

Application of a mathematical model to South African migration data, 1975-1980

B.E. Hofmeyr

Human Sciences Research Council

INTRODUCTION

Migration studies in a number of countries have shown a common age-dependent characteristic of the type illustrated in Figure 1, which shows the fundamental age pattern of migration with peaks occurring at infancy, during young adulthood, and at retirement.

Shaw (1975:18) refers to a large number of studies which corroborate the proposition that adult persons in the age group 20-29 years have the greatest propensity to migrate. It also has been demonstrated that in a number of the more industrialized countries an upturn in migration appears at about age 65, which has been referred to by some researchers as retirement migration (see Shaw 1975:19). It is evident from Figure 2 that migration data for whites in South Africa show a similar pattern. However, attempts to summarize and codify such regularities have received little attention in demographic studies. Up to the present, model schedules, i.e. the mathematical modelling of migration patterns, have not been applied to South African migration data.

With the increased emphasis on applied and applicable research, purely qualitative descriptions could become inadequate and it is obvious that the ability to quantify patterns, especially in the form of mathematical expressions, is of great significance.

Rogers and Castro (1981) have developed a model migration schedule that was applied to observed migration data of several countries and the aim of the work described in this article was to see whether this model schedule could be applied to white South African migration patterns. In the following section it is shown that it could be applied successfully by obtaining best fits to the relevant parameters.

REGULAR FEATURES OBSERVED IN WHITE SOUTH AFRICAN MIGRATION DATA

The country is divided into 44 National Planning Regions (relatively large areal units normally comprising a number of magisterial districts) and for the purpose of this study a migrant is defined as a person who has left a particular National Planning Region (NPR) during the 5-year period 1975-1980 to reside in a different NPR. For the purpose of this study seven self-governing black areas (or homelands) within the RSA were also regarded as NPRs. The data were obtained from the 5% sample of the 1980 population census in which the question about place of usual residence in 1975 was posed. This excluded the category under 5 years of age (in 1980) from the survey.¹

Migration movement within one NPR is not registered, but according to empirical evidence, differences in the size of the geographic units would only influence

FIGURE 1: THE MODEL MIGRATION SCHEDULE

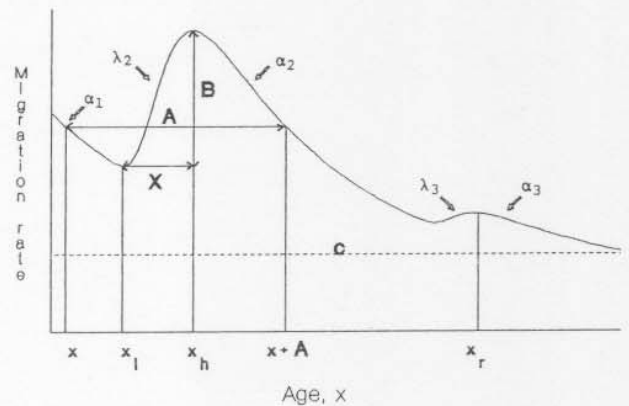
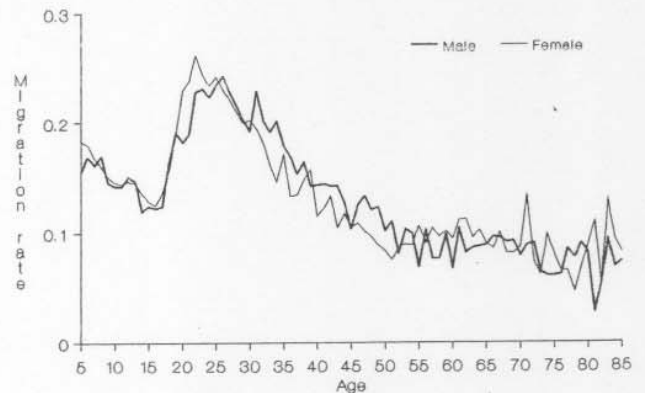


FIGURE 2: OBSERVED QUINQUENNIAL MIGRATION RATES BY SINGLE YEARS OF AGE: SOUTH AFRICA: 1975-1980



the level and not the shape or profile of an age-specific schedule of migration data (Rogers, Raquillet and Castro 1978:476). Similarly, the inclusion of immigrants to the country during 1975-1980 boosts the level without materially affecting the general profile of the curve (compare Figures 3 and 4).

The 5-year age-specific migration rates are defined as the ratio of the number of migrants, aged x years, to the population aged x years exposed to the risk of becoming migrants during the specified time interval. These rates are probability rates which give a measure of the proportion of the population moving at least once during the given migration interval (United Nations 1970:40). The 5-year migration rates were adjusted to obtain effective annual migration rates ($M(x)$), which

will tend to be a lower estimate, since the count of migrants over a longer time interval usually will understate the number of annual interarea migrants due to multiple migrations, non-surviving migrants and return migrants not being recorded.

With respect to South African national migration patterns the following findings, as can be seen in Figures 3 and 4, seem to support the general pattern, as illustrated in Figure 1, with parameters labelled according to Rogers and Castro (1981):

- (a) Migration rates among young children reflect the relatively high rates of their parents, who are young adults in their late twenties. Young teens reveal a local *low point* (x_l in Figure 1) around age 15. Young adults in their early twenties generally have the highest migration rates, attaining a *high peak* (x_h) at about age 22 (females) or 25 (males). After this age the migration rate as a function of age declines monotonically until it reaches the so-called *retirement peak* (x_r) at ages around 60 (females) and 65 (males).
- (b) The age profiles of male and female migration show a distinct difference. The high peak of the female schedule is a little higher and precedes that of the male schedule by approximately 3 years (Figure 2), which appears to be in line with the difference between the average ages at marriage of the two sexes and the fact that females usually move to their husband's place of residence. The male rates are higher from the mid-twenties to the

FIGURE 3A: INTERNAL MALE MIGRANTS

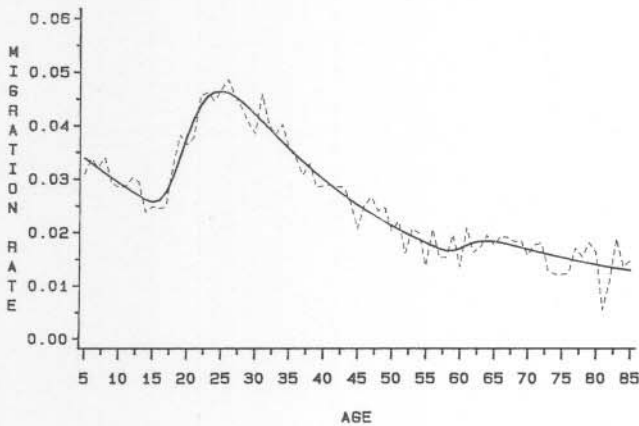


FIGURE 3B: INTERNAL FEMALE MIGRANTS

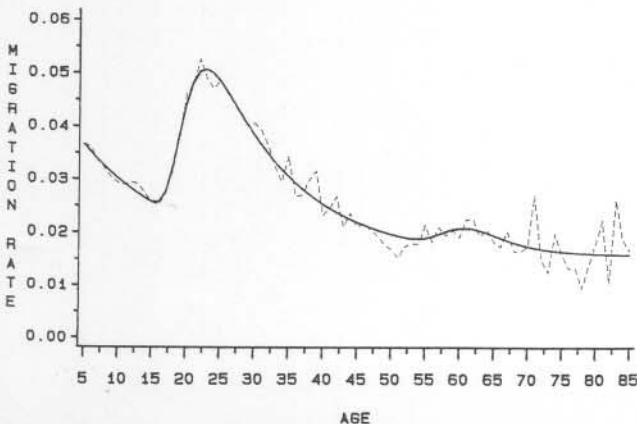


FIGURE 4A: MALE MIGRANTS (INCLUDING IMMIGRANTS)

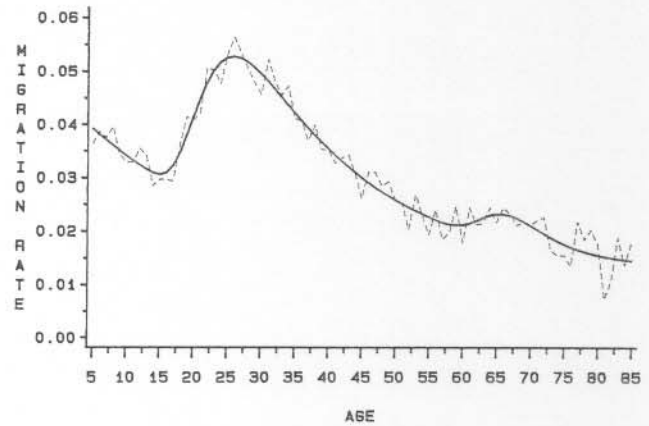
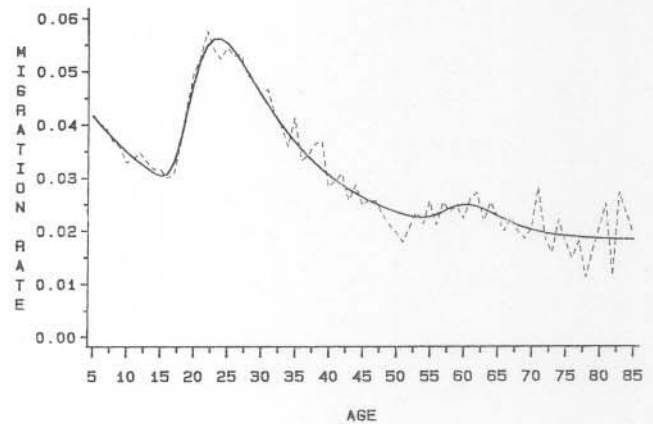


FIGURE 4B: FEMALE MIGRANTS (INCLUDING IMMIGRANTS)



mid-fifties and lower at most other ages. However, the absolute numbers of migrants coincide between the male ages of 33 and 57 if the female curve is displaced to a higher age by about 3 years, which might fit in with a dominant migration of couples. The retirement peak among females is somewhat higher and broader and starts at an earlier age than for males and is also shifted by about 3 years with respect to males. This would be consistent with both migration by couples and the fact that women usually retire 5 years earlier in South Africa. When comparing numbers, after the age of 55 consistently more women migrate than men, in line with their lower mortality and resulting widowhood.

THE MIGRATION MODEL SCHEDULE

Rogers and Castro (1981) have described these regularities in the age profile of migrants by means of a model schedule (Figure 1), which is defined by a total of 11 parameters and which, for the sake of clarity, will be summarized. The double-exponential curve which they applied, had previously been used by Coale and McNeil (1972) to describe patterns of nuptiality and fertility.

$$\hat{M}(x) = a_1 \exp(-\alpha_1 x) + a_2 \exp(-\alpha_2(x-\mu_2) - \exp(-\lambda_2(x-\mu_2))) + a_3 \exp(-\alpha_3(x-\mu_3) - \exp(-\lambda_3(x-\mu_3))) + c \dots (1)$$

where $x = 0, 1, 2, 3, \dots, z$ (z represents the open age interval of 85 years and over).

This mathematical expression is the sum of four components:

- A single negative exponential curve during the pre-occupation ages, the so-called pre-labour force component, with its exponential parameter α_1 , which determines the rate of descent.
- A left-skewed unimodal curve for the labour force ages positioned around mean age μ_1 on the age axis (the so-called labour force component), with its ascent and descent-related parameters λ_2 and α_2 .
- An approximately bell-shaped curve for the retirement ages centred at μ_3 on the age axis (the so-called post-labour force component), with its corresponding parameters λ_3 and α_3 .
- A constant curve c (the constant or age-independent component), which is required to improve the fit of the mathematical expression to the observed schedule.

Rogers and Castro (1981) distinguish between three families of age profiles:

Firstly, the full 11-parameter *basic model migration schedule*, which exhibits a definite retirement peak;

secondly, the 7-parameter *reduced form* of this basic model, from which the retirement peak is eliminated, and

thirdly, the 9-parameter model migration schedule, where the retirement ages are characterized by an *upward slope*.

The full model was found to be applicable in the case under study.

Seven of the eleven parameters, α_1 , μ_2 , α_2 , λ_2 , μ_3 , α_3 and λ_3 define the *profile* of the model schedule, while the four remaining parameters, a_1 , a_2 , a_3 , and c , refer only to its *level*.

APPLICATION OF THE MODEL SCHEDULE TO OBSERVED MIGRATION RATES

The application of this model schedule (which is non-linear in the parameters) to migration rates of whites between NPRs in South Africa is described below.

The programming procedure used to estimate the parameters of the model migration schedule is the Marquardt method (Marquardt 1963), which seeks out a parameter vector that minimizes the residual vector. This procedure performs a non-linear least-squares regression by using an iterative method to solve for least-squares estimates. It starts off with initial estimates of the parameters (which require careful selection), then keeps regressing the residuals on the partial derivatives of the model with respect to the parameters until the iterations converge (Sall, 1981: 7-1).

The index of goodness-of-fit, E , that the model schedule provides when it is applied to the observed data, is calculated as follows:

$$E = \frac{1/n \sum_x |\hat{M}(x) - M(x)|}{1/n \sum_x M(x)} 100,$$

i.e. the mean of the absolute difference between estimated and observed values, expressed as a percentage of the observed mean. In Table 1 the results of the application to the South African data are given, the indices of goodness-of-fit E being 7,8 for males and 7,3 for females, (6,9 and 6,8 respectively for males and females including immigrants). Figures 3 and 4 graphically illustrate this goodness-of-fit of the model schedule (solid line) to the observed migration rates (broken

line). By comparison, in the summary of model parameters and variables which Rogers and Castro (1981: 25-26) calculated for reduced sets (i.e. schedules without retirement peaks) of observed migration schedules for Sweden, the United Kingdom, and Japan, the median value of E was 13,1 and 12,3 for males and females respectively. This measure indicates that the fit of the model to the observed South African data is relatively good.

The obtained parameters for the observed model migration schedule, i.e. the empirical data fitted by equation (1), are presented in Table 1, of which the salient features are described next.

Besides the ages associated with the *low point*, x_1 , the *high peak* x_h , and the *retirement peak*, x_r , (see Paragraph 2 and Table 1), Rogers and Castro (1981) identify several descriptive measures which characterize the model migration schedule:

(a) The "gross migraproduction rate" (GMR) represents the sum of the age-specific migration rates over all ages during a specific time period, i.e. $GMR = \sum M(x)$. It reflects the level or intensity of migration during the given period. Since it combines the migration rates over all ages of a lifespan, it may also be viewed as representing the average migrations per lifetime of a member of a synthetic cohort of migrants, ignoring the effects of mortality on migration. The observed migration levels for males and females are similar, showing a GMR of about 2,3 migrations for internal migrants (2,7 including immigrants). Because of the similarity in the GMRs no normalization is required for the visual comparison of Figures 3 and 4 (Rogers and Castro, 1981: 13).

(b) The percentage of GMR in the age intervals 0-14 years (typically 21-22%), 15-64 years (typically 62-65%), and 65+ years (typically 14-15%), shows a comparatively high relative migration rate for the aged, similar to the United Kingdom, as observed by Rogers and Castro (1981).

(c) The mean age of migration, \bar{n} , is defined as

$$\bar{n} = \frac{\sum_x x \hat{M}(x)}{\sum_x \hat{M}(x)}$$

The values obtained were 34,9 for males (35,3 including immigrants) and 34,8 for females (35,2 including immigrants), which seem to indicate that the observed migration schedules are relatively "old", compared with the average mean value (30,7) for males and females for Sweden, the United Kingdom and Japan which Rogers and Castro (1981) computed for reduced sets. However, it should be noted that the average age of migrants at the time of migration was about 2,5 years younger than the reported age in 1980. Taking this into account, the curves might be called "early peaking" curves and this is reflected by the low values of μ_2 . Rogers and Castro term curves with the parameter μ_2 taking on a value below age 19 "early peaking" and those above 22 "late peaking" curves.

(d) The *labour force shift*, X , is defined as the difference in years between the ages at the low point and the high peak, i.e. $X = x_h - x_1$. This gives an indication of the speed at which the work force settles down. The South African pattern resembles that in Europe (larger for males than for females); by contrast X is smaller for males in Japan. On average the South African figures are similar to those in Europe, but the male-female difference of 2-3 years is larger than the typical value of about one year.

TABLE 1 PARAMETERS AND VARIABLES DEFINING THE OBSERVED MIGRATION SCHEDULE (standard deviation in brackets)

Parameters and variables	Internal migrants		Including immigrants	
	Male	Female	Male	Female
GMR	2,28	2,31	2,69	2,70
E	7,8	7,3	6,9	6,8
a1	0,036 (0,010)	0,031 (0,005)	0,033 (0,004)	0,033 (0,005)
alpha1	0,033 (0,019)	0,072 (0,019)	0,042 (0,015)	0,067 (0,018)
a2	0,041 (0,007)	0,050 (0,006)	0,052 (0,007)	0,056 (0,007)
mu2	19,75 (0,71)	19,50 (0,45)	21,07 (0,78)	19,91 (0,55)
alpha2	0,043 (0,017)	0,089 (0,013)	0,059 (0,012)	0,085 (0,012)
lamda2	0,337 (0,083)	0,436 (0,093)	0,260 (0,051)	0,357 (0,072)
a3	0,006 (0,003)	0,010 (0,021)	0,014 (0,024)	0,011 (0,032)
mu3	60,42 (1,94)	62,72 (32,5)	68,07 (25,77)	62,87 (29,98)
alpha3	0,005 (0,038)	0,269 (0,714)	0,259 (0,557)	0,297 (0,693)
lamda3	0,539 (0,673)	0,200 (0,546)	0,193 (0,391)	0,200 (0,490)
c	0,003 (0,011)	0,015 (0,001)	0,012 (0,003)	0,018 (0,001)
mean age	34,9	34,8	35,3	35,2
% (0-14)	21,19	22,48	20,91	21,98
% (15-64)	64,70	62,44	64,78	62,88
% (65+)	14,11	15,08	14,31	15,14
delta1c	11,30	1,99	2,73	1,86
delta12	0,888	0,611	0,642	0,591
delta32	0,137	0,190	0,261	0,201
beta12	0,754	0,814	0,714	0,795
sigma2	7,75	4,90	4,43	4,21
sigma3	114	0,742	0,745	0,673
x(low)	15	16	15	15
x(high)	25,86	23,14	26,80	23,94
x(retirement)	64	61	65	60
X(shift)	10	7	11	9
A	30,69	25,31	31,29	26,38
B	0,021	0,025	0,017	0,026

(e) The *parental shift*, A, is given by the sum of the age differences between adults and children representing the same migration rates on the descending slopes of the curve. Due to the lack of statistics on the children under 5 years in the present study, the parental shift is defined as:

$$A = 1/9 \sum_{x=5}^{13} A_x$$

The values obtained thus correspond in magnitude and in the pattern (larger for males) with typical values in other countries, but the male-female differential is larger than in all above-mentioned countries except Japan (Rogers, 1981). Stoto's (1977) suggestion that the parental shift is linked to the mean age of child-bearing was confirmed, the mean child-bearing age of white females (including immigrants) during the period 1975-80 being 26,8 years.

(f) The *jump*, B, is the increase in the migration rate of individuals aged x_2 over those aged x_1 , i.e. $B = \hat{M}(x_2) - \hat{M}(x_1)$. The values given in Table 1 correspond in magnitude and in the pattern (larger for females) with typical values in Western Europe (Rogers 1981).

(g) The ratios of some of the fundamental parameters, as defined by Rogers and Castro, obtained in this study, are presented in Table 1. They are summarized as follows:

- (i) $\delta_{1c} = a_1/c$ measures the level of the pre-labour force component in terms of the constant component. The relatively large spread between the observed values may not be particularly meaningful, since migrants under 5 years of age could not be accurately assessed.
- (ii) The value $\delta_{12} = a_1/a_2$ reflects the relative weight factors of the pre-labour force compo-

nent and the labour force component of the curve. Low values define a labour-dominant curve and high values point to a family-dominant curve. Rogers and Castro term this ratio the "index of child dependency", which measures the pace at which children migrate with their parents. The relatively high ratios obtained show a pronounced child dependence of the curves, but the same caution expressed under (i) remains.

- (iii) $\delta_{32} = a_3/a_2$ measures the relative weight of the post-labour force component in terms of the labour force component. The relatively large non-zero values obtained in this study indicate the presence of a pronounced retirement component, which is very small or absent in many countries.
- (iv) $\beta_{12} = \alpha_1/\alpha_2$ is termed the index of parental-shift regularity, which is close to unity if migration rates of children mirror those of their parents. The obtained indices are in all cases near the lower borderline of the region 0,8-1,2, which Rogers and Castro term regular.
- (v) The index of labour asymmetry, $\delta_2 = \lambda_2/\alpha_2$, refers to the slope of the labour component of the curve. The high values obtained, reflect the asymmetric age profile of the labour force component, but all the curves fall within the average range (1-8) observed by Rogers and Castro (1981). The labour asymmetry is more pronounced in the curve representing the internal migrants.
- (vi) The index of retirement asymmetry, $\sigma_3 = \lambda_3/\alpha_3$, is defined analogous to σ_2 . This feature is much more pronounced in the curve of the internal male migrants than for any other category of migrants included in this analysis.

Compared with the values obtained by Rogers and Castro for the countries which they analysed, a feature of all four South African curves are the relatively low values of α_1 and α_2 , which reflect the relatively moderate descending slopes of the curves.

The female values for a_2 , α_2 and λ_2 are larger than those for males. The reverse is true for μ_2 , which reflects the tendency for girls to leave home at an earlier age than boys. This corresponds with the general trend observed in a number of countries (excepting Japan) by Rogers and Castro (1981).

CONCLUSION

South African migration patterns of whites for the period 1975-1980, as obtained from census data, correspond in general with those observed in most other developed countries and the mathematical model of Rogers and Castro could thus be applied. Quantitative values have been found for the various parameters defined by Rogers and Castro. In Note 2 an example is given of how the obtained parameters are used without further information.

Once the regularities in observed migration schedules have been mathematically established, it is visualized that limited data obtained through sample surveys can be used to obtain the appropriate migration schedule, which will provide quantitative descriptions on a probability basis, i.e. it should be possible to predict the volume of migration with respect to specific categories of the population (e.g. race, age, sex) for a given period even with such limited available data. There are important application possibilities, for example for planning purposes. The computational and analytical advantages are self-evident.

According to Rogers and Castro (1981: 45) "regularities in age profiles lead naturally to the development of hypothetical model migration schedules that might be suitable for studies of populations with inadequate or defective data".³ Rogers and Castro (1981: 45-47) point out that if observed migration data are available "the model migration schedule may be used to *graduate* observed data, thereby smoothing out irregularities and ascribing to the data summary measures that can be used for comparative analysis. It may be used to *interpolate* to single years of age, observed migration schedules that are reported for wider age intervals. Assessments of the *reliability* of empirical migration data and indications of appropriate strategies for their *correction* are aided by the availability of standard families of migration schedules. Finally, such schedules also may be used to help resolve problems caused by *missing* data."

This is the first application of a scientific mathematical model to South African migration data.

This study describes purely the total national migration (including and excluding immigrants).

other geographical units and interesting comparative studies are envisaged for other population groups in order to obtain a comprehensive model for migration in South Africa.

NOTES

1 Using the place of birth for the latter group (instead of place of residence in 1975) yielded anomalously high values for migration possibly due to a certain percentage of movement for the purpose of confinement and the resulting apparent migration. This has been proved by also considering this migration criterion for older age groups, e.g. 5-10 years, which also revealed approximately double the rate obtained by the above-mentioned method. Attempts to introduce corrections to obtain annual rates for the group under 5 years did not yield sufficiently consistent results. The children under 5 years were therefore not modelled, except by extrapolation, for the purpose of calculating averages where they have only a small weight.

2 To calculate the estimated annual migration rates replace the various parameters with the values in Table 1, for example for age 20 (internal male migrants):

$$\begin{aligned} \hat{M}(20) &= 0,036 * \exp(-0,033 * 20) \\ &+ 0,041 * \exp(-0,043 * (20 - 19,75) - \exp(-0,337 * (20 - 19,75))) \\ &+ 0,006 * \exp(-0,005 * (20 - 60,42) - \exp(-0,539 * (20 - 60,42))) \\ &+ 0,003 \\ &= 0,038. \end{aligned}$$

However, these values are based on data observed during 1975-1980 and changing socioeconomic conditions will make it advisable to obtain limited new observations to adjust the amplitude accordingly, while the slope is kept unchanged. This is done if the ratios between the amplitude parameters ($a_1:a_2:a_3:c$) are retained while adapting their absolute values to fit the observed data.

3 The application of the model is illustrated in Rogers and Castro 1981: 34-45.

REFERENCES

- Coale, A.J., McNeil, D.R. 1972. *The distribution by age of the frequency of first marriage in a female cohort*. Journal of the American Statistical Association 67: 743-749.
- Marquardt, D.W. 1963. *An algorithm for least-squares estimation of nonlinear parameters*. Journal of the Society for Industrial and Applied Mathematics, 11, 431-441.
- Rogers, A., Castro, L.J. 1981. *Model migration schedules*. RR-81-30. Laxenburg, Austria: International Institute for Applied Systems Analysis.
- Rogers, A., Raquillet, R., Castro, L.J. 1978. *Model migration schedules and their applications*. Environment and Planning A 10: 475-502.
- Sall, J. 1981. *SAS Regression applications*. SAS Institute: Technical Report A-102.
- Shaw, R.P. 1975. *Migration theory and fact: a review and bibliography of current literature*. Philadelphia: Regional Science Research Institute.
- Stoto, M.A. 1977. *On the relationship of childhood to labor force migration rates*. RM-77-55, International Institute for Applied Systems Analysis. Laxenburg, Austria.