Andrei Rogers

Estimating Interregional Population and Migration Operations from Interregional Population Distributions

Institute of Urban & Regional Development
University of California, Berkeley

Reprint No. 36

Reprinted from DEMOGRAPHY
Vol. 4, No. 2, 1967

Printed in U.S.A.

ESTIMATING INTERREGIONAL POPULATION AND MIGRATION OPERATORS FROM INTERREGIONAL POPULATION DISTRIBUTIONS

ANDREI ROGERS*

RESUMEN

Una limitación común en el análisis de los sistemas interregionales de población, es la ausencia de datos confiables que describan el comportamiento de los componentes fundamentales de los cambios de población. Como resultado, frecuentemente, los demógrafos tienen que recurrir a medidas imperfectamente construídas del incremento natural y de la migración neta.

Recientes intentos de expresar el crecimiento de la población interregional en forma de una matriz sugieren, sinembargo, un método para estimar el régimen de crecimiento de un sistema multirregional sólo en base a datos históricos sobre distribuciones interregionales de población. Este artículo describe e ilustra tal método. Desde que los resultados son bastante insatisfactorios, gran parte de este trabajo está dedicada a analizar las condiciones que determinan un bajo rendimiento del método, evaluando, en particular, índices de estimacion alternativos y considerando la posibilidad de "suavizar los datos."

SUMMARY

A common constraint in analyses of interregional population systems is the absence of reliable data for describing the behavior of the fundamental components of population change. As a result, demographers frequently have had to rely on crudely constructed measures of natural increase and net migration.

Recent efforts to express interregional population growth in matrix form, however, suggest a method for estimating the regime of growth of a multiregional system solely on the basis of historical data on interregional population distributions. This paper describes and illustrates such a method. Since the results are largely unsatisfactory, a large portion of the paper is devoted to an analysis of the conditions leading to the poor performance of the method, by evaluating, in particular, alternative estimators and by considering the possibility of "data smoothing."

Efforts to analyze and forecast interregional population growth and distribution frequently are severely constrained by a general paucity of reliable data for describing the behavior of the fundamental components of population change. Migration data, for example, are simply not available for most regional subdivisions of the nation. As a result, demographers have had to rely on rather crudely constructed "residual" methods of estimating migration. Recent efforts to express the population growth process in matrix form, however, suggest a means whereby the growth regime of an interregional system may be estimated solely on the basis of a time series of interregional population distributions. In this paper, we present such a method and illustrate it using data on a two-region system consisting of California and the rest of the United States.

For ease of exposition we shall focus on the simplest possible population model. Such a model is the cohort-survival model with only a single cohort: the total population. This model may be defined by matrix equation (1).[1]

* University of California, Berkeley. The author wishes to acknowledge the invaluable services rendered by his research assistant, Mr. Edwin Kampmann, who performed most of the data processing chores involved in this paper. Thanks are due the Center for Planning and Development Research and the Institute of Social Sciences, both of the University of California, Berkeley, for financial support.

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

$$\begin{pmatrix} w_1^{(t+1)} \\ w_2^{(t+1)} \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ w_m^{(t+1)} \end{pmatrix} = \begin{pmatrix} g_{11} & g_{21} & \cdot & \cdot & \cdot & \cdot & g_{m1} \\ g_{12} & g_{22} & & & & & \cdot \\ \cdot & & \cdot & & & & \cdot \\ \cdot & & & \cdot & & & \cdot \\ \cdot & & & & \cdot & & \cdot \\ \cdot & & & & & \cdot & \cdot \\ g_{1m} & \cdot & \cdot & \cdot & \cdot & \cdot & g_{mm} \end{pmatrix} \begin{pmatrix} w_1^{(t)} \\ w_2^{(t)} \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ w_m^{(t)} \end{pmatrix}, \tag{1}$$

where $w_i^{(t)}$ is the total population of region i at time t; g_{ij} is the proportion of people who were in region i at time t and in region j at time $t+1$; and g_{ii} is the proportion of people who were in region i at time t and remained there until $t+1$ *plus* the rate of natural increase in region i during the unit time interval $(t, t+1)$.

POPULATION OPERATOR ESTIMATORS

Consider the population growth process, in a two-region system, that occurs over two sequential periods. Assume that an unknown schedule of growth, defined by the matrix G in equation (1), remains unchanged during this period of time. We have then:

$$\begin{bmatrix} w_1^{(t+1)} \\ w_2^{(t+1)} \end{bmatrix} = \begin{bmatrix} g_{11} & g_{21} \\ g_{12} & g_{22} \end{bmatrix} \begin{bmatrix} w_1^{(t)} \\ w_2^{(t)} \end{bmatrix}$$

and

$$\begin{bmatrix} w_1^{(t+2)} \\ w_2^{(t+2)} \end{bmatrix} = \begin{bmatrix} g_{11} & g_{21} \\ g_{12} & g_{22} \end{bmatrix} \begin{bmatrix} w_1^{(t+1)} \\ w_2^{(t+1)} \end{bmatrix}.$$

If the w_i's have been observed and are therefore known, the matrix population growth operator, G, may be derived by solving the following system of four equations in four unknowns:[2]

$$w_1^{(t+1)} = g_{11}w_1^{(t)} + g_{21}w_2^{(t)}$$
$$w_2^{(t+1)} = g_{12}w_1^{(t)} + g_{22}w_2^{(t)}$$
$$w_1^{(t+2)} = g_{11}w_1^{(t+1)} + g_{21}w_2^{(t+1)}$$
$$w_2^{(t+2)} = g_{12}w_1^{(t+1)} + g_{22}w_2^{(t+1)}$$

Thus, for example, given 1950–52 population data for the two-region system of California and the rest of the United States (Table 1), we have:

$$11130 = 10643g_{11} + 141225g_{21}$$
$$142852 = 10643g_{12} + 141225g_{22}$$
$$11638 = 11130g_{11} + 142852g_{21}$$
$$144755 = 11130g_{12} + 142852g_{22}.$$

The more general problem of estimating such a G on the basis of a time series of distributions, however, requires a criterion of fit, since more equations than unknowns are available. Three estimators have been suggested in the literature: the

[2] Assuming, of course, that the following matrix is non-singular:

$$\begin{bmatrix} w_1^{(t)} & w_2^{(t)} \\ w_1^{(t+1)} & w_2^{(t+2)} \end{bmatrix}.$$

Table 1.—ESTIMATED TOTAL POPULATION OF CALIFORNIA AND THE REST OF THE UNITED STATES, 1950–60[a]

(In thousands)

Year (July 1)	California	Rest of the United States	National total
1950	10,643	141,225	151,868
1951	11,130	142,852	153,982
1952	11,638	144,755	156,393
1953	12,101	146,855	158,956
1954	12,517	149,367	161,884
1955	13,004	152,065	165,069
1956	13,581	154,507	168,088
1957	14,177	157,010	171,187
1958	14,741	159,408	174,149
1959	15,288	161,637	176,925
1960	15,863	163,284	179,147

[a] Continental U.S. only; that is, excluding Alaska and Hawaii.

SOURCE: Economic Development Agency, California Statistical Abstract (Sacramento, California: State of California, Documents Section, 1963), p.44.

unrestricted least squares estimator,[3] the minimum absolute deviations estimator,[4] and the restricted least squares estimator.[5] The first adopts the well-known estimation technique traditionally used in regression analysis; the latter two require a mathematical programming formulation. In this paper only the first two estimation techniques are considered. The restricted least squares solution will be described in a forthcoming paper.

THE UNRESTRICTED LEAST SQUARES (ULS) ESTIMATOR

Assume that the observed vectors, w, denoting the population in each of m regions over n years, are generated by the linear statistical model (2):

$$w^{(t+1)} = Gw^{(t)} + \epsilon^{(t)} \tag{2}$$

where $\epsilon^{(t)}$ is a $(mn \times 1)$ vector of error terms. Expressing Equation (2) in the more familiar linear regression form, we have:

$$y = X_g + \epsilon \tag{3}$$

where, for our two-region example,

$$y = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \quad g = \begin{bmatrix} g_1 \\ g_2 \end{bmatrix} \quad \epsilon = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \end{bmatrix}$$

$$X = \begin{bmatrix} X_1 & 0 \\ 0 & X_2 \end{bmatrix} X_1 = X_2$$

or, more specifically,

$$\underset{2n \times 1}{y} = \left[\begin{array}{c} w_1^{(t+1)} \\ w_1^{(t+2)} \\ \cdot \\ \cdot \\ w_1^{(t+n)} \\ \hline w_2^{(t+1)} \\ w_2^{(t+2)} \\ \cdot \\ \cdot \\ w_2^{(t+n)} \end{array}\right] \qquad \underset{2^2 \times 1}{g} = \left[\begin{array}{c} g_{11} \\ g_{21} \\ \hline g_{12} \\ g_{22} \end{array}\right] \qquad \underset{2n \times 1}{\epsilon} = \left[\begin{array}{c} \epsilon_{11} \\ \epsilon_{12} \\ \cdot \\ \cdot \\ \epsilon_{1n} \\ \hline \epsilon_{21} \\ \epsilon_{22} \\ \cdot \\ \cdot \\ \epsilon_{2n} \end{array}\right]$$

and

$$\underset{2n \times 2^2}{X} = \left[\begin{array}{cc|cc} w_1^{(t)} & w_2^{(t)} & 0 & 0 \\ w_1^{(t+1)} & w_2^{(t+1)} & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ w_1^{(t+n-1)} & w_2^{(t+n-1)} & 0 & 0 \\ \hline 0 & 0 & w_1^{(t)} & w_2^{(t)} \\ 0 & 0 & w_1^{(t+1)} & w_2^{(t+1)} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & w_1^{(t+n-1)} & w_2^{(t+n-1)} \end{array}\right].$$

[3] A. Madansky, "Least Squares Estimation in Finite Markov Processes," *Psychometrika*, XXV (1959), 137–44; George A. Miller, "Finite Markov Processes in Psychology," *Psychometrika*, XVII (1952), 149–67; and Lester G. Telser, "Least Squares Estimates of Transition Probabilities," in *Measurement in Economics*, ed. F. Christ (Palo Alto, California: Stanford University Press, 1963), pp. 272–92.

[4] W. D. Fisher, "A Note on Curve Fitting with Minimum Deviations by Linear Programming," *Journal American Statistical Association*, LVI (1961), 359–62; and T. C. Lee, G. G. Judge, and T. Takayama, "On Estimating the Transition Probabilities of a Markov Process," *Journal of Farm Economics*, XLIII (1965), 742–62.

[5] G. G. Judge and T. Takayama, "Inequality Restrictions in Regression Analysis," *Journal American Statistical Association*, LXI (1966), 166–81; and Henry Theil and Guido Rey, "A Quadratic Programming Approach to the Estimation of Transition Probabilities," *Management Science*, XII (1966), 714–21.

ESTIMATOR: $\underline{g} = (X'X)^{-1}X'\underline{y}$

DATA:

$$X = \left[\begin{array}{cc|cc}
10,643 & 141,225 & 0 & 0 \\
11,130 & 142,852 & 0 & 0 \\
11,638 & 144,755 & 0 & 0 \\
12,101 & 146,855 & 0 & 0 \\
12,517 & 149,367 & 0 & 0 \\
13,004 & 152,065 & 0 & 0 \\
13,581 & 154,507 & 0 & 0 \\
14,177 & 157,010 & 0 & 0 \\
14,741 & 159,408 & 0 & 0 \\
15,288 & 161,637 & 0 & 0 \\
\hline
0 & 0 & 10,643 & 141,225 \\
0 & 0 & 11,130 & 142,852 \\
0 & 0 & 11,638 & 144,755 \\
0 & 0 & 12,101 & 146,855 \\
0 & 0 & 12,517 & 149,367 \\
0 & 0 & 13,004 & 152,065 \\
0 & 0 & 13,581 & 154,507 \\
0 & 0 & 14,177 & 157,010 \\
0 & 0 & 14,741 & 159,408 \\
0 & 0 & 15,288 & 161,637
\end{array}\right] \qquad \underline{y} = \left[\begin{array}{c}
11,130 \\ 11,638 \\ 12,101 \\ 12,517 \\ 13,004 \\ 13,581 \\ 14,177 \\ 14,741 \\ 15,288 \\ 15,863 \\
\hline
142,852 \\ 144,755 \\ 146,855 \\ 149,367 \\ 152,065 \\ 154,507 \\ 157,010 \\ 159,408 \\ 161,637 \\ 163,284
\end{array}\right]$$

ESTIMATE:

$$\hat{\underline{g}} = \left[\begin{array}{c} 1.0149 \\ 0.0022 \\ \hline -0.0557 \\ 1.0194 \end{array}\right]$$

FIG. 1.—Unrestricted least squares estimate of the annual migration flow operator for California and the rest of the United States: 1950–60 total population without natural increase time series.

Recalling that the least squares estimator of g is one which minimizes the sum of squared deviations of observed from predicted values, we minimize

$$S = \Sigma(\hat{y}_i - y_i)^2 = [y - Xg]'[y - Xg] \quad (4)$$

with respect to g, and we find that

$$\hat{g} = (X'X)^{-1}X'y\,. \quad (5)$$

Applying Equation (5) to our two-region example in Table 1, we have the estimation process illustrated in Figure 1 and the results presented in Figure 2. Table 2 provides data for comparing the discrepancies between the observed and predicted values.

Results.—The results, presented in Figure 2, are very unrealistic. First, the presence of a negative migration rate in two of the estimated growth operators is unacceptable by definition, since it implies a negative migration flow. Second, in the one case where this particular migration

A. 1950-60 Time Series:

$$\begin{bmatrix} 1.0149 & 0.0022 \\ -0.0557 & 1.0194 \end{bmatrix}$$

$R^2 = 1.0000$

B. 1950-55 Time Series:

$$\begin{bmatrix} 0.9456 & 0.0076 \\ 0.7839 & 0.9522 \end{bmatrix}$$

$R^2 = 1.0000$

C. 1955-60 Time Series

$$\begin{bmatrix} 0.9603 & 0.0072 \\ -0.6169 & 1.0700 \end{bmatrix}$$

$R^2 = 1.0000$

FIG. 2.—Unrestricted least squares estimates: total population.

Table 2.—OBSERVED AND PREDICTED TOTAL POPULATION OF CALIFORNIA AND THE REST OF THE UNITED STATES, 1950-60

(In thousands)

Total population	Year (July 1)									
	1951	1952	1953	1954	1955	1956	1957	1958	1959	1960
California										
Observed[a]	11,130	11,638	12,101	12,517	13,004	13,581	14,177	14,741	15,288	15,863
Predicted[b]										
1950-60	11,110	11,608	12,128	12,602	13,030	13,530	14,121	14,732	15,309	15,809
1950-55	11,139	11,612	12,106	12,560	12,973	...	...	...	...	...
1955-60	...	...	...	...	...	13,587	14,158	14,749	15,308	15,849
Rest of the United States										
Observed[a]	142,852	144,755	146,855	149,367	152,065	154,507	157,010	159,408	161,637	163,284
Predicted[b]										
1950-60	143,367	144,998	146,910	149,025	151,562	154,285	156,743	159,261	161,674	163,916
1950-55	142,819	144,750	146,961	149,323	152,041	...	...	...	...	...
1955-60	...	...	...	...	...	154,682	156,939	159,249	161,467	163,515

[a] See Table 1.

[b] Derived by estimated growth operations in parts A, B, and C of Figure 2, respectively.

rate is nonnegative it is unrealistically large. Third, the diagonal element denoting the growth of California in the absence of in-migration is less than unity on two occasions—a most unlikely event. Thus, it appears that the ULS estimator is not very satisfactory for our purposes. We therefore turn next to a weighted estimator in order to explore whether the peculiar results are in any way related to the great difference in the size of the populations of our two regions.

THE WEIGHTED UNRESTRICTED LEAST SQUARES (WULS) ESTIMATOR

The unweighted ULS estimator assumes that

$$E(\epsilon\epsilon') = \sigma^2 I \; ; \tag{6}$$

that is, that the error terms are independently distributed with a constant variance. Both assumptions are likely to be violated in our estimation problem. Because of the lagged variable relationship in our model, a dependence exists between the error terms and the explanatory variables. And since the populations of the regions differ considerably in size, it is unlikely that the condition of homoscedasticity is present. One might reasonably expect, for example, the variance of the error term to be related to the size of the population being estimated.

The problem of lagged variables is a thorny one, and its resolution in our problem context is beyond the scope of this paper. We must bear in mind, therefore, that our least squares estimates are likely to be biased. The presence of heteroscedasticity, however, may be taken into account by a suitable transformation of the data and, therefore, will be briefly explored here.

Assume, for example, that the standard deviation of the error term in the estimating equation for $w_i^{(t+1)}$ is directly proportional to the size of $w_i^{(t)}$. Then in our model

$$E(\epsilon\epsilon') = \sigma^2 W^2 \tag{7}$$

where W is given in equation 8.

If now we premultiply our original model by W^{-1}, we have

$$W^{-1}y = W^{-1}Xg + W^{-1}\epsilon \tag{9}$$

or

$$y^* = X^*g + \epsilon^* \tag{10}$$

where $y^* = W^{-1}y$, $X^* = W^{-1}X$, and $\epsilon^* = W^{-1}\epsilon$. The variance-covariance matrix for the transformed vector of error terms, ϵ^*, is

$$\begin{aligned} E(\epsilon^*\epsilon^{*\prime}) &= E(W^{-1}\epsilon\epsilon' W^{-1}) \\ &= W^{-1}E(\epsilon\epsilon')W^{-1} \\ &= W^{-1}\sigma^2 W^2 W^{-1} \\ &= \sigma^2 I \, . \end{aligned} \tag{11}$$

Hence we may apply the ULS estimator directly to the transformed model in Equation (10). Table 3 outlines the computational sequence for transforming the data, Figure 3 illustrates the weighted estimation procedure, and Figure 4 presents the three estimated population growth operators.

$$W = \left[\begin{array}{ccccc|ccccc} w_1^{(t)} & \cdot & \cdots & \cdot & 0 & 0 & \cdot & \cdots & \cdot & 0 \\ \vdots & w_1^{(t+1)} & & & \vdots & \vdots & \ddots & & & \vdots \\ \vdots & & \ddots & & \vdots & \vdots & & \ddots & & \vdots \\ 0 & \cdot & \cdots & \cdot & w_1^{(t+n-1)} & 0 & \cdot & \cdots & \cdot & 0 \\ \hline 0 & \cdot & \cdots & \cdot & 0 & w_2^{(t)} & \cdot & \cdots & \cdot & 0 \\ \vdots & \ddots & & & \vdots & \vdots & w_2^{(t+1)} & & & \vdots \\ \vdots & & \ddots & & \vdots & \vdots & & \ddots & & \vdots \\ 0 & \cdot & \cdots & \cdot & 0 & 0 & \cdot & \cdots & \cdot & w_2^{(t+n-1)} \end{array}\right] . \tag{8}$$

Table 3.—WEIGHTED TOTAL POPULATION OF CALIFORNIA AND THE REST OF THE UNITED STATES, 1950–60
(In thousands)

Year	Unweighted populations				Weights		Weighted populations					
	Cal.	R.U.S.	Cal.	R.U.S.	Cal.	R.U.S.	Cal.	R.U.S.	Cal.	R.U.S.	Cal.	R.U.S.
	$\underline{w}_1$	$\underline{w}_2$	$\underline{y}_1$	$\underline{y}_2$			$\underline{w}_{11}$	$\underline{w}_{21}$	$\underline{w}_{12}$	$\underline{w}_{22}$	$\underline{y}_1$	$\underline{y}_2$
t	$w_1^{(t)}$	$w_2^{(t)}$	$w_1^{(t+1)}$	$w_2^{(t+1)}$	$w_1^{(t)}$	$w_2^{(t)}$	$w_1^{(t)}/w_1^{(t)}$	$w_2^{(t)}/w_1^{(t)}$	$w_1^{(t)}/w_2^{(t)}$	$w_2^{(t)}/w_2^{(t)}$	$w_1^{(t+1)}/w_1^{(t)}$	$w_2^{(t+1)}/w_2^{(t)}$
1950.....	10,643	141,225	11,130	142,852	10,643	141,225	1	13.2693	.0754	1	1.0458	1.0115
1951.....	11,130	142,852	11,638	144,755	11,130	142,852	1	12.8349	.0779	1	1.0456	1.0133
1952.....	11,638	144,755	12,101	146,855	11,638	144,755	1	12.4381	.0804	1	1.0398	1.0145
1953.....	12,101	146,855	12,517	149,367	12,101	146,855	1	12.1358	.0824	1	1.0344	1.0171
1954.....	12,517	149,367	13,004	152,065	12,517	149,367	1	11.9331	.0838	1	1.0389	1.0181
1955.....	13,004	152,065	13,581	154,507	13,004	152,065	1	11.6937	.0855	1	1.0444	1.0161
1956.....	13,581	154,507	14,177	157,010	13,581	154,507	1	11.3767	.0879	1	1.0439	1.0162
1957.....	14,177	157,010	14,741	159,408	14,177	157,010	1	11.0750	.0903	1	1.0398	1.0153
1958.....	14,741	159,408	15,288	161,637	14,741	159,408	1	10.8139	.0925	1	1.0371	1.0140
1959.....	15,288	161,637	15,863	163,284	15,288	161,637	1	10.5728	.0946	1	1.0376	1.0102

ESTIMATOR: $\underline{g} = (X'X)^{-1}X'\underline{y}$

DATA:

$$X = \left[\begin{array}{cc|cc}
1 & 13.2693 & 0 & 0 \\
1 & 12.8349 & 0 & 0 \\
1 & 12.4381 & 0 & 0 \\
1 & 12.1358 & 0 & 0 \\
1 & 11.9331 & 0 & 0 \\
1 & 11.6937 & 0 & 0 \\
1 & 11.3767 & 0 & 0 \\
1 & 11.0750 & 0 & 0 \\
1 & 10.8139 & 0 & 0 \\
1 & 10.5728 & 0 & 0 \\
\hline
0 & 0 & 0.0754 & 1 \\
0 & 0 & 0.0779 & 1 \\
0 & 0 & 0.0804 & 1 \\
0 & 0 & 0.0824 & 1 \\
0 & 0 & 0.0838 & 1 \\
0 & 0 & 0.0855 & 1 \\
0 & 0 & 0.0879 & 1 \\
0 & 0 & 0.0903 & 1 \\
0 & 0 & 0.0925 & 1 \\
0 & 0 & 0.0946 & 1
\end{array}\right] \qquad \underline{y} = \left[\begin{array}{c}
1.0458 \\ 1.0456 \\ 1.0398 \\ 1.0344 \\ 1.0389 \\ 1.0444 \\ 1.0439 \\ 1.0398 \\ 1.0371 \\ 1.0376 \\
\hline
1.0115 \\ 1.0133 \\ 1.0145 \\ 1.0171 \\ 1.0181 \\ 1.0161 \\ 1.0162 \\ 1.0153 \\ 1.0140 \\ 1.0102
\end{array}\right]$$

ESTIMATE:

$$\hat{\underline{g}} = \left[\begin{array}{c} 1.0138 \\ 0.0023 \\ \hline -0.0249 \\ 1.0167 \end{array}\right]$$

FIG. 3.—Unrestricted least squares estimate of the annual population growth operator for California and the rest of the United States: 1950–60 weighted total population time series.

A. 1950–60 Time Series:

$$\begin{bmatrix} 1.0138 & 0.0023 \\ -0.0249 & 1.0167 \end{bmatrix}$$

$R^2 = 1.0000$

B. 1950–55 Time Series:

$$\begin{bmatrix} 0.9446 & 0.0077 \\ 0.7899 & 0.9517 \end{bmatrix}$$

$R^2 = 1.0000$

C. 1955–60 Time Series:

$$\begin{bmatrix} 0.9593 & 0.0073 \\ -0.6042 & 1.0688 \end{bmatrix}$$

$R^2 = 1.0000$

FIG. 4.—Unrestricted least squares estimates: weighted total population.

Results.—The weighted ULS estimates in Figure 4 do not appreciably differ from the unweighted ULS estimates in Figure 2. Thus, it appears that weighting does not significantly alter the estimation results. We therefore turn to a different criterion of fit: one which gives less weight to observations with large deviations from the mean.

THE MINIMUM ABSOLUTE DEVIATIONS (MAD) ESTIMATOR

The minimum absolute deviations estimator finds the vector of parameter estimates which minimizes the sum of the *absolute values* of the deviations between observed and predicted values. That is, it derives the vector g which minimizes

$$S = \Sigma |y_i - \hat{y}_i| = |y - Xg|'1, \quad (12)$$

subject to the constraints

$$y = X_g + \epsilon \quad (13)$$

$$g \geqq 0 . \quad (14)$$

To transform equation (12) into a form that is solvable by linear programming methods, we express each ϵ_i as the difference of two nonnegative variables u_i and v_i. Thus, we now seek the vector g which minimizes

$$S = [u + v]'l\,, \tag{15}$$

subject to the constraints

$$y = Xg + \epsilon = Xg + [I, -I]\begin{bmatrix} u \\ v \end{bmatrix} \tag{16}$$

$$g, u, v \geqq 0\,. \tag{17}$$

The problem can be solved by using the simplex algorithm on the tableau presented in Table 4. Since, in equation (16), the vectors u and v are linearly dependent, elements of one or the other, but not both, will enter the optimal solution.[6]

A simplex tableau made up of the population data in Table 1 appears in Figure 5. Figure 6 presents the three population growth operators that result from the application of the unweighted and weighted estimators, respectively.

Results.—The MAD estimators appear to resolve one of the problems presented by the ULS and WULS estimators;

[6] If a particular u_i and v_i were both strictly positive, with $u_i > v_i$ say, the solution could not be optimal since it would be possible to find a smaller value for S in equation (15) by replacing u_i by $u_i - v_i$ and setting the original v_i equal to zero.

Table 4.—SIMPLEX TABLEAU FOR FINDING THE MINIMUM ABSOLUTE DEVELOPING ESTIMATE

	$\underline{0}'$	$\underline{0}'$	...	$\underline{0}'$	$\underline{1}'$	$\underline{1}'$	...	$\underline{1}'$	$\underline{1}'$	$\underline{1}'$	...	$\underline{1}'$
B	$\underline{g}_1'$	$\underline{g}_2'$	...	$\underline{g}_r'$	$\underline{u}_1'$	$\underline{u}_2'$	...	$\underline{u}_r'$	$\underline{v}_1'$	$\underline{v}_2'$	...	$\underline{v}_r'$
$\underline{y}_1$........	X_1				I_n				$-I_n$			
$\underline{y}_2$........		X_2				I_n				$-I_n$		
...			...				...				...	
$\underline{y}_r$........				X_r				I_n				$-I_n$

	0	0	0	0	1	1	....	1	1	....
B	g_{11}	g_{21}	g_{12}	g_{22}	u_{11}	u_{12}	....	v_{11}	v_{12}	....
11,130	10,643	141,225			1			-1		
11,638	11,130	142,852				1			-1	
12,101	11,638	144,755								
12,517	12,101	146,855								
13,004	12,517	149,367								
13,581	13,004	152,065								
14,177	13,581	154,507								
14,741	14,177	157,010								
15,288	14,741	159,408								
15,863	15,288	161,637								
142,852			10,643	141,225						
144,755			11,130	142,852						
146,855			11,638	144,755						
149,367			12,101	146,855						
152,065			12,517	149,367						
154,507			13,004	152,065						
157,010			13,581	154,507						
159,408			14,177	157,010						
161,637			14,741	159,408			1			-1
163,284			15,288	161,637			1			-1

FIG. 5.—Simplex tableau for minimum deviations estimate of the annual population growth operator for California and the rest of the United States: 1950–60 total population time series.

namely, that of the possible entry of negative elements into the growth operator. However, the MAD estimator does not seem to provide improved results in two important respects.

First, although the MAD estimator is sure to produce a non-negative population growth operator, experiments conducted in the course of preparing this paper suggest that what appears as a negative element in the ULS and WULS estimates often merely assumes a zero value in the MAD estimate. Zero migration, though a more reasonable result than negative migration, is, nevertheless, still a rather improbable event. Second, the elements of the population growth matrix occasionally take on very unlikely dimensions. The diagonal elements frequently are less than unity, in particular, and the off-diagonal entries, in order to compensate for this, become excessively large.

Now, we shall analyze the causes of such behavior and find a means for "smoothing" the data, prior to the estimation process, in order to obtain more reasonable estimates. To do this, we draw on the frequently used result in regression analysis which establishes that any multiple linear regression problem can be decomposed and solved as a series of simple linear regressions.[7] Such a technique forms the basis of most stepwise regression procedures.

ESTIMATION WITH SMOOTHED DATA

THE POPULATION MODEL AS A SERIES OF SIMPLE LINEAR REGRESSIONS

Consider the fitted regression equation,

$$Y_i = b_0 + b_1 X_{1i} + b_2 X_{2i} \qquad (18)$$

[7] For a lucid exposition see N. R. Draper and H. Smith, *Applied Regression Analysis* (New York: John Wiley, 1966), 107–15.

A. 1950–60 Time Series:

Unweighted

$$\begin{bmatrix} 1.0057 & 0.0030 \\ 0.0774 & 1.0083 \end{bmatrix}$$

Z = 0.3123

Weighted

$$\begin{bmatrix} 1.0054 & 0.0030 \\ 0.0808 & 1.0080 \end{bmatrix}$$

Z = 0.0431

B. 1950–55 Time Series:

Unweighted

$$\begin{bmatrix} 0.9504 & 0.0072 \\ 0.7753 & 0.9531 \end{bmatrix}$$

Z = 0.0284

Weighted

$$\begin{bmatrix} 0.9500 & 0.0072 \\ 0.7857 & 0.9523 \end{bmatrix}$$

Z = 0.0101

C. 1955–60 Time Series:

Unweighted

$$\begin{bmatrix} 0.9577 & 0.0074 \\ 0.0 & 1.0153 \end{bmatrix}$$

Z = 0.1350

Weighted

$$\begin{bmatrix} 0.9575 & 0.0074 \\ 0.0 & 1.0153 \end{bmatrix}$$

Z = 0.0122

FIG. 6.—Minimum absolute deviations estimates: total population and weighted total population

We can obtain the vector of coefficients, b, simultaneously, by using the matrix least squares estimator,

$$b = (X'X)^{-1}X'y\,. \qquad (19)$$

Alternatively, we may adopt the following stepwise procedure to obtain the same result. (*a*) Regress Y and X_1 and compute the residuals $Y_i - \hat{Y}_i$ where $\hat{Y}_i = a_0 + a_1X_{1i}$. (*b*) Regress X_2 on X_1 and compute the residuals $X_{2i} - \hat{X}_{2i}$ where $\hat{X}_{2i} = c_0 + c_1X_{1i}$. (*c*) Regress the residuals $Y_i - \hat{Y}_i$ on $X_{2i} - \hat{X}_{2i}$ by fitting the model,

$$(Y_i - \hat{Y}_i) = \delta(X_{2i} - \hat{X}_{2i}) + \epsilon_i\,. \qquad (20)$$

No intercept term is required, since we are dealing with two sets of residuals whose sums are zero.

If we now consider the fitted equation,

$$(Y_i - \hat{Y}_i) = d(X_{2i} - \hat{X}_{2i})\,, \qquad (21)$$

and denote $\hat{Y}$ and $\hat{X}_2$ as functions of X_1, we have

$$[Y_i - (a_0 + a_1X_1)] = d[X_{2i} - (c_0 + c_1X_{1i})]\,, \qquad (22)$$

or, collecting terms and attaching a caret to Y to identify it as a fitted value,

$$\begin{aligned}\hat{Y}_i &= (a_0 - dc_0) + (a_1 - dc_1)X_{1i} + dX_{2i} \\ &= b_0 + b_1X_{1i} + b_2X_{2i}\,.\end{aligned} \qquad (23)$$

Applying the above decomposition procedure to our population model, we find that

$$w_1^{(t+1)} = (a_1 - g_{21}c_1)w_1^{(t)} + g_{21}w_2^{(t)}\,, \qquad (24)$$

and

$$w_2^{(t+1)} = (a_2 - g_{12}c_2)w_2^{(t)} + g_{12}w_1^{(t)}\,. \qquad (25)$$

Since in our population model the linear equation is forced through the origin, all intercept terms are zero.

Figure 7 illustrates the decomposition for the 1955–60 population growth operator that appears in part C of Figure 2. Using the 1955–60 time series, we obtain[8]

$$\begin{aligned}w_1^{(t+1)} &= [1.04022 - 0.00725(11.06124)]w_1^{(t)} + 0.00725w_2^{(t)} \\ &= .9603w_1^{(t)} + 0.0073w_2^{(t)}\,,\end{aligned}$$

and

$$\begin{aligned}w_2^{(t+1)} &= [1.01426 + 0.61726\ (0.09029)]w_2^{(t)} - 0.6173w_1^{(t)} \\ &= 1.0700w_2^{(t)} - 0.6173w_1^{(t)}\,.\end{aligned}$$

ANALYSIS OF THE NEGATIVE MIGRATION RATE

Consider the negative migration rate $g_{12} = -0.6169$. The least squares estimator for the slope of a line going through the origin of an (X, Y) coordinate system is

$$b = \frac{\Sigma X_i Y_i}{\Sigma X_i^2}\,. \qquad (26)$$

Hence, the least squares estimator for g_{12} is

$$g_{12} = \frac{\Sigma[(w_1^{(t)} - c_2w^{(2)})(w_2^{(t+1)} - a_2w_2^{(t)})]}{\Sigma[w_1^{(t)} - c_2w_2^{(t)}]^2}\,. \qquad (27)$$

Since the denominator is never negative, g_{12} is negative only when the numerator in equation (27) is negative.[9] This will occur when there is a preponderance of observations for which either,

$$w_2^{(t+1)} < a_2w_2^{(t)} \quad \text{and} \quad w_1^{(t)} > c_2w_2^{(t)}\,, \qquad (28)$$

or

$$w_2^{(t+1)} > a_2w_2^{(t)} \quad \text{and} \quad w_1^{(t)} < c_2w_2^{(t)}\,. \qquad (29)$$

The conditions which will lead to a positive g_{ij} may be illustrated graphically. Consider g_{12} once again, for example. For g_{12} to be positive we must have either the situation illustrated by part A or part B of Figure 8. That is, the data points in the

[8] Discrepancies are owing to rounding.

[9] Notice that g_{12} is zero if $w_2^{(t+1)} = a_2w_2^{(t)}$, and is undefined if $w_1^{(t)} = c_2w_2^{(t)}$.

$(w_2^{(t+1)}, w_2^{(t)})$ coordinate system must follow the same pattern as those in the $(w_1^{(t)}, w_2^{(t)})$ coordinate system. Either both must exhibit a declining rate of growth or both must follow an increasing rate of growth (for example, the broken lines in Figure 8). This will ensure that the conditions specified by equations (28) and (29) are not satisfied and, hence, that g_{12} is positive.

The necessary conditions for diagonal elements to be greater than unity may be established in an analogous manner. However, the problem is a bit more difficult and is beyond the scope of this paper.

SMOOTHING THE DATA

The following smoothing operation is suggested by the conditions illustrated in Figure 8. We begin by fitting quadratic functions (forced through the origin) to the data. We have, then, the following four equations.

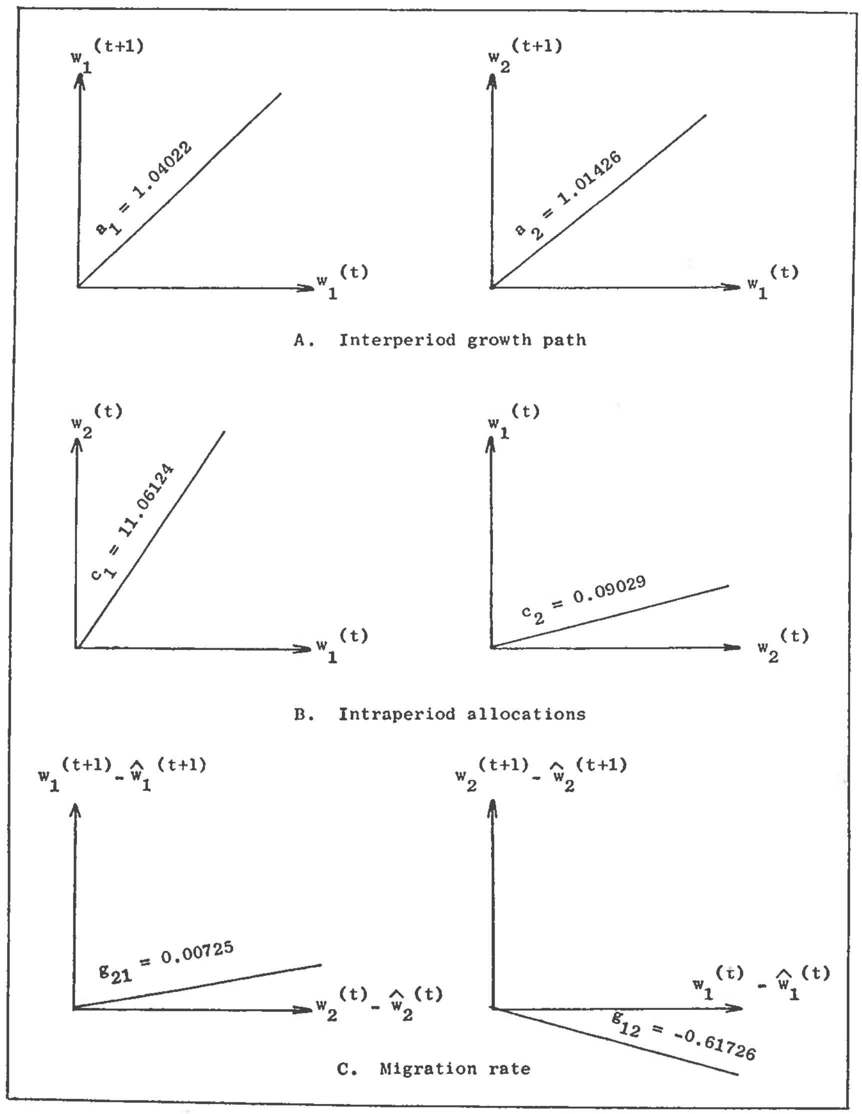

FIG. 7.—The population model as a series of simple linear regressions

$$w_1^{(t+1)} = \alpha_1 w_1^{(t)} + \beta_1 [w_1^{(t)}]^2 \,, \quad (30)$$

$$w_2^{(t)} = \sigma_2 w_1^{(t)} + \beta_2 [w_1^{(t)}]^2 \,, \quad (31)$$

$$w_2^{(t+1)} = \alpha_3 w_2^{(t)} + \beta_3 [w_2^{(t)}]^2 \,, \quad (32)$$

$$w_1^{(t)} = \alpha_4 w_2^{(t)} + \beta_4 [w_2^{(t)}]^2 \,. \quad (33)$$

Next, we examine the fitted equations to determine whether equations (30) and (31) have the same form and then repeat the examination for equations (32) and (33). Returning to our analysis of the 1955–60 time series, for example, we find,

$$w_1^{(t+1)} = 1.09029 w_1^{(t)} - 0.0000035 [w_1^{(t)}]^2 \quad (34)$$

$$w_2^{(t)} = 18.01309 w_1^{(t)} - 0.0004878 [w_1^{(t)}]^2 \quad (35)$$

$$w_2^{(t+1)} = 1.0608 w_2^{(t)} - 0.0000006 [w_2^{(t)}]^2 \quad (36)$$

$$w_1^{(t)} = -0.05800 w_2^{(t)} + 0.0000009 [w_2^{(t)}]^2 \quad (37)$$

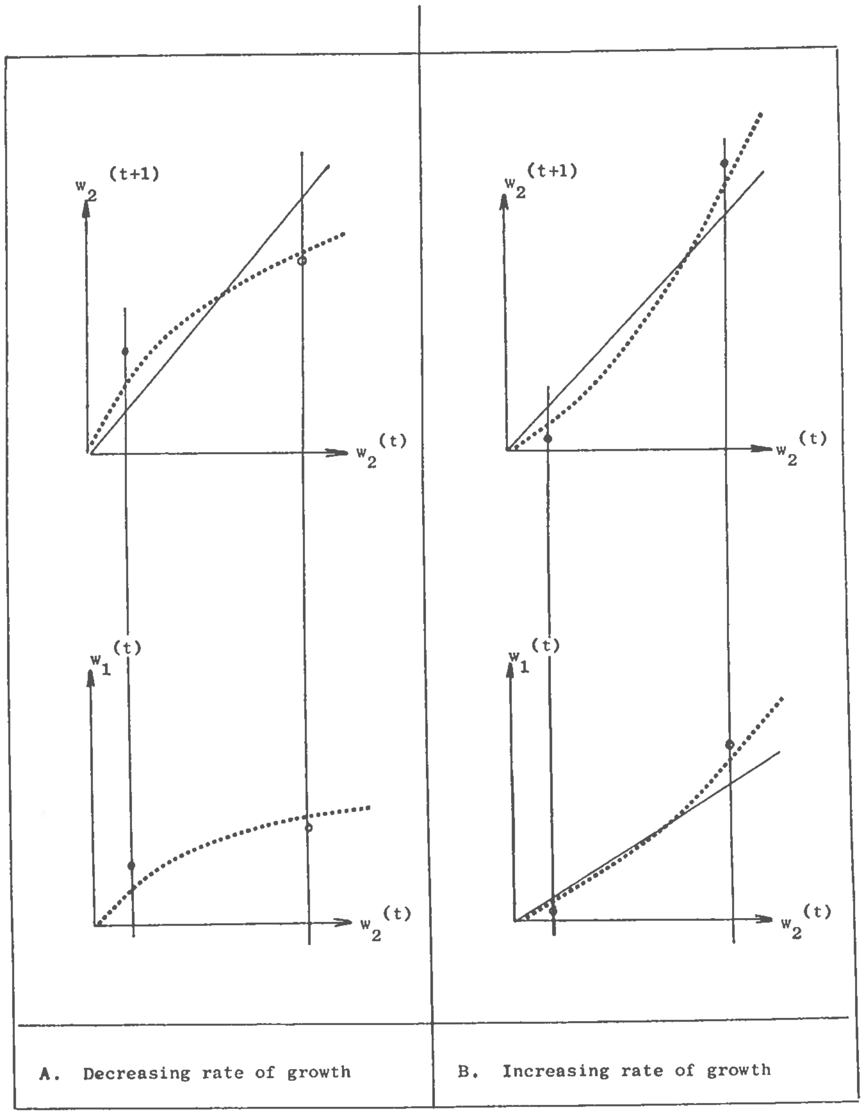

FIG. 8.—Conditions for a positive migration rate from Region 1 to Region 2

Clearly equations (34) and (35) are of the same form. Equations (36) and (37), however, are not. This is to be expected since g_{12} was negative. We select equation (36) for smoothing and transform it from a decreasing rate-of-growth function to an increasing rate-of-growth function by ensuring that both coefficients in equation (33) remain strictly positive. Our procedure from here on is quite arbitrary and is only meant to be illustrative.

We select for the coefficient α_3, in equation (32), the lowest $w^{(t+1)}/w^{(t)}$ ratio in our time series. Thus we have

$$\alpha_3 = \frac{163284}{161637} = 1.01019 .$$

To derive β_3, we take the first two data points and solve for the β_3 that will reproduce the second data point when the first is entered into equation (32); that is,

$$154507 = 1.01019[152065] + \beta_3[152065]^2 ,$$

and

$$\beta_3 = 0.00000004 .$$

Thus we replace equation (36) with the "smoothed" fitted equation (38):

$$w_2^{(t+1)} = 1.01019 w_2^{(t)} + 0.00000004[w_2^{(t)}]^2 , \qquad (38)$$

and derive the smoothed time series for the rest of the United States that appears in column A of Table 5. The series presented in columns B and C of Table 5 result from setting $\beta_3 = 0.00000002$ and $\beta_3 = 0.00000003$, respectively.

Results.—Figure 9 presents the ULS and MAD estimates of the 1955–60 population growth operator that correspond to the three alternate smoothed-data series

A. Series A

ULS Estimate

$$\begin{bmatrix} 0.9573 & 0.0075 \\ 0.0450 & 1.0123 \end{bmatrix}$$

$R^2 = 1.0000$

MAD Estimate

$$\begin{bmatrix} 0.9575 & 0.0074 \\ 0.0431 & 1.0124 \end{bmatrix}$$

$Z = 0.0062$

B. Series B

ULS Estimate

$$\begin{bmatrix} 0.9663 & 0.0067 \\ 0.0170 & 1.0118 \end{bmatrix}$$

$R^2 = 1.0000$

MAD Estimate

$$\begin{bmatrix} 0.9656 & 0.0067 \\ 0.0167 & 1.0118 \end{bmatrix}$$

$Z = 0.0060$

C. Series C

ULS Estimate

$$\begin{bmatrix} 0.9618 & 0.0071 \\ 0.0298 & 1.0122 \end{bmatrix}$$

$R^2 = 1.0000$

MAD Estimate

$$\begin{bmatrix} 0.9616 & 0.0071 \\ 0.0286 & 1.0123 \end{bmatrix}$$

$Z = 0.0061$

FIG. 9.—Unrestricted least squares and minimum absolute deviations estimates: smoothed data, 1955–60 time series.

in Table 5. A glance reveals the sensitivity of g_{12} to a unit change in the seventh decimal place in the value of β_3. This underscores the importance of establishing a more rational method for smoothing the data points than is presented here.

MIGRATION OPERATOR ESTIMATORS

Up to this point we have assumed that, for our interregional system, data on births and deaths are totally unavailable. This is a level of data availability that exists in many less-developed parts of the world. However, it hardly represents conditions in the United States. Data on births and deaths are regularly collected for most subdivisions of the nation down to the county level. Where such information exists, it should be incorporated into the estimation process and used to reduce

Table 5.—ESTIMATED TOTAL POPULATION OF CALIFORNIA AND SMOOTHED TOTAL POPULATION OF THE REST OF THE UNITED STATES, 1955–60

(In thousands)

Year (July 1)	California	Rest of the United States[a]			
		Observed	A	B	C
1955.................	13,004	152,065	152,065	152,065	152,065
1956.................	13,581	154,507	154,516	154,077	154,308
1957.................	14,177	157,010	157,022	156,122	156,595
1958.................	14,741	159,408	159,584	158,200	158,926
1959.................	15,288	161,637	162,203	160,313	161,303
1960.................	15,863	163,284	164,882	162,461	163,727

[a] Series A: β_3 = 0.00000004

Series B: β_3 = 0.00000002

Series C: β_3 = 0.00000003

Table 6.—NATURAL POPULATION INCREASE IN CALIFORNIA AND THE REST OF THE UNITED STATES, 1950–60

(In thousands)

Interval (July 1)	National total	California	Rest of the United States
1950-51..............	2,287	151	2,136
1951-52..............	2,349	165	2,184
1952-53..............	2,421	180	2,241
1953-54..............	2,558	192	2,366
1954-55..............	2,614	198	2,416
1955-56..............	2,597	207	2,390
1956-57..............	2,730	220	2,510
1957-58..............	2,630	225	2,405
1958-59..............	2,651	227	2,424
1959-60..............	2,583	234	2,349

SOURCE: Financial and Population Research Section, California Population, 1964 (Sacramento, California: Department of Finance, 1964), p.9 and U.S. Department of Commerce, Bureau of the Census, Current Population Reports, Series P-25, (July 1966). (January 1 data were linearly interpolated to provide July 1 data.)

ESTIMATOR: $\underline{g} = (X'X)^{-1}X'\underline{y}$

DATA:

$$X = \left[\begin{array}{cc|cc}
10{,}643 & 141{,}225 & 0 & 0 \\
11{,}130 & 142{,}852 & 0 & 0 \\
11{,}638 & 144{,}755 & 0 & 0 \\
12{,}101 & 146{,}855 & 0 & 0 \\
12{,}517 & 149{,}367 & 0 & 0 \\
13{,}004 & 152{,}065 & 0 & 0 \\
13{,}581 & 154{,}507 & 0 & 0 \\
14{,}177 & 157{,}010 & 0 & 0 \\
14{,}741 & 159{,}408 & 0 & 0 \\
15{,}288 & 161{,}637 & 0 & 0 \\
\hline
0 & 0 & 10{,}643 & 141{,}225 \\
0 & 0 & 11{,}130 & 142{,}852 \\
0 & 0 & 11{,}638 & 144{,}755 \\
0 & 0 & 12{,}101 & 146{,}855 \\
0 & 0 & 12{,}517 & 149{,}367 \\
0 & 0 & 13{,}004 & 152{,}065 \\
0 & 0 & 13{,}581 & 154{,}507 \\
0 & 0 & 14{,}177 & 157{,}010 \\
0 & 0 & 14{,}741 & 159{,}408 \\
0 & 0 & 15{,}288 & 161{,}637
\end{array}\right] \qquad
\underline{y} = \left[\begin{array}{c}
10{,}991 \\ 11{,}468 \\ 11{,}910 \\ 12{,}296 \\ 12{,}761 \\ 13{,}323 \\ 13{,}926 \\ 14{,}488 \\ 15{,}050 \\ 15{,}571 \\
\hline
140{,}876 \\ 142{,}514 \\ 144{,}483 \\ 146{,}660 \\ 149{,}123 \\ 151{,}540 \\ 154{,}162 \\ 156{,}699 \\ 159{,}099 \\ 161{,}353
\end{array}\right]$$

ESTIMATE:

$$\hat{\underline{g}} = \left[\begin{array}{c} 0.9832 \\ 0.0034 \\ \hline 0.0171 \\ 0.9964 \end{array}\right]$$

FIG. 10.—Unrestricted least squares estimate of the annual population growth operator for California and the rest of the United States: 1950–60 total population time series.

A. 1950-60 Time Series

ULS Estimate

$$\begin{bmatrix} 0.9832 & 0.0034 \\ 0.0171 & 0.9964 \end{bmatrix}$$

$R^2 = 1.0000$

MAD Estimate

$$\begin{bmatrix} 0.9689 & 0.0048 \\ 0.0311 & 0.9952 \end{bmatrix}$$

$Z = 0.0852$

B. 1950-55 Time Series

ULS Estimate

$$\begin{bmatrix} 0.8711 & 0.0122 \\ 0.1295 & 0.9877 \end{bmatrix}$$

$R^2 = 1.0000$

MAD Estimate

$$\begin{bmatrix} 0.8824 & 0.0113 \\ 0.1176 & 0.9887 \end{bmatrix}$$

$Z = 0.0196$

C. 1955-60 Time Series

ULS Estimate

$$\begin{bmatrix} 0.9562 & 0.0059 \\ 0.1614 & 0.9832 \end{bmatrix}$$

$R^2 = 1.0000$

MAD Estimate

$$\begin{bmatrix} 0.9288 & 0.0085 \\ 0.0712 & 0.9915 \end{bmatrix}$$

$Z = 0.0243$

FIG. 11.—Unrestricted least squares and minimum absolute deviations estimates: total population without natural increase.

the problem to one of estimating migration only. This removes variation that may be attributable to fluctuation in birth and death rates. Moreover, it allows us to introduce a side constraint in the estimation process; namely, that the rows of the migration operator sum to unity. As a result, only $m^2 - m$ parameters need to be estimated. The diagonal elements of the operator are strictly determined by the side constraint.

To transform the population distributions of our two-region system into a form suitable for estimating the migration operator, we need to remove the effects of natural increase from the data. This can be done by subtracting the time series data on natural increase presented in Table 6 from the population data appearing in Table 1 and then normalizing the results so that the national population after the application of the migration operator (that is, the elements of the vector y) remains equal to the national population as represented by the operand (that is, the row elements of X). We thereby ensure that during the migration operation the population remains constant. The results of such a data transformation for our two-region system appear in Table 7.

The ULS *and* MAD *estimators.*—Figure 10 illustrates the ULS estimation process for the migration operator. Figure 11 presents the three migration operators as estimated by the ULS and MAD estimators, respectively.

Results.—The results are considerably improved. The problem of negative elements does not seem to arise and the off-diagonal elements, although still much too large in two instances, now take on more plausible values. Finally, the problem of diagonal elements being less than unity, of course, no longer is present.

CONCLUSIONS

In this paper we have demonstrated that it is possible to estimate the structure of interregional population growth and migration solely on the basis of historical data on interregional population distributions. The application of the method to data generated by a changing growth operator, however, provides unsatisfastory results. Recourse to data "smoothing," therefore, appears to be necessary in instances where the operator undergoes considerable changes over time. Furthermore, data permitting, estimation of migration operators rather than population growth operators seems to lead to improved results.

Table 7.—TOTAL POPULATION WITHOUT NATURAL INCREASE FOR CALIFORNIA AND THE REST OF THE UNITED STATES, 1950–60[a]

(In thousands)

Year (July 1)	California	Rest of the United States
1951	10,991	140,876
1952	11,468	142,514
1953	11,910	144,483
1954	12,296	146,660
1955	12,761	149,123
1956	13,323	151,540
1957	13,926	154,162
1958	14,488	156,699
1959	15,050	159,099
1960	15,571	161,353

[a] Data were derived by subtracting the natural increases in Table 6 from the populations in Table 1 and then normalizing the results so that the national population at time $t + 1$ added up to the national population at time t; for example, for 1951 we have 10,979 + 140,716 = 151,695 and, since the national population in 1950 was 151,868, we normalize to find 10,991 + 140,876 = 151,868.