RM-77-55

ON THE RELATIONSHIP OF CHILDHOOD TO LABOR FORCE MIGRATION RATES

Michael A. Stoto

November 1977

Preface

Interest in human settlement systems and policies has been a critical part of urban-related work at IIASA since its inception. Recently this interest has given rise to a concentrated research effort focusing on migration dynamics and settlement patterns. Four sub-tasks form the core of this research effort:

I. the study of spatial population dynamics;

II. the definition and elaboration of a new research area called demometrics and its application to migration analysis and spatial population forecasting;

III. the analysis and design of migration and settlement policy;

IV. a comparative study of national migration and settlement patterns and policies.

This paper, the twelfth in the spatial population dynamics series, investigates some features of the migration schedule. Using stable population theory, an attempt is made to model the observation that young children move with their parents.

Michael Stoto is a graduate student at Harvard University. He spent three months at IIASA as part of a Summer Student Program supported in part by the Ford Foundation.

Related papers and other publications of the Migration and Settlement Study are listed in the back of this report.

Andrei Rogers
Chairman
Human Settlements
and Services Area

September 1977

Abstract

Based on stable population theory, a formula is derived relating childhood migration rates to adult migration rates. A simple approximation of this formula is tested on Swedish migration data and is found to work well. The reasons for its failure in some instances are also explored.

Acknowledgements

This work was done while I was at IIASA as a member of its summer graduate student program, and I am grateful to the Institute for organizing this program and the Ford Foundation for funding me. Thanks are also due to Andrei Rogers, in whose department I worked. Finally, I would like to thank Richard Raquillet for help in starting this project and for many helpful discussions along the way.

On The Relationship Of Childhood
To Labor Force Migration Rates

1. INTRODUCTION

One of the most obvious characteristics of a plot of age-specific migration rates is that there is a close correspondence between the rates for young children and the "labor force peak", that is the peak in the rates which usually occurs at about age 20 to 25. This is clearly seen in exhibit 1. It has been noted (Rogers and Castro, 1976) that there is an obvious reason why this should be so: young children do not move by themselves but move with their parents.

Based on certain assumptions, demographic theory can describe this correspondence exactly. If the theory is generally applicable it could provide a partial solution to the problem of estimating age-specific migration rates and understanding their structure. This paper presents some results from stable population theory and evaluates the performance of these formulas in describing Swedish migration data.

2. STABLE POPULATION THEORY

Let us use the conventional notation B(t) is the births at time t, $\ell(a)$ is the probability of surviving to age a, and m(a) is the instantaneous age-specific fertility rate. The earliest and latest possible ages of childbearing are, respectively, α and β.

Under the assumption of a stable population the number of births at time t is

$$B(t) = \int_{\alpha}^{\beta} B(t - a)\ell(a)m(a)da \quad ,$$

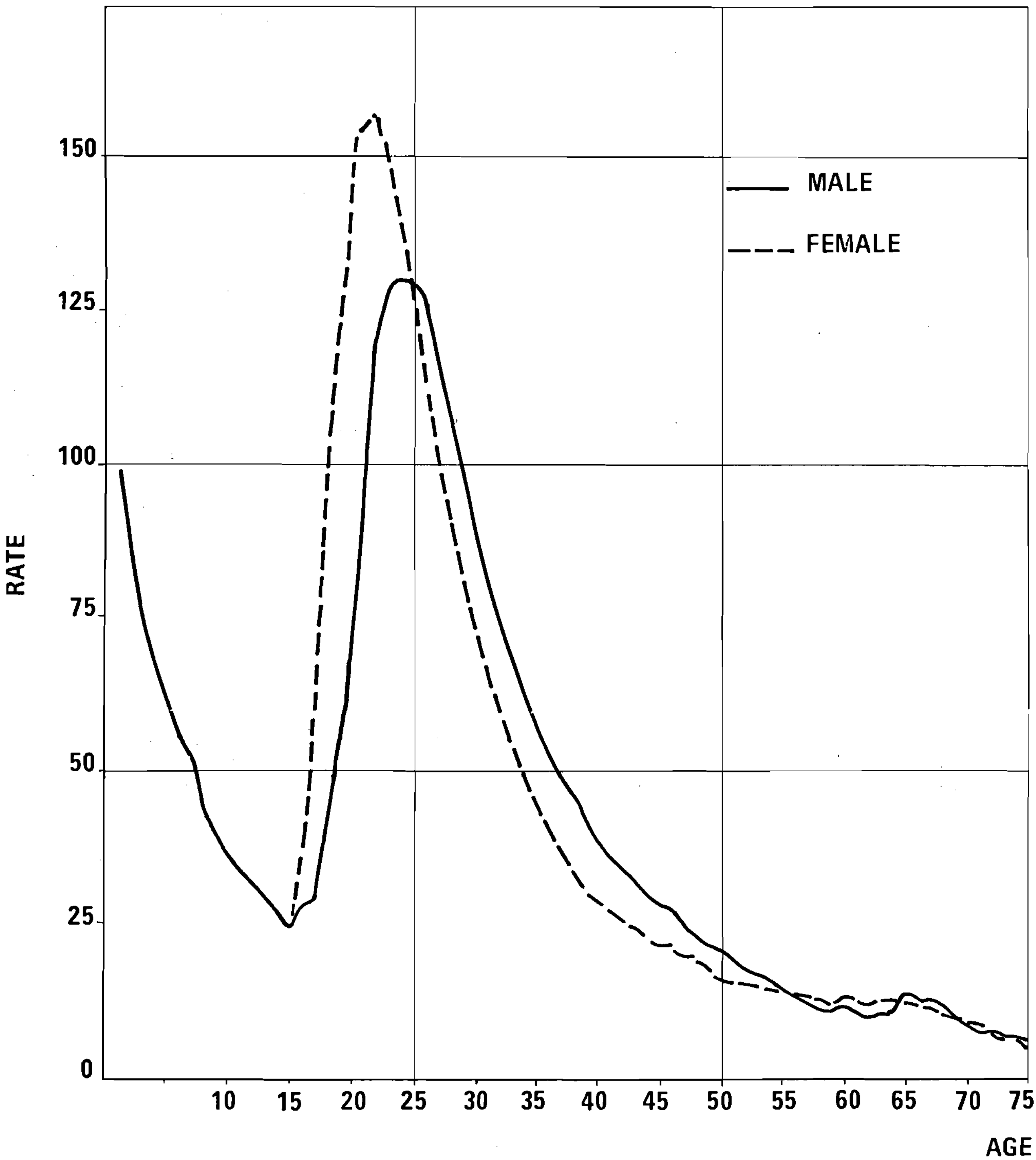
$$= \int_{\alpha}^{\beta} Qe^{r(t-a)}\ell(a)m(a)da \quad ,$$

of which

$$Qe^{r(t-a)}\ell(a)m(a)da$$

are from mothers aged a to a + da.

EXHIBIT 1



AGE-SPECIFIC ANNUAL MIGRATION RATES OF THE SWEDISH POPULATION BY SEX: AVERAGE OF 1968-1973.

Source: Internal Migration in Sweden 1968-1973, 1974, p. 10.

The proportion of all newborn babies whose mothers are aged a to a + da is then

$$\frac{Qe^{r(t-a)}\ell(a)m(a)da}{B(t)} = \frac{Qe^{r(t-a)}\ell(a)m(a)da}{Qe^{rt}}$$

$$= e^{-ra}\ell(a)m(a)da \quad .$$

For a child aged x to have a living mother aged a:

1) The mother must have been aged a - x when the child was born.
   The distribution of this is

$$e^{-r(a-x)}\ell(a - x)m(a - x)da \quad .$$

2) The mother must have survived the intervening x years. This has probability

$$\ell(a)/\ell(a - x) \quad .$$

Let us make the assumption that the probability a child migrates is the product of the probabilities that his or her mother is alive and she migrates,

$$\gamma(a)\ell(a)/\ell(a - x) \quad .$$

Note that this assumed independence of migration and fertility status. Then the child migration rate is simply the average of this proportion over all ages of mother

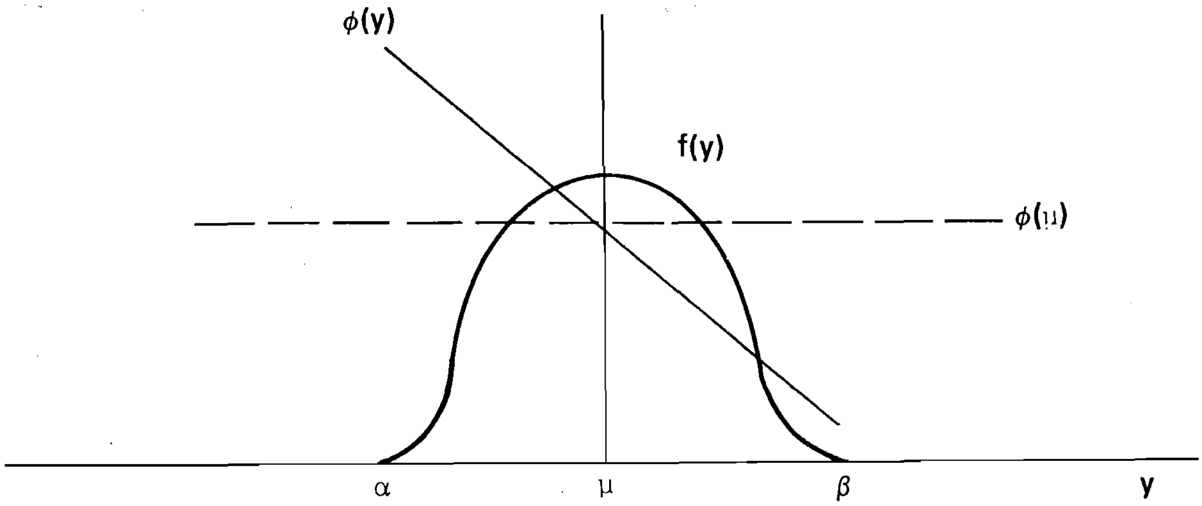
$$\gamma(x) = \frac{\displaystyle\int_{\alpha+x}^{\beta+x} \gamma(a)\,\frac{\ell(a)}{\ell(a - x)}\,e^{-r(a-x)}\ell(a - x)m(a - x)da}{\displaystyle\int_{\alpha+x}^{\beta+x} e^{-r(a-x)}\ell(a - x)m(a - x)da}$$

which, upon a change of variables y = a - x, becomes

$$\gamma(x) = \frac{\int_{\alpha}^{\beta} \gamma(y+x)\frac{\ell(y-x)}{\ell(y)}\,\ell^{-ry}\ell(y)m(y)dy}{\int_{\alpha}^{\beta} \ell^{-ry}\ell(y)m(y)dy} \qquad (1)$$

We first see that childhood migration rates are independent of fertility (with the exception of the dependence in r). This makes sense because rates are calculated on a per child basis, and the fact that a mother moves all of her children does not matter.

Suppose that $f(y) = \ell^{-ry}\ell(y)m(y)$ was symmetric around some point A, and that $\phi(y) = \gamma(y+x)\ell(y+x)/\ell(y)$ was linear in y. A schematic sketch is shown below.



Then by expressing $\phi(y)$ as $[\phi(y) - \phi(\mu)] + \phi(\mu)$,

$$\int_{\alpha}^{\beta} \phi(y)f(y)dy = \int_{\alpha}^{\beta} [\phi(y) - \phi(\mu)]f(y)dy + \int_{\alpha}^{\beta} \phi(\mu)f(y)dy$$

$$= 0 + \phi(\mu)\int_{\alpha}^{\beta} f(y)dy$$

by symmetry of f(y). In general, $\phi(y)$ need not be linear, but only $\phi(y) - \phi(\mu)$ need to be anti-symmetrics. This means that equation (1) could be approximated by

$$\gamma(x) = \alpha(\mu + x)\frac{\ell(\mu + x)}{\ell(\mu)} \tag{2}$$

Where μ is the mean age of fertility in the stable population,

$$\mu = \frac{\int_{\alpha}^{\beta} ye^{-ry}\ell(y)m(y)dy}{\int_{\alpha}^{\beta} e^{-ry}\ell(y)m(y)dy}$$

Although this derivation depends specifically on the assumption of a stable population and symmetry, it may be that formula (2) is more robust. This is comparable to the assumption that all births to a woman take place at exactly the mean age of childbearing, and this approximation has worked well in many cases. If the population is not stable some though should be given to the definition of μ.

In many countries, $\ell(\mu + x)/\ell(\mu)$ may be sufficiently close to one to be ignored.

3. DESCRIPTION OF DATA

In order to test the applicability of the foregoing theory, we have obtained single year age-specific migration data for Sweden in 1974 from the Swedish National Central Bureau of Statistics as part of IIASA's Comparative Migration and Settlement Study. The data consist of numbers of births, deaths, outmigrants by region and population on July 1 by sex and single year of age for each of the 24 Swedish counties (län). On the suggestion of Arne Arvidsson of the National Central Bureau of Statistics, they were aggregated into 8-county block (Viksområden) as shown in exhibit (2). County of destination was ignored and pooled outmigration rates were computed.

Although the formulas given in section 2 are stated in terms of mothers, theoretically they could have been derived for fathers as well. We did not attempt to verify a theory based on paternity for two reasons. First, the distribution of the ages of fathers of newborn boys is not usually as concentrated as the distribution of the ages of mothers. Second military service has a large effect on men very close to the peak age of migration. For these reasons we considered rates based on a closed female population.

EXHIBIT 2

Aggregation of Swedish Migration Data

Region (viksområden)	Counties (län)
1. Stockholm	Stockholm
2. Östra mellansverige (East-middle)	Uppsala Södermanland Östergötland Örebro Västmanland
3. Småland och öarna (East-middle)	Jönköping Kronoberg Kalmar Gotland
4. Sydsverige (South)	Blekinge Kristianstad Malmöhus
5. Västsverige (West)	Halland Göteborg och Bohus Älusborg Skaraborg
6. Norra mellansverige (North-middle)	Värmland Kopparberg Gävleborg
7. Mellersta norrland (Middle north)	Västernorrland Jämtland
8. Ovre norrland (Upper north)	Västerbotten Norrbotten

We speak loosely of the peak migration at around age 20 or 25 as "labor force migration". For males, a large part of the movements are probably caused by seeking new jobs. But for females, a large part of this could conceivably be women following their husbands and not directly related to employment. Since we are dealing only with the relationship of child to mother rates, the actual cause of the mother's migration is not important, and "labor force migration" is as good a term as any to describe this peak.

4. SMOOTHING OF AGE-SPECIFIC RATES

Section 2 suggests that the curve of childhood migration rates may be close to a pure translation of labor-force migration rates. Graphically, this hypothesis can be tested by simply superimposing the actual and translated curves. Alternatively, for each point in the childhood curve, $\gamma(x)$, we can calculate by interpolation the point $x + A_x$, such that $\gamma(x) = \gamma(x + A_x)$ and these values of A_x should be close to μ, the average age of childbearing in the stable population. For the smooth curve in exhibit 1, there is no problem in carrying out either of these comparisons, but when the data are disaggregated into 8 regions, the stochastic variation becomes relatively larger and the curves are not as smooth, and the comparison becomes more difficult.

A number of solutions to this problem are possible, including running means, fitting simple polynominals or splines to the curves and more complicated kinds of curve fitting. We have chosen to use a method described in Tukey (1977) called "non-linear smoothing". Compared to running means, it has the advantage of not being sensitive to a few large deviations from the curve, and unlike other curve fitting techniques very few assumptions are made concerning the shape of the fitted curve.

Because the technique is relatively new, and not well known to demographers, a short description may be helpful. A more detailed version is given in chapters 7 and 16 of Tukey (1977).

There are three main tools: "running medians", "splitting", and "hanning". "Running medians" works as follows. Given a sequence of x_1, i = 1,n, each x_i is replaced by the median of the three numbers x_{i-1}, x_i, and x_{i+1}. This procedure is repeated until no further changes take place. A special procedure is applied to the end point x_1 and x_n. This procedure tends to create many "plateaus" of two adjacent points with the same value. "Splitting" is a means of dealing with this. Finally, the data is "hanned" by replacing each x_i by $1/4\ x_{i-1} + 1/2\ x_i + 1/4\ x_{i+1}$. To get a sequence of x_i which more closely follows the original data but is still smooth, the entire process is repeated in the residuals, that is the differences between the original and smoothed sequence. The smoothed residuals are added to the original smoothed data to obtain the final smoothed curve. More details and a fully worked out example are given in Appendix A.

To illustrate the effect of these techniques, the raw and smoothed age-specific migration rates of each of the eight regions are presented in exhibit 3.

5. RESULTS

As mentioned in the previous section, two basic approaches were used to test the validity of the theory. In the first case, the actual rates are compared graphically and numerically to rates obtained by the full formula (1). Secondly, for each age from 2 to 11 inclusive, the quantity A_x such that $\gamma(x) = \gamma(x + A_x)$ is calculated. These values are then compared to the mean ages of childbearing in both the actual and stable populations.

From the beginning, we see that the factor $\ell(\mu + x/\ell(\mu)$ can be ignored in studying migration in Sweden. The value of $1 - \ell(\mu + x)/\ell(\mu)$ ranges from 0.07% to 0.9% in the present data. For other data this adjustment might be necessary.

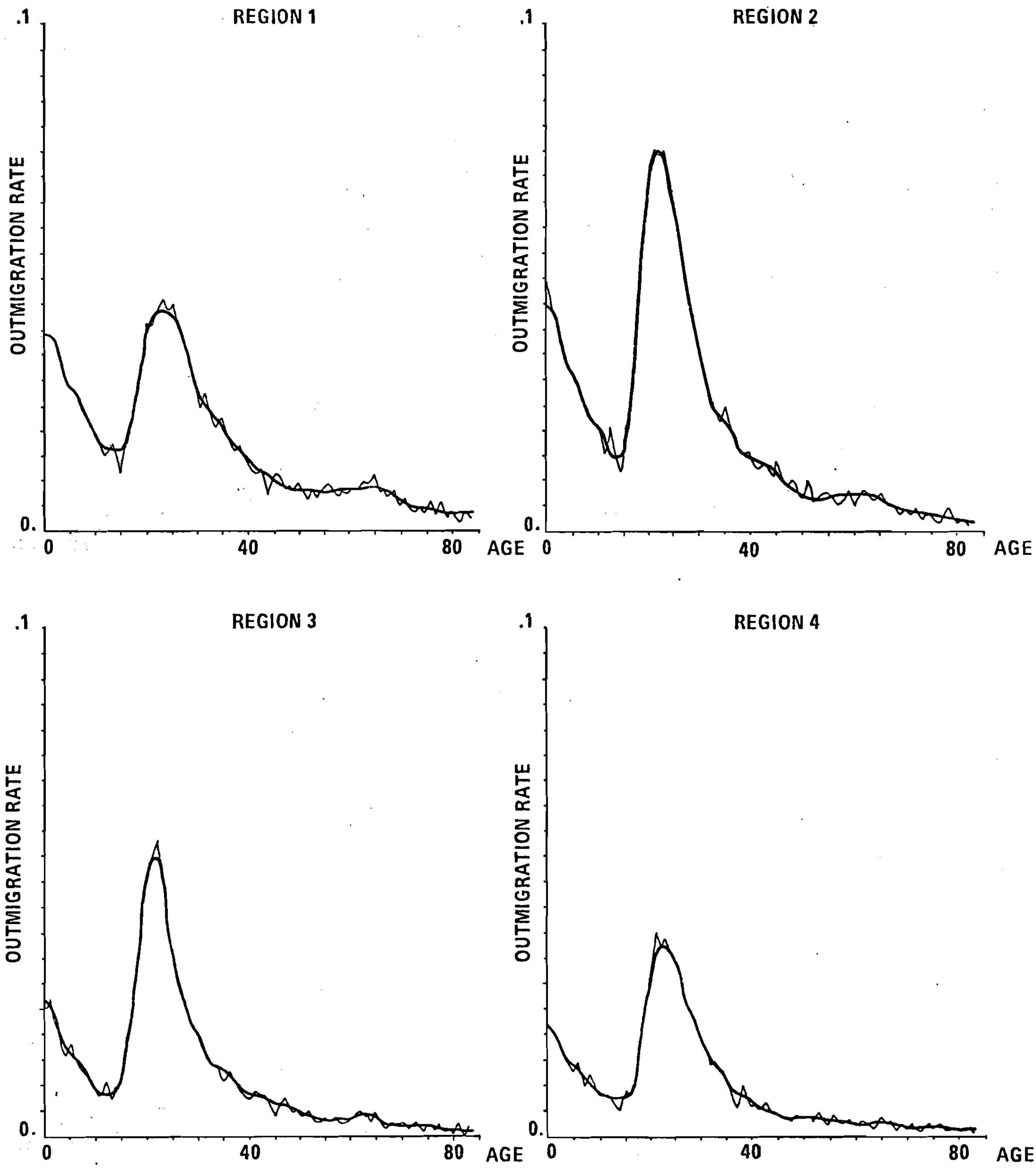
5.1 Comparison of Rates

For each region a life table was computed and the intrinsic growth rate calculated as if there were no migration. The mean age of childbearing, μ, was then calculated based on the actual age-specific birth rates and the resultant stable population. After the age-specific rates were smoothed, the "actual" (smoothed) rates were plotted in exhibit 4 along with the "translated" rates $\gamma(x + \mu)$. The full formula (1) was used to obtain the "calculated" rates and these are also plotted in exhibit 4.

For most regions, it seems that the translated rates behave quite well in estimating the actual rates. The residual plots show the residuals generally oscillating around the zero mark. On the other hand, the calculated rates are usually greater than the actual rates. The major exceptions are regions one and eight where the residuals are, respectively, all negative and almost all positive. To study these residuals quantitatively, a technique called median polish is used. A matrix, is formed from the residuals (that is the translated minus the actual rates) in exhibit 4.

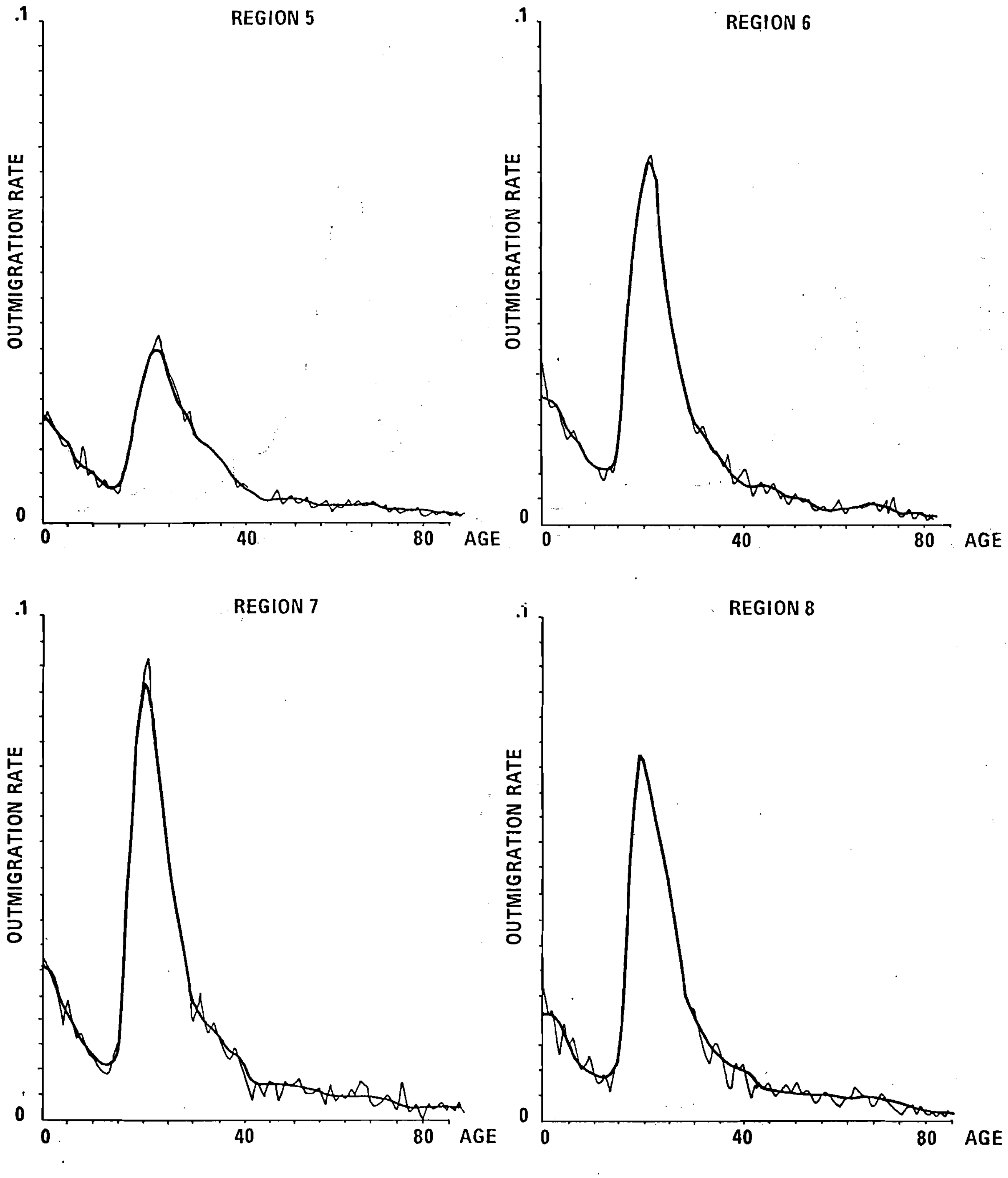
In finding the row and column effects in an analysis of variance, the means of each row and column are compiled. In a median polish these means are replaced by row and column medians. The main advantage is that the medians are not as sensitive as means to a few stray values. Unlike with means, the process can be improved by iteration. One first subtracts the column medians

EXHIBIT 3



AGE-SPECIFIC MIGRATION RATES, RAW AND SMOOTHED

EXHIBIT 3 (CONTINUED)



SPECIFIC MIGRATION RATES, RAW AND SMOOTHED from the original data, and then subtracts the row medians from the residuals. This process is then repeated on the residuals until no further change takes place. The final result is that the data is separated into

data = row effect + column effect + residual.

One way to summarize the ability of this linear model to explain the data is to compare the sum of the absolute values of the residuals with the sum of the absolute deviations of the original data from its median. For a more detailed discussion of this technique, see Tukey (1977). A worked example is given in Appendix B.

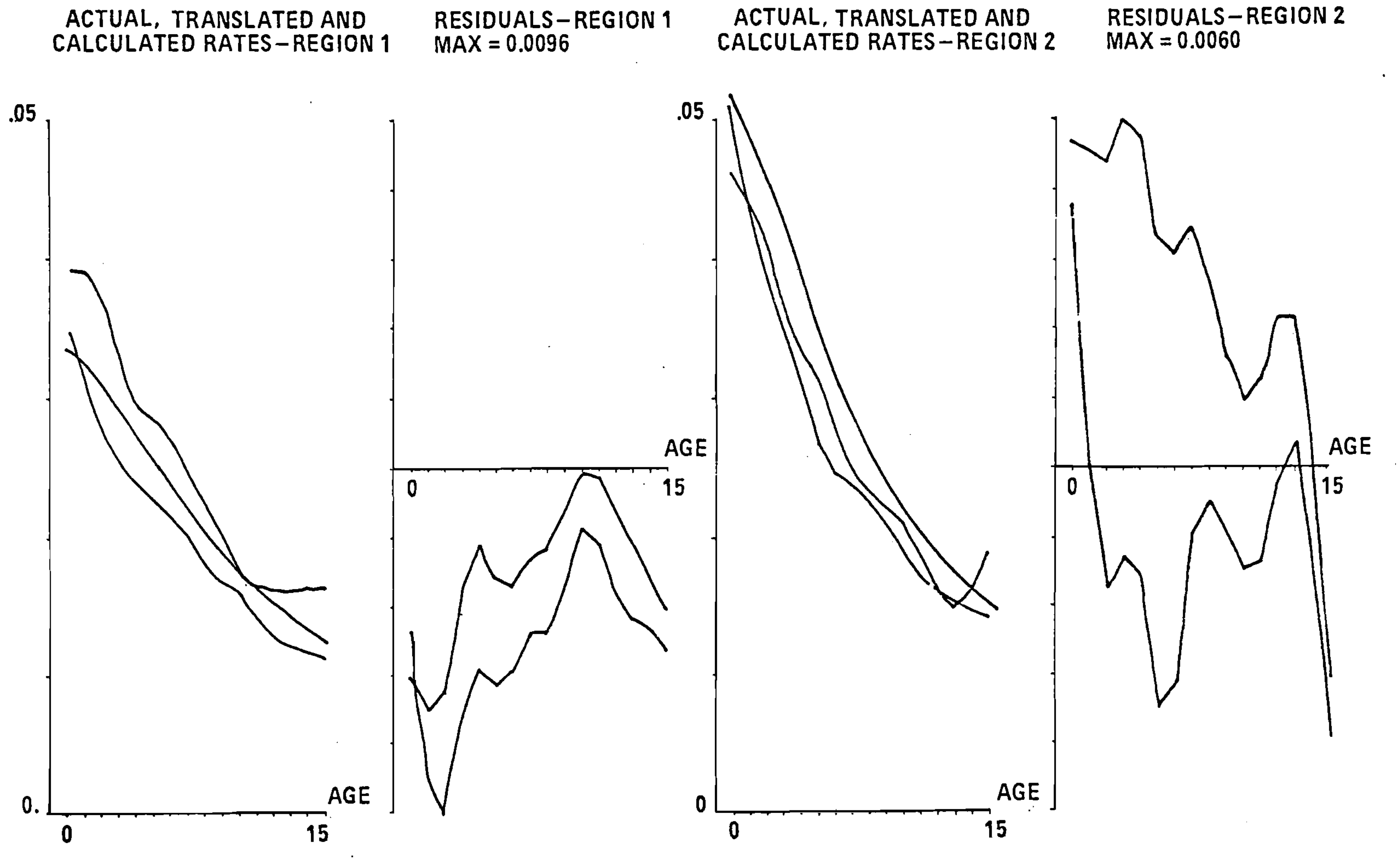
Exhibit 5 gives the results of median polishing the residuals in exhibit 4. Overall, 52% of the sum of absolute deviation was explained. The column of "region" effects show that the translated rates are low by about .005 for region 2, high by .0016 for region 8 and much closer for the other regions. This is about 15% lower for region 1 and 5% higher for region 8. For the other regions the translation obviously works better. The row or "age" effects summarize the difference in shape of the actual and translated rates by looking for regularities in the residuals. It is obvious that the translated rates are too low at ages 14 or 15 and too high at ages 0 and 1. Perhaps this means that 14 or 15 is the beginning of the labor-force peak and that parents of very young children are less likely to move than other people their age. It would take more careful study to prove these assertions. There is also a subtle pattern over the intermediate ages although there should be none if the theory holds. But the effect is small compared to the size of the actual rates.

5.2 Comparison of Translations

After the rates have been smoothed, for each age x we can calculate the translation A_x such the $\gamma(x) = \gamma(x + A_x)$. This has been done for ages 2 through 11, and the resultant values provide another means of testing the "translation" formula (2). The average A_x should reflect the mean age of childbearing in the stable population and there should be no dependence of the A_x on x, for this would indicate a failure of the model. The A_x values for each region are plotted as a function of x in exhibit 6.

The average A_x, defined as $Av = \sum A_x/10$, and μ, the mean age of childbearing in the stable population are plotted there as well. We first notice that there is a tendency for a V-shaped pattern in the A_x. This could indicate that the curve for children has a slightly different shape than for labor-force rates.

EXHIBIT 4



+ – Rates translated by equation 2

Δ – Rates calculated by equation 1

EXHIBIT 4 (CONTINUED)

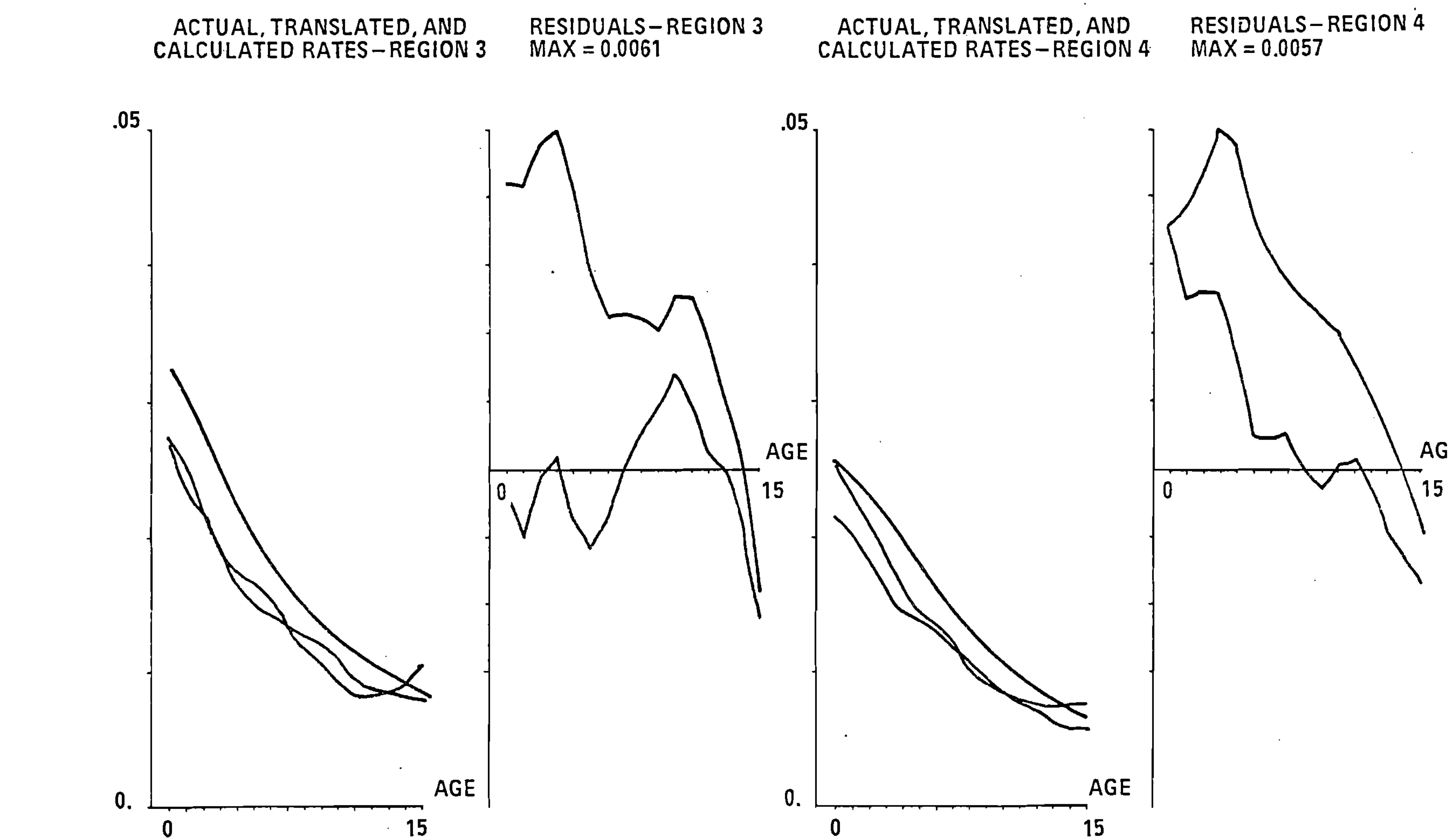


EXHIBIT 4 (CONTINUED)

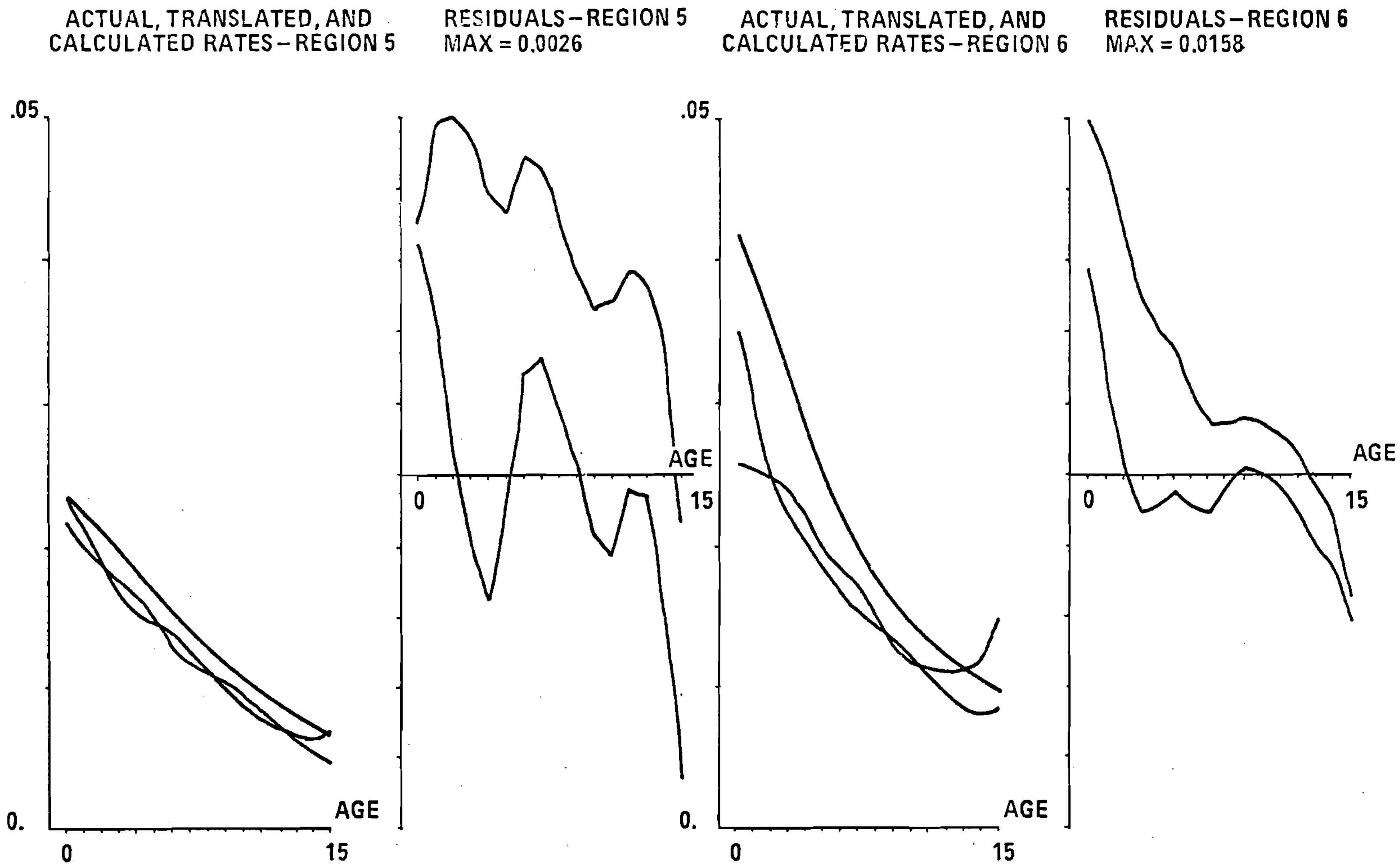


EXHIBIT 4 (CONTINUED)

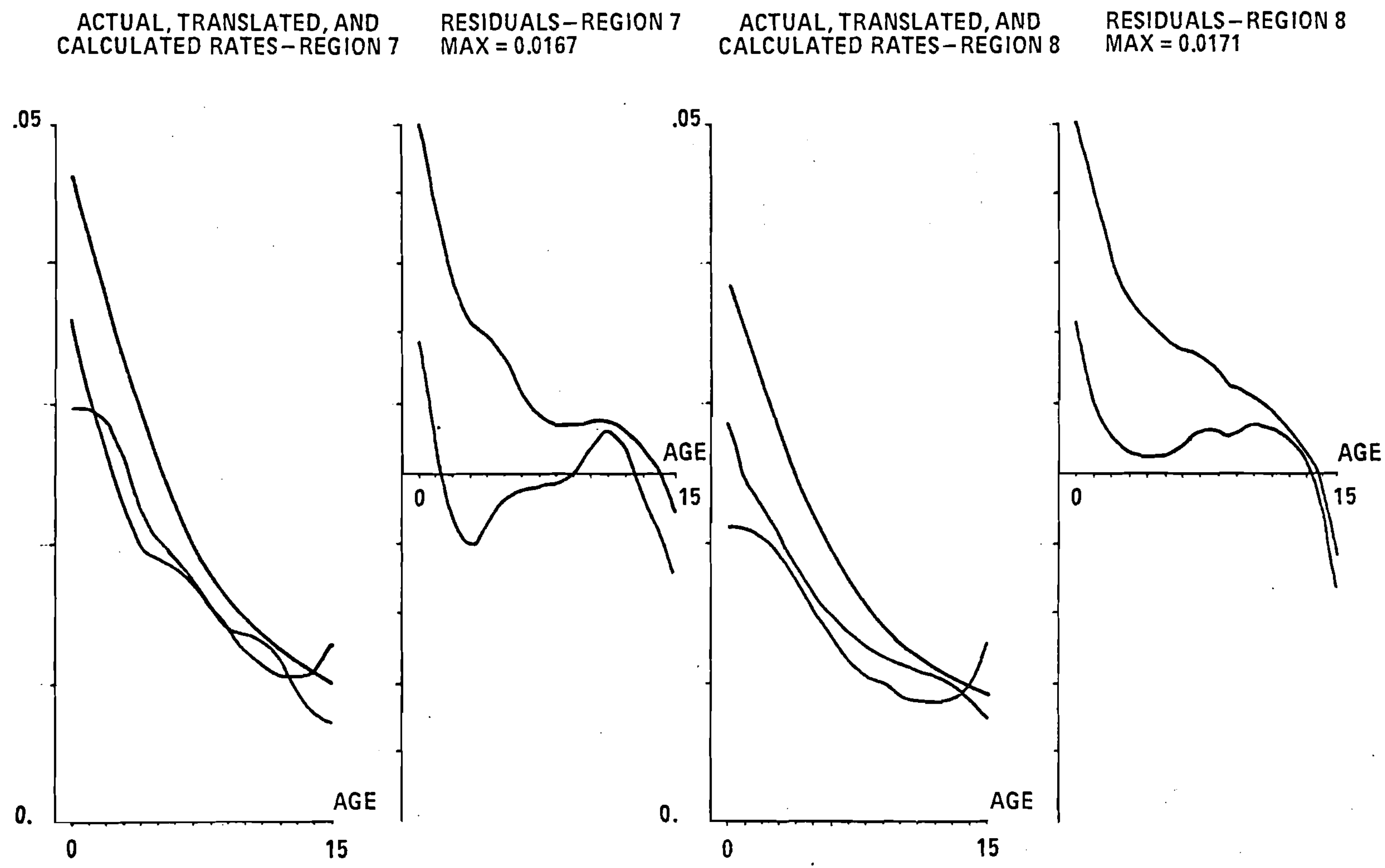


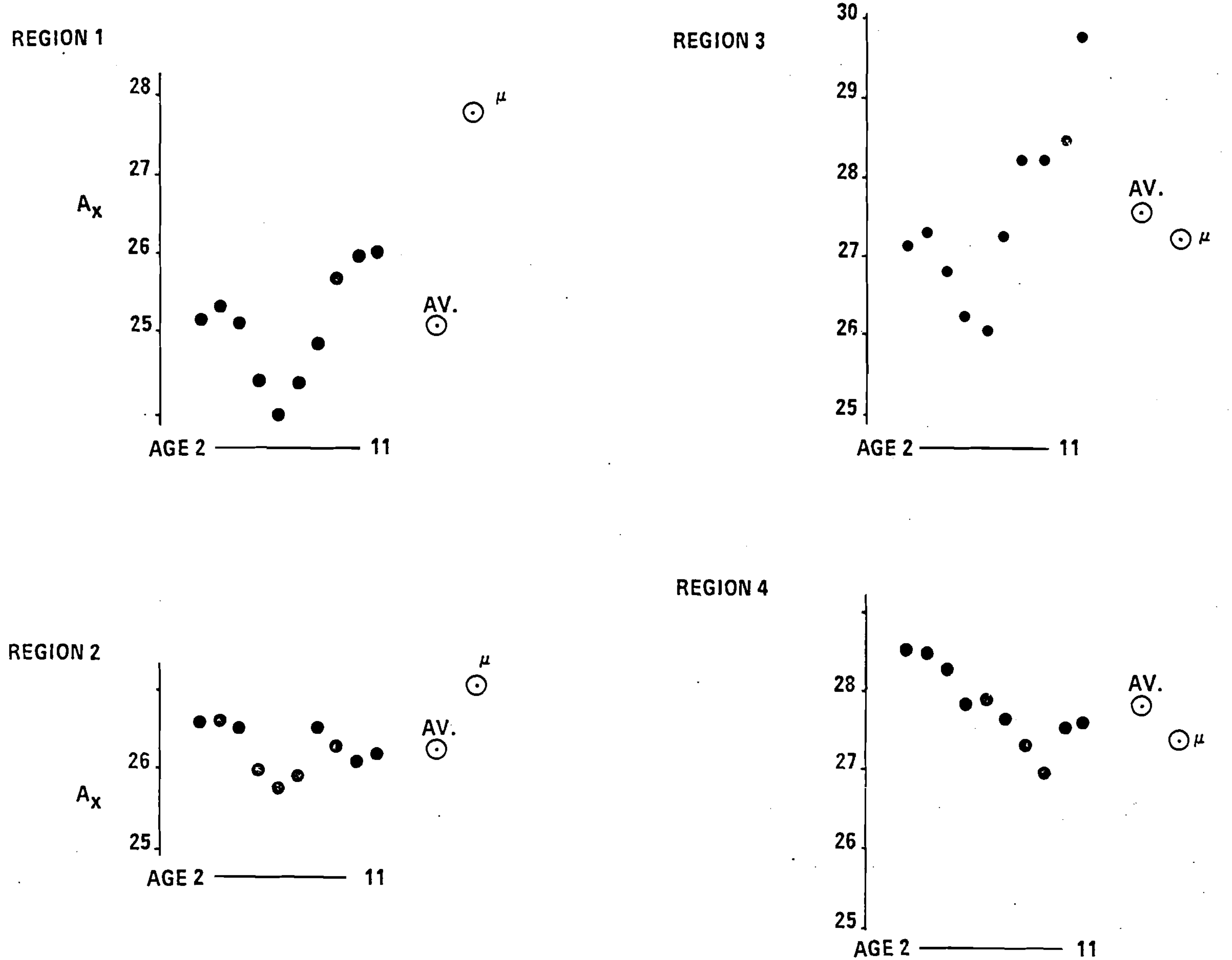
EXHIBIT 5

Results of Median Polish of Residuals in Exhibit 4

Region		Age	
1	-.00486	0	.00476
2	-.00154	1	.00140
3	.00014	2	-.00014
4	.00010	3	-.00057
5	.00005	4	-.00082
6	-.00084	5	-.00064
7	-.00064	6	-.00054
8	.00164	7	.00018
		8	.00034
		9	.00048
		10	.00082
		11	.00045
		12	.00013
		13	-.00038
		14	-.00141
		15	-.00218

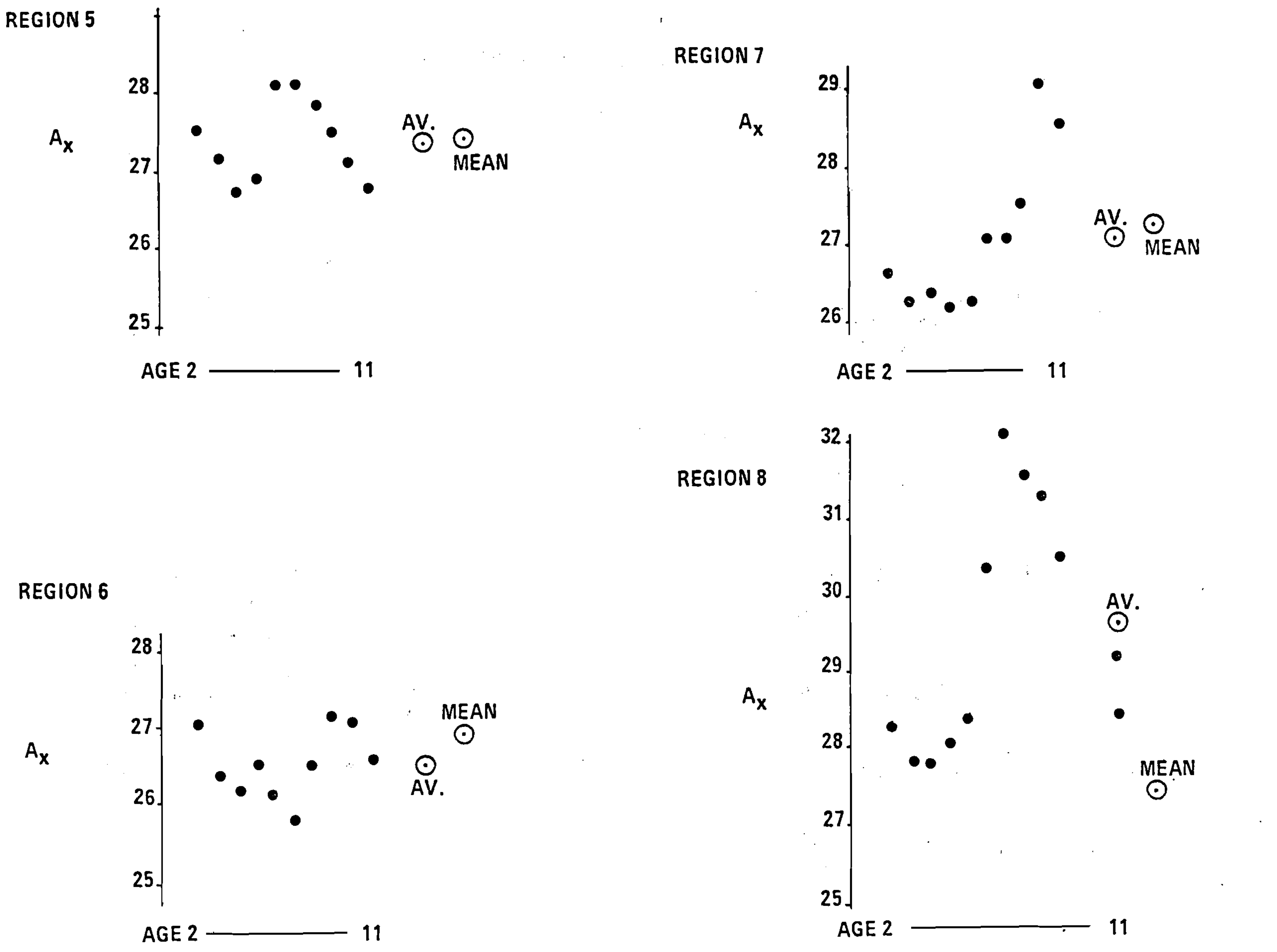
Reduction in sum of absolute deviations = 52%.

EXHIBIT 6



RELATIONSHIP OF A_x TO AGE.

EXHIBIT 6 (CONTINUED)



RELATIONSHIP OF A_x TO AGE

EXHIBIT 7

Comparison of Average A_x and μ

Region	Av.	μ	Difference
1	25.07	27.79	-2.72
2	26.22	27.02	-0.80
3	27.52	27.20	0.32
4	27.79	27.32	0.47
5	27.37	27.42	-0.05
6	26.54	26.92	-0.38
7	27.06	27.25	-0.19
8	29.62	27.41	2.21

$$\text{Av.} = \frac{1}{10} \sum A_x$$

$$\mu = \frac{\int_{\alpha}^{\beta} y\, e^{-ry} \ell(y) n(y)\, dy}{\int_{\alpha}^{\beta} e^{-ry} \ell(y) m(y)\, dy}$$

EXHIBIT 8

Results of Median Polish of Ax Values in Exhibit 6

Effects

Regions		Ages	
1	25.04	2	.08
2	26.28	3	-.08
3	27.40	4	-.40
4	28.08	5	-.48
5	27.34	6	-.58
6	26.52	7	-.23
7	26.78	8	.22
8	29.44	9	.64
		10	.72
		11	.50

Reduction in sum of absolute deviations 51%.

The mean A_x for each region, the corresponding mean age of childbearing in the stable population, μ, and their difference are given in exhibit 7. The median difference is -0.12 years, and the median of the absolute differences is only 0.40 years. As in the previous approach, there are two glaring deviations. The average Av is 2.72 years too low for region 1 and 2.21 years to high for region 8.

To explore for age patterns in the A_x, we can median polish the matrix of A_x (rows corresponding to age and columns to region). The results of doing this are given in exhibit 8. 49% of the absolute deviation is explained by the median polish. As in the median polish of the residual rates, there is a hint of a pattern in the age effect.

The indication of this analysis is that in six of eight cases the translation technique works well, and in two it is far off. In addition, there is a weak indication that the shapes of the childhood and adult curves are different.

5.3 Lack of Fit

To investigate the failure of the method for regions 1 and 8, we look now at the main assumption: stability. Exhibit 9 plots the normalized stable and actual population for each region and the differences of the two curves. The differences among the eight regions are seen most clearly in the residual plots, and the differences seem to be mainly in the prescence and relative size of three bumps between ages 0 and 35.

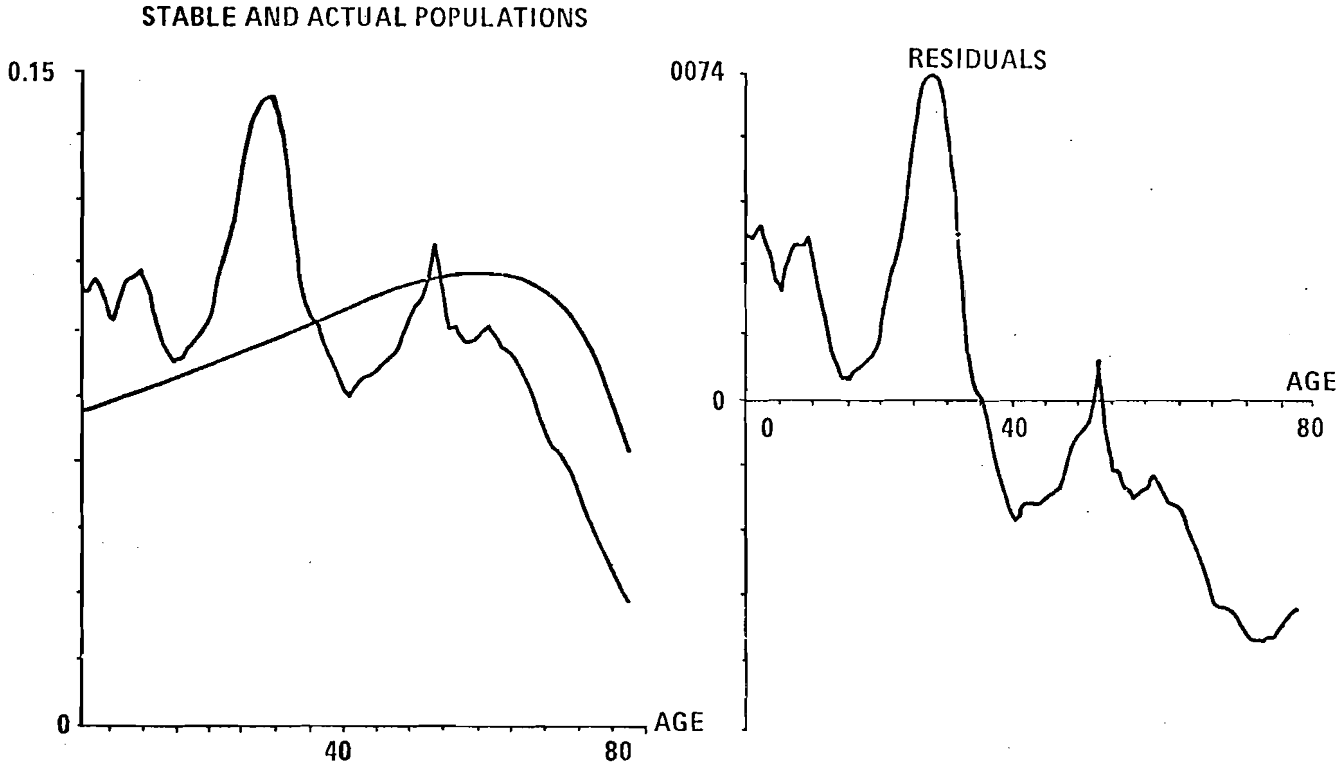
To attack this problem numerically, a number of indicators of stability (that is, closeness of the actual and stable population) are calculated and are given in exhibit 10. The main indicator of lack of fit is the difference between Av. and μ. A simple indicator of the conformance of the actual population to the stable population is just the sum over all ages of the absolute values of the differences of the normalized actual and stable populations. When the difference in average ages is plotted against the index (exhibit 11) no clean pattern emerges.

We can also calculate two other indices of stability by weighting the sum of the deviations by age-specific birth rates and migration rates. The results of these calculations are shown in exhibits 10 and 11. They indicate very little difference from the above.

When the difference in average ages is plotted against the intrinsic growth rate, r, a strong relationship shows up. Exhibit 11 shows that very negative values of r (as in Region 1) lead to actual A_x's smaller than the mean age of childbearing in the stable population and vice versa. This is reasonable in terms of the justification of the approximation formula (2).

EXHIBIT 9

REGION 1



REGION 2

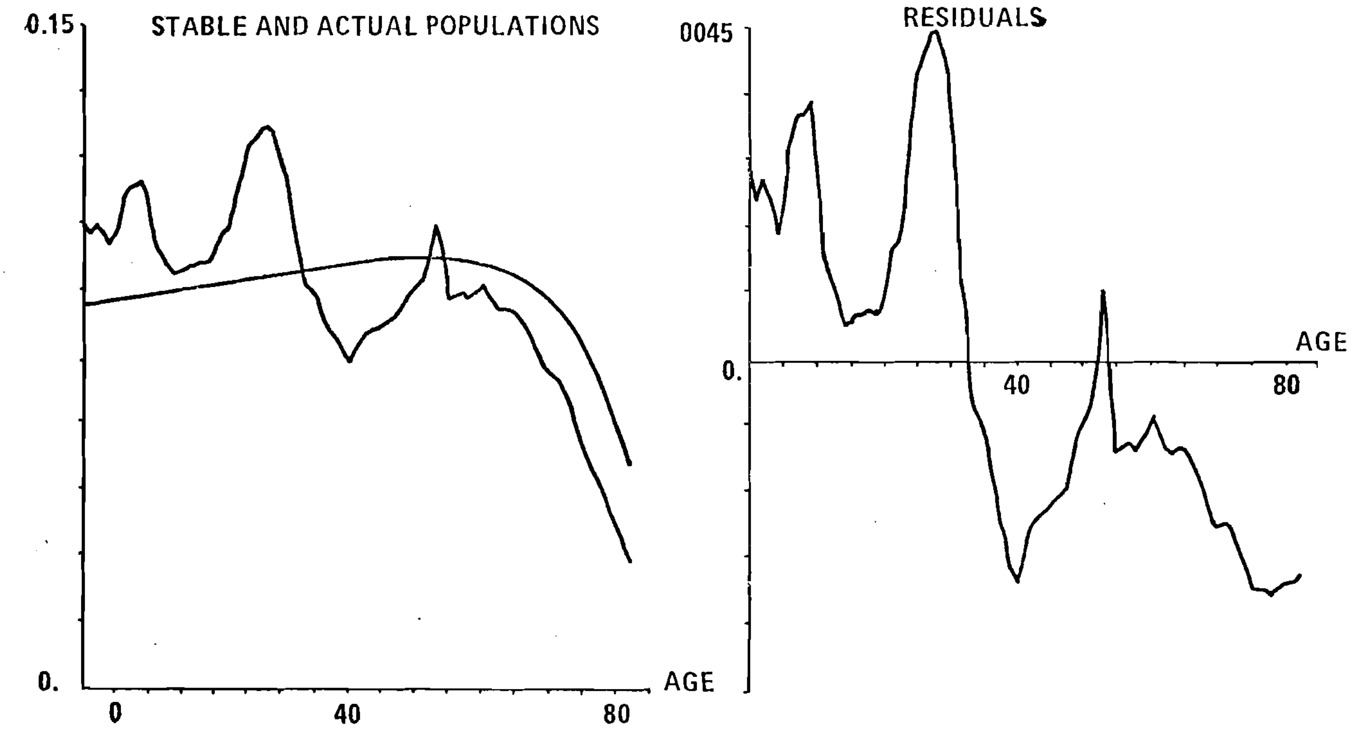
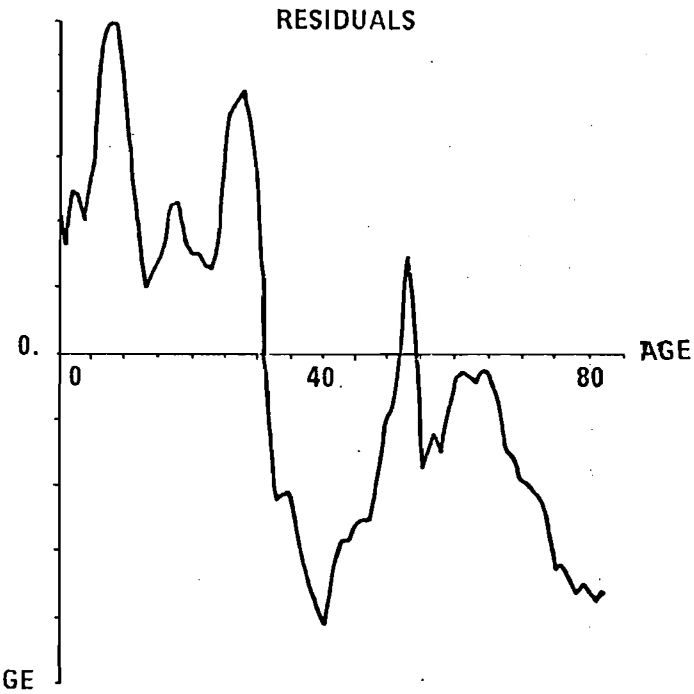


EXHIBIT 9 (CONTINUED)

REGION 3



REGION 4

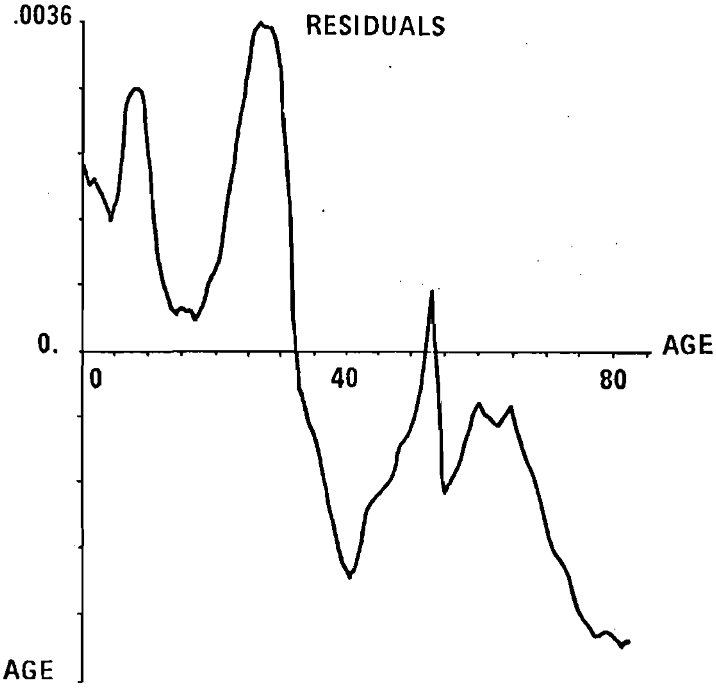
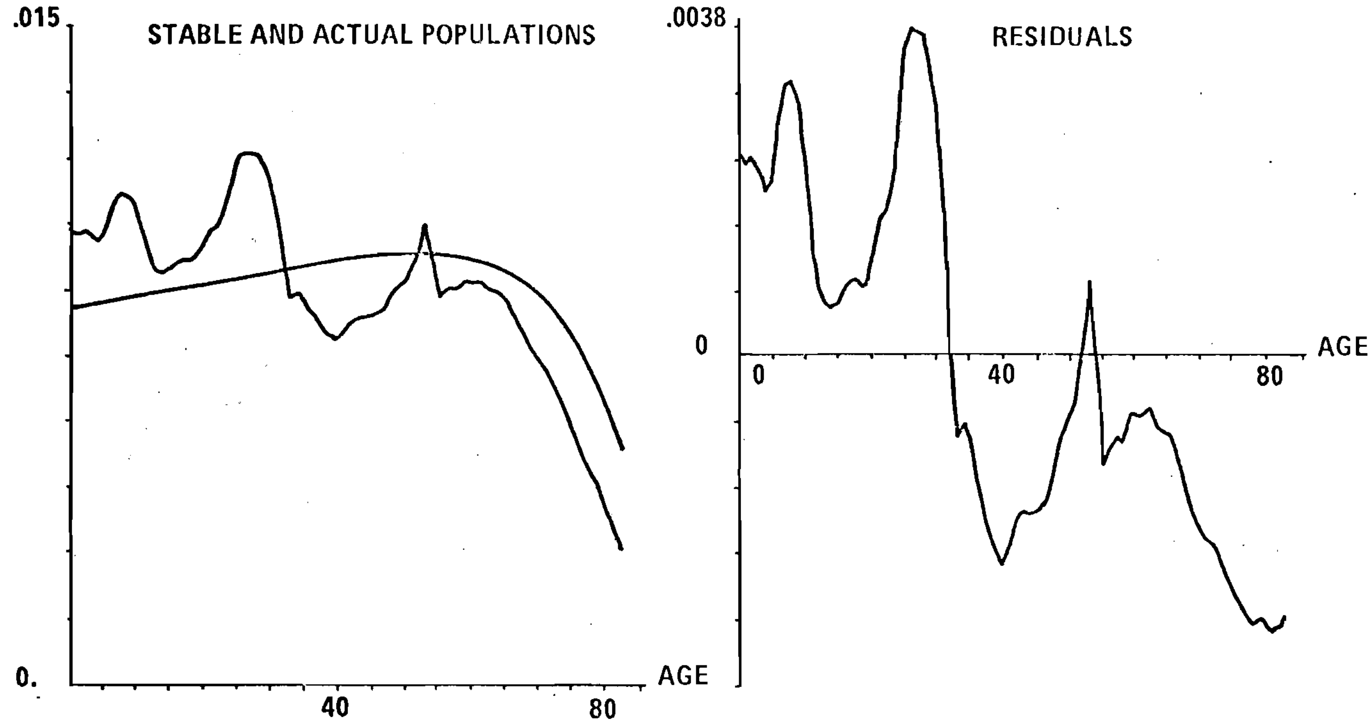


EXHIBIT 9 (CONTINUED)

REGION 5



REGION 6

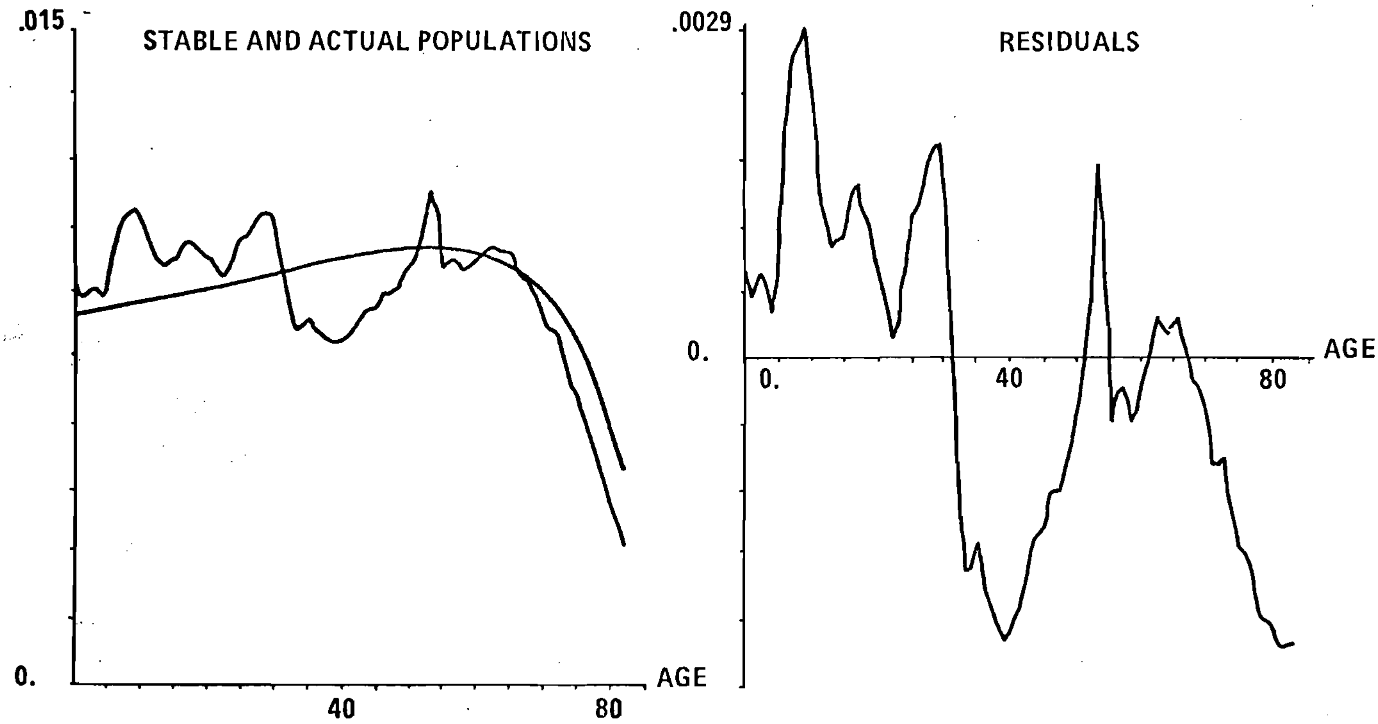
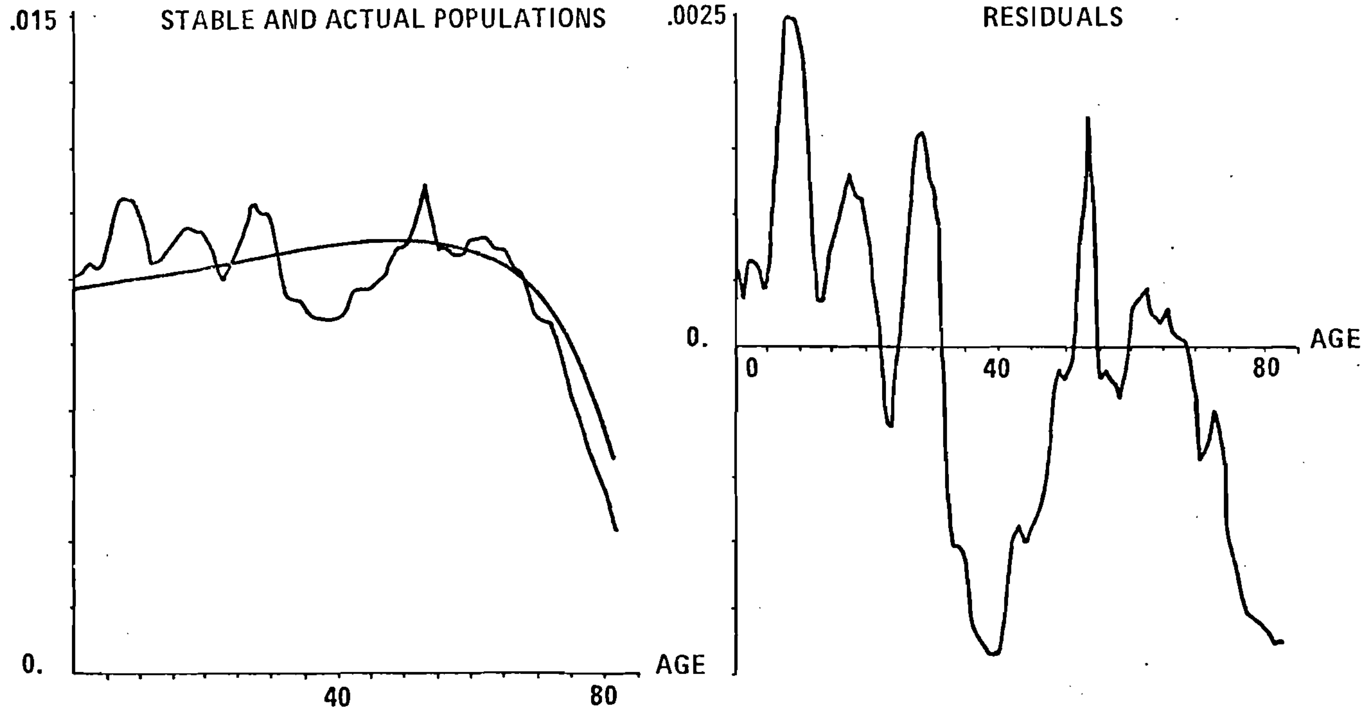


EXHIBIT 9 (CONTINUED)

REGION 7



REGION 8

EXHIBIT 10

Indicators of Stability

Region	Av. - μ	$\Sigma\lvert dev\rvert$	$\frac{\Sigma br\lvert dev\rvert}{\Sigma br}$	$\frac{\Sigma mr\lvert dev\rvert}{\Sigma mr}$	r(%)
1	-2.72	.251	4.92	3.25	-.790
2	-0.80	.173	2.80	2.13	-.344
3	0.32	.126	1.50	1.39	-.221
4	0.47	.152	2.32	1.86	-.397
5	-0.05	.159	2.43	1.99	-.344
6	-0.38	.116	1.18	1.16	-.420
7	-0.19	.099	0.92	0.96	-.344
8	2.21	.145	1.72	1.17	-.165

$$\text{Av.} = \frac{1}{10} \sum A_x$$

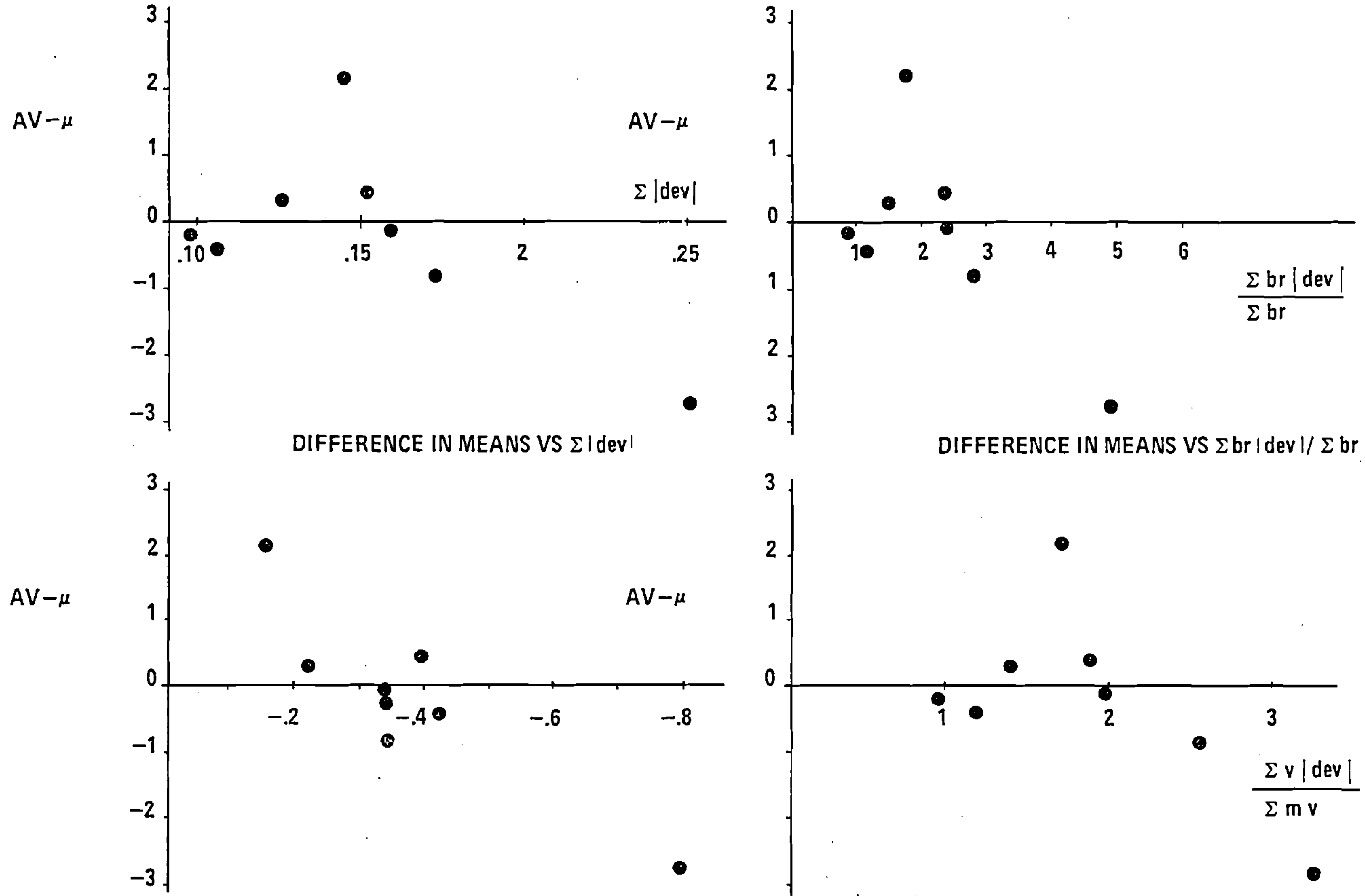
$$\mu = \frac{\int_{\alpha}^{\beta} y\, e^{-ry} \ell(y) m(y)\, dy}{\int_{\alpha}^{\beta} e^{-ry} \ell(y) m(y)\, dy}$$

$$|dev| = |c(x) - p(x)| \quad ,$$

where $c(x)$ is the proportion of the stable population aged x and $p(x)$ is the proportion of the actual population aged x.

br = age-specific birth rate
mr = age-specific migration rate

EXHIBIT 11



DIFFERENCE IN MEANS VS Σ |dev|

DIFFERENCE IN MEANS VS Σ br |dev|/ Σ br

DIFFERENCE IN MEANS VS γ

DIFFERENCE IN MEANS VS, Σ m v |dev|/ Σ m v

High negative values of r would make the curve f(y) [see the discussion before equation (2)] skew to the left, thus emphasizing lower values of $\phi(y)$ on the right. This would make both the translated and calculated childhood migration rates lower than the observed.

But a closer examination shows that changes of skewness are not the direct cause. One indicator of skewness is the third central moment of a distribution divided by the second,

$$\frac{\int_{\alpha}^{\beta} (y - \mu)^3 f(y)dy}{\int_{\alpha}^{\beta} (y - \mu)^2 f(y)dy} .$$

This was calculated for each region and for both the net maternity curve $f(y) = \ell(y)m(y)$ and the maternity curve in the stable population $f(y) = e^{-ry}\ell(y)m(y)$. The result, together with the differences of the two calculated skewness indices are presented in exhibit 12. The median difference is .029. For comparison purposes the deviations of the skewness calculations for the stable population from their median were also calculated. The median absolute deviation is .132, an order of the magnitude larger than the median skewness increment due to r. Since the skewness introduced by r is so small compared to the variation of skewness between regions, it is clear that r does not play a direct role in the determination of the skewness of the maternity curve in the stable population.

Moreover, skewness alone does not determine the applicability of the translation formula. For example, region 1 has the least skew net maternity curve but the worst failure of the model.

Another possible cause for the failure of the methods in region 1 and 8 can be explored by reviewing its derivation. Heuristically, equation 2 says that childhood migration occurs as if all children are born at exactly the mean age of childbearing and children move if and only if their mothers do. But if the population is not stable, perhaps the proper translation is not the mean age of childbearing in the stable population, μ, but rather in the actual population Z_1. These averages are compared in exhibit 13. Another possibility is simply that the mean age of childbearing in the stable multiregional population Z_M, should be used. These averages are also given in exhibit 13.

By comparing the differences of the average shift, Av, and each of the population averages, we easily see that neither Z_1

EXHIBIT 12

Skewness Comparison

Region	Skewness (1) Net Maternity	Skewness (2) Stable Population	Diff. (1)-(2)	Deviations (2)-Med(2)
1	2.150	2.084	.064	-.623
2	2.898	2.881	.017	.174
3	3.058	3.046	.012	.340
4	2.319	2.283	.036	-.423
5	2.761	2.739	.022	.033
6	2.674	2.616	.058	-.090
7	2.829	2.788	.041	.082
8	2.691	2.674	.017	-.032

EXHIBIT 13

Comparison of Average Ages

Region	Av.	μ	Z_1	Z_M	Av-μ	Av-Z_1	Av-Z_M
1	25.07	27.79	28.53	28.34	-2.72	-3.46	-3.27
2	26.22	27.02	27.83	27.60	-0.80	-1.61	-1.38
3	27.52	27.20	27.93	28.23	.32	-0.41	-0.71
4	27.79	27.32	28.13	28.37	.47	-0.34	-0.58
5	27.37	27.42	28.13	28.35	-.05	-0.76	-0.98
6	26.54	26.92	27.64	27.75	-.38	-1.10	-1.21
7	27.06	27.25	28.03	27.97	-.19	-0.97	-0.91
8	29.62	27.41	28.15	28.28	2.21	1.47	1.34

$$Av = \frac{1}{10} \sum A_x$$

$$\mu = \frac{\int_\alpha^\beta y e^{-ry} \ell(y) m(y) dy}{\int_\alpha^\beta e^{-ry} \ell(y) m(y) dy}$$

$$z = \frac{\sum ap(a)m(a)}{\sum p(a)m(a)} ,$$

where p(a) and m(a) are the actual population distribution and fertility rates.

Z_M = Mean age of childbearing in stable multi-regional population.

nor Z_M are better than μ. In fact, both tend to be about the same and in almost every case, worse than the corresponding μ.

So that lack of fit for regions 1 and 8 can not be directly related to lack of stability, skewness, or use of the wrong average age. The reason must be more complicated; region 1 is Stockholm and has been the destination of a number of migrants in the recent past, (1968-70 net inmigration was 20,587), but the trend was reversed (1971-73 net inmigration was -10,131). Region 8, the upper north, had severe out-migration in 1968-70 (-16,826), but this trend was reduced in 1971-73 (-1,998).[1] Perhaps the difficulty is due to this change, or some other peculiarity of the first and eighth regions. The answer does not seem to be in the data we have at hand.

6. CONCLUSION

After close examination of some internal migration data from Sweden, we can conclude that the translation formula derived in section 2 works reasonably well in calculating childhood migration rates from labor-force migration rates. An indirect link between lack of fit and extreme values of r were found, but no concrete explanation is available.

Trying out a formula on data from one country for one year hardly constitutes a test of a scientific theory. But the type of data needed for this analysis is hard to obtain, and this test, while it does not verify the theory does not invalidate it either.

If no other data become available and the formula turns out to hold over a wide range of circumstances, it will be a valuable addition to theoretical and practical migration studies.

[1] These figures are taken from a set of tables provided by Arne Arvidsson of the Swedish National Central Bureau of Statistics. Figures for region 8 include migration from Västerbotten to Norvbotten and vice versa, but this effect is probably small.

References

Rogers, A. and L.J. Castro, Model Multiregional Tables and Stable Populations, RR-76-09, International Institute for Applied Systems Analysis, Laxenburg, Austria, May 1976.

Swedish Central Bureau of Statistics (Statistiska Centralbyrån), (1974) Internal Migration in Sweden 1968-1973, No. 9, Stockholm.

Tukey, J.W., Exploratory Data Analysis, Addison Wesley, Reading, Massachusetts, 1977.

APPENDIX A: NON-LINEAR SMOOTHING

Perhaps the best way to understand the mechanisms of non-linear smoothing is to work through an example. Exhibit A1, column 2, displays the age-specific out-migration rates for region 7 from ages 52 to 68 (multiplied by 1000 for convenience). This age range was chosen to be short enough to handle but long enough to display all of the important ideas.

The first step is to take "running medians". Each value in the series, x_i, is replaced by the median of x_{i-1}, x_i, x_{i+1}. For example, $x_2 = 52$ is replaced by med $\{x_2,x_3,x_4\}$ = med $\{52, 50,58\} = 52$. Since med $\{x_2,x_3,x_4\}$ = med $\{50,58,59\} = 58$, x_3 is not changed. This process is carried out for each value in the series except the first and the last, which need a special procedure. The results of this procedure are shown in column 3. Only the changes have been noted.

After one pass through the series has been made, there will ordinarily still be groups of three points which do not have their middle point equal to their median. The process is, therefore, repeated as in column 4 and 5, until no further changes can be made.

Obviously, the same rule cannot be applied to the first or the last point. One alternative is as follows: replace x_1 by med $\{x_1,x_2,x_2 + (x_2 - x_3)\}$, and the corresponding thing for the last point. For our series, this rule says to replace 52 by the median of $x_1 = 52$ = the original end point, $x_2 = 52$ = the new adjacent point, and $x_2 + (x_2 - x_3) = 52 + (52 - 58) = 46$. This last number is the result of extrapolating a line through x_2 and x_3 to get an estimate of x_1. The result is then med $\{52,52,46\}$ = 52. For the last value we have med $\{x_n = 42, x_{n-1} = 27, x_{n-1} + (x_{n-1} - x_{n-2}) = 27 + (27 - 27) = 27\}$, so the end point changes from 42 to 27.

With running medias of this type, it often happens that resulting series have a number of pairs of adjacent identical values. A technique called "splitting" sometimes breaks these up and gives a smoother appearance. Consider the pair of 67's at ages 63 and 64. Break the series between this pair and apply the end point rule of the previous paragraph to the new ends $\{67,61,50,...\}$ and $\{67,48,27,...\}$. The results are $x_{63} = \{67,61,61 + (61-50) = 72\} = 67$ and x_{64} = med $\{67,48,48 + (48-27) = 69\} = 67$. In this case, no change is made. If the 67's had instead been, say, 77's x_{63} would be 71 and x_{64} would be 69. The results of the end point and splitting rules are given in column 6.

In general, this procedure is repeated until no further changes take place, as with the running medians. The procedure is a bit more complicated in some cases, for instance two adjacent pairs, and these further details are found in Tukey (1977, Chapter 7).

The final technique is known as "hanning". It is a form of running mean which gives a final, close smooth to the data. Each data point x_i is simply replaced by $1/4\ x_{i-1} + 1/2\ x_i + 1/4\ x_{i+1}$. End points are not changed. The results of this procedure are shown in column 7, rounded to the same number of digits as the previous numbers.

To see what we have done the "rough", that is the original series minus the smoothed series is shown in column 8. The whole procedure described above was repeated on these data; the result is given in column 13. It is usually the case that a second pass like this finds a bit more of a pattern in the residuals of the first pass. For a final smoothed series, we add the original smooth (column 7) and the "smooth of the rough" (column 13) to get the result in column 14.

The motivation of the technique is very simple: a method is sought which can smooth a rough series of numbers without making assumptions about their shape or being excessively influenced by a few stray points. The key step is the use of running medians. If a single point is far from its two adjacent values, it is reduced to one of the two less extreme points. The splitting and hanning steps work on a less extreme series and give it a polished look.

Exhibit A1

Non-Linear Smoothing Example

(1) Age	(2) #	(3) R	(4) R	(5) R	(6) E&S	(7) H	(8) Rough	(9) R	(10) R	(11) R	(12) E&S	(13) H	(14) Sm
52	52				52	52	0				0	0	52
	50	52				54	-4	0				0	54
	58					56	2	1				1	57
35	59	58				58	1	1				1	59
	29	59	58			55	-29	1				1	56
	70	40	47			50	23	-7	0			0	50
	40	47	46	47		47	-7	0	-2	0		0	47
	47	46	47			47	0	-2	0			0	47
60	46	47				48	-2	0				0	48
	61	50				52	9	-2	0			1	53
	50	61				60	-10	8	5			4	64
	74	67			67	66	8	5				5	71
	67				67	62	5	5				4	66
65	48					48	0	0				1	49
	27					32	-5	-1				-1	31
	26	27				27	-1					-1	26
68	42				27	27	15				-1	-1	26

Source: Age-specific migration rates for region 7, multiplied by 1000 for convenience.

APPENDIX B: MEDIAN POLISH

Exhibit B1 gives a worked out step by step example of a median polish. The data are three "averages" for each of the eight regions: the average of the 10 translations of the migration curve Av, the mean age of childbearing in the stable population z.

The procedure could start either way, but since most of the variation is probably from region to region (column to column) column medians are first calculated. These are shown at the bottom of panel A. The values in panel 2 are obtained by subtracting the medians from all of the values in their column. For instance, the upper left hand element is 25.07 - 27.79 = -2.72. For clarity and convenience, all numbers in panels 2 through 4 are multiplied by 100.

The procedure is then repeated for rows. The medians are shown on the right of panel 2 and the numbers in panel 3 are obtained by subtraction. The column medians are not zero, so they are subtracted to yield the values in panel 4. In this case, the row medians are zero to the procedure. In general, the procedure is carried out until there is little or no change in the residuals. This usually takes a few more steps than this example.

By adding up all of the row medians into a "row-effect" and the column medians into a "column effect" we can write

data = row effect + column effect + residual

For example, the column effect for region 3 is 27.52 - .31 = 27.21 and the row effect for Av is -.12. Then Av for region 3 can be broken down 27.52 = -.12 + 27.21 .43.

There is of course, an ambiguity here. Any constant can be added to all of the row effects and subtracted from all of the column effects and they would still add up to the original data. One solution is to subtract a constant from each so that the medians of the new effects and of the column effects are zero. This is done at the bottom of exhibit B1.

The effects are the summary measures. We can say that Av tends to be -.12 years lower and Z .72 years higher than μ. The average age is 27.33 plus as much as .46 years for some regions or minus as much as .41 for others.

The residuals tell us about how good the fit is. For an overall measure we can compare the sum of the absolute values of the residuals (7.22) to the sum of the absolute deviations (14.79) of the original data from the overall median 27.47. The reduction is 1 - 7.22/14.79 = 49%. A look at the residuals themselves tells us a few things as well. First of all the residuals in the second and third row are much smaller than the first.

This means that once the difference of .72 years between μ and Z is accounted for, these two measures are quite close. There is a lot more variation in Av. In particular, the Av residuals for region 1 (-2.60) and region 8 (2.31) indicate a certain peculiarity in these two regions.

Exhibit B1

Median Polish Example

Panel 1

Region	1	2	3	4	5	6	7	8
Av	25.07	26.22	27.52	27.79	27.37	26.54	27.06	29.62
μ	27.79	27.02	27.20	27.32	27.42	26.92	27.25	27.41
Z	28.53	27.83	27.93	28.13	28.13	27.64	28.03	28.15
	27.79	27.02	27.52	27.79	27.42	26.92	27.25	28.15

Panel 2

-272	-80	0	0	-5	-38	-19	147	-12
0	0	-32	-47	0	0	0	-74	0
74	81	41	34	71	72	78	0	72

Panel 3

-260	-68	12	12	7	-26	-7	159
0	0	-32	-47	0	0	0	-74
2	9	-31	-38	-1	0	6	-72
0	0	-31	-38	0	0	0	-72

Panel 4

	1	2	3	4	5	6	7	8	
	-260	-68	43	50	7	-26	-7	231	0
	0	0	-1	-9	0	0	0	-2	0
	2	9	0	0	-1	0	6	0	0
Column effects	27.79	27.02	27.21	27.41	27.42	26.92	27.25	27.43	
Normalized column effects	.46	-.31	-.12	.08	.09	-.41	-.08	.10	

	Av.	μ	Z
Row effects	-.12	0	-.72

Overall effect 27.33

Σ|deviations from median| = 14.79
Σ|residuals| = 7.22
reduction 49%

Note: decimal points dropped from panels 2-4.

Papers of the Migration and Settlement Study

September 1977

I. Papers in the Dynamics Series

1. Andrei Rogers and Frans Willekens, "Spatial Population Dynamics," RR-75-24, July, 1975, published in *Papers Regional Science Association*, Vol. 36, 1976, pp 3-34.

2. Andrei Rogers and Jacques Ledent, "Multiregional Population Projection," internal working paper, August 1975, published in *Optimization Techniques: Modelling and Optimization in the Service of Man*, Part 1, ed. Jean Cea, Springer-Verlag, Berlin, 1976, pp 31-58.

3. Andrei Rogers and Jacques Ledent, "Increment-Decrement Life Tables: A Comment," internal working paper, October 1975, published in *Demography*, *13* (1976), pp 287-290.

4. Andrei Rogers, "Spatial Migration Expectancies," RM-75-57, November 1975.

5. Andrei Rogers, "Aggregation and Decomposition in Population Projection," RM-76-11, February 1976, published in revised form in *Environment and Planning A*, *8* (1976), pp 515-541.

6. Andrei Rogers and Luis J. Castro, "Model Multiregional Life Tables and Stable Populations," RR-76-09, May 1976.

7. Andrei Rogers and Frans Willekens, "Spatial Zero Population Growth," RM-76-25, April 1976.

8. Frans Willekens, Sensitivity Analysis," RM-76-49, June 1976. First part published in revised form as "Sensitivity Analysis in Multiregional Demographic Models, *Environment and Planning*, *A9* (1977), pp 653-674.

9. Andrei Rogers and Frans Willekens, "The Spatial Reproductive Value and the Spatial Momentum of Zero Population Growth," RM-76-81, December 1976.

10. Frans Willekens, "The Spatial Reproductive Value: Theory and Applications," RM-77-09, February 1977.

11. Jacques Ledent, "Intrinsic Rates and Stable Age-Specific Mortality (and Migration) Rates of the Growth Matrix Operator in the Single Region (Multiregion) Population Model", RM-77-37, July 1977.

12. Michael Stoto, "On the Relationship of Childhood to Labor Force Migration Rates", RM-77-44, September 1977.

II. Papers in the Demometrics Series

1. John Miron, "Job-Search Migration and Metropolitan Growth," RM-76-68, August 1976.

2. Andrei Rogers, "The Demometrics of Migration and Settlement," RM-76-68, August 1976, forthcoming in Papers of the Regional Science Association, British Section, Pion Ltd., London.

III. Papers in the Policy Analysis Series

1. Yuri Evtushenko and Ross D. MacKinnon, "Non-Linear Programming Approaches to National Settlement System Planning, "RR-75-26, July 1975, published in revised form in Environment and Planning A, 8 (1976), pp. 637-653.

2. R.K. Mehra, "An Optimal Control Approach to National Settlement System Planning, "RM-75-58, November 1975.

3. Frans Willekens, "Optimal Migration Policies," RM-76-50, June 1976.

4. Anatoli Propoi and Frans Willekens, "A Dynamic Linear Programming Approach to National Settlement System Planning," RM-77-08, February 1977.

IV. Papers in the Comparative Study Series

1. Ross D. MacKinnon and Anna Maria Skarke, "Exploratory Analyses of the 1966-1971 Austrian Migration Table," RR-75-31, September 1975.

2. Galina Kiseleva, "The Influence of Urbanization on the Birthdate and Mortality Rate for Major Cities in the USSR," RM-75-68, December 1975.

3. George Demko, "Soviet Population Policy," RM-75-74, December 1975.

4. Andrei Rogers, "The Comparative Migration and Settlement Study: A Summary of Workshop Proceedings and Conclusions, "RM-76-01, January, 1976.

5. Andrei Rogers, "Two Methodological Notes on Spatial Population Dynamics in the Soviet Union," RM-76-48, June 1976.

6. Frans Willekens and Andrei Rogers, "Computer Programs for Spatial Demographic Analysis," RM-76-58, July 1976.

7. Frans Willekens and Andrei Rogers, "More Computer Programs for Spatial Demographic Analysis," RM-77-30, June, 1977.