two remarks that, ideally, data on interregional migration streams should be produced by asking in censuses direct questions concerning the place of residence at a fixed prior data as well as the place of birth. But, in the case that some constraints would allow for only one of the two questions, our first remark above suggests that the former question (place of residence at a fixed prior date) be asked in preference to the latter one (birthplace).*

*Note that, in contrast to this, it is the question on birthplace which is most commonly used: it is given priority in the United States recommendations (United Nations 1970). The fact is that people can easily name their birthplace whereas they very often have difficulty recalling where they were living at some arbitrary date in the past.

REFERENCES

Eldridge, H., and Y. Kim (1968) *The Estimates of Internal Migration from Birth Residence Statistics: A Study of Data for the United States, 1950 and 1960.* Philadelphia, Pennsylvania: Population Studies Center, University of Pennsylvania.

Harary, F., B. Lipstein, and E. Styan (1970) A Matrix Approach to Nonstationary Chains. *Operations Research 18:1168-81.*

Ledent, J. (1978) *Some Methodological and Empirical Considerations in the Construction of Increment-Decrement Life Tables.* RM-78-25. Laxenburg, Austria: International Institute for Applied Systems Analysis.

Ledent, J. (1980) *Constructing Multiregional Life Tables Using Place-of-Birth-Specific Migration Data.* WP-80-96. Laxenburg, Austria: International Institute for Applied Systems Analysis.

Lee, E., et al. (1957) *Population Redistribution of Economic Growth, United States, 1970-1950, Volume 1-Methodological Considerations and Reference Tables.* Philadelphia, Pennsylvania: American Philosophical Society.

Rogers, A. (1968) *Matrix Analysis of Interregional Population Growth and Distribution.* Berkeley, California: University of California Press.

Rogers, A. (1975) *Introduction to Multiregional Mathematical Demography.* New York, New York: John Wiley.

Rogers, A. (1979) Migration patterns and population redistribution. *Regional Science and Urban Economics 9:275-310.*

Rogers, A., and B. von Rabenau (1971) Estimation of interregional migration streams from place-of-birth-by-residence data. *Demography 8:185-194.*

United Nations (1970) *Methods of Measuring Internal Migration.* Population Studies No. 47, Department of Economic and Social Affairs, United Nations, New York, New York.

U.S. Bureau of the Census (1973) *Census of Population: 1970: Lifetime and Recent Migration.* Final Report PC(2)-2D, U.S. Government and Printing Office, Washington, D.C.

Willekens, F. (1977) *The Recovery of Detailed Migration Patterns from Aggregate Data: An Entropy Maximizing Approach.* RM-77-58. Laxenburg, Austria: International Institute for Applied Systems Analysis.

APPENDIX: PROOF OF INEQUATION (40)

Let Z denote the smaller of the two differences $m' - m$ and $n' - n$. Then, from (37), we have

$$m - \hat{M} > (1 - \beta)\ Z\ \frac{(1 - \alpha) + \beta\ x}{\alpha + \beta - 1} \tag{A1}$$

The problem here is one of finding a lower bound for

$$y = \frac{1 - \alpha + \beta\ x}{\alpha + \beta - 1} \tag{A2}$$

Let us suppose that α is given and let us study the variations of y in terms of β. Differentiating (A2) with respect to β yields

$$\frac{dy}{d\beta} = - \frac{(1 - \alpha)(1 + x)}{(\alpha + \beta - 1)^2} \quad , \tag{A3}$$

which shows that z is a decreasing function of β. Thus the minimal value of z, z_{min}, is obtained for $\beta = 1$, i.e.,

$$z_{min} = \frac{1 - \alpha + x}{\alpha} \quad \text{(A4)}$$

It can be established that the minimal value of the right-hand side of (A4) is greater than x for all values of β comprised between $1 - \alpha$ and 1 [the fact that $\beta > 1 - \alpha$ reflects the aforementioned observation that the denominator of the right-hand side of (29) is positive]. Thus, we have

$$m - \hat{M} > (1 - \beta) Z x \quad (40)$$

SELECTED PAPERS ON MIGRATION AND SETTLEMENT AT IIASA

Andrei Rogers, editor, *Migration and Settlement: Selected Essays.* RR-78-6. Reprinted from a special issue of *Environment and Planning A* 10(5):469-617.

Andrei Rogers and Frans Willekens, *Migration and Settlement: Measurement and Analysis.* RR-78-13.

Frans Willekens and Andrei Rogers, *Spatial Population Analysis: Methods and Computer Programs.* RR-78-18.

Andrei Rogers, *Migration Patterns and Population Redistribution.* RR-80-7. Reprinted from *Regional Science and Urban Economics* 9(1970):275-310.

Andrei Rogers, editor, *Essays in Multistate Mathematical Demography.* RR-80-10. Reprinted from a special issue of *Environment and Planning A* 12(5):485-622.