A STATISTICAL APPROACH TO THE PROBLEM OF ESTIMATING THE NUMBER OF INTERNAL NET MIGRANTS AND THE INTERNAL NET MIGRATION RATES BY CENSUS SURVIVAL RATE METHOD*

G. K. Kripalani**, Western Michigan University

1. The problem is to estimate for a color-sex-age group: (a) the number of net migrants for state i and (b) the net migration rate for state i subject to the following structural elements of the model:

(i) There are census enumeration errors.

(ii) U. S. population is subject to and is acted upon by only one cause of change viz. mortality.

(iii) State i population is subject to and is simultaneously acted upon by two causes of change viz. mortality and net migration (in- or out-).

2. Notation:

A = Number of persons in a color-sex-age (CSA) category, all aged "x" in U.S. as enumerated at the census at the beginning of a decade $t = 0$.

B = Number of persons in the same color-sex (CS) category, all aged "x + 10" in U.S. as enumerated at the census at the end of the decade $t = 1$.

A_i = Numbers of persons in the same CSA category, all aged "x" in state i as enumerated at the census at the beginning of the decade $t = 0$.

B_i = Number in the same CS category, all aged "x + 10" in state i as enumerated at the census at the end of the decade $t = 1$.

$R = B/A$ the census survival ratio for U.S. for the cohort.

$R_i = B_i/A_i$ the census survival ratio for state i for the same cohort.

A', B', A_i', B_i', denote corresponding 'true' unknown numbers.

3. A number of formulae using census survival ratios are currently in use for estimating internal net migration numbers and rates for individual color-sex-age groups in a state. Three of these formulae known as Forward, Reverse and Average formulae are given by Hamilton and Siegel (1942), and Hamilton (1959). The fourth formula is the one used by Everett S. Lee et al. (1957), in their momentous work, "Population Redistribution and Economic Growth, United States, 1870-1950"; and, the fifth is the formula which takes the numerator of the Forward formula for its numerator and the initial decade population for its denominator. These existing formulae are based on an arithmetical approach and thus provide single point estimates. The alternative method outlined below uses a statistical approach and hence provides an estimate of the standard error of the estimate of net migration rates and numbers. In addition further merits of the proposed method are:

(a) It calculates the net migration rate as an independent rate which excludes the effects of all other operating causes viz. mortality and census enumeration errors.

(b) Uses a procedure that takes into account the errors of differential under-enumeration and keeps such errors out of net migration rates and numbers. (While the estimates of net migration rate given by the three formulae of Siegel and Hamilton are free from the effects of census enumeration errors, their estimates of the number of net migrants are not.)

(c) Takes into account mortality differences between states and the national average.

(d) Provides an estimate of the measure of the impact of census enumeration errors considered as a separate cause acting on the model, besides the causes of net migration and mortality. (See paragraphs 10, 11 and 12)

---

* The major part of this work was done as a part of the author's work on a National Science Foundation Project "Area-Population Adjustments in Relation to Economic Activity" at North Carolina State University during 1963-65. The work was revised subsequently in 1967 when the author worked at the University of Chicago on a Corps of Engineers Project "Employment-Population Effects of Location of Water Resource Development Projects".

** Fellow, Institute of Actuaries, London.

4. The basic structural condition of our model is that a group of lives is being continuously and simultaneously acted upon by two causes of change viz. mortality and net migration, (ignoring, for the time-being the problem of census enumeration errors). The principal objection to all these formulae is that they are derived by a procedure that is valid only if the group of lives under consideration is subject to one cause of change at a time. This procedure is invalid when the group is subject to continuous and simultaneous operation of two or more causes of change. The basic structure of our problem is that we start with an initial group of lives at time $t = 0$ and that this group is being depleted by mortality and by net out-migration (or depleted by mortality and augmented by net in-migration) and that both causes of change operate simultaneously and continuously throughout the census decade up to time $t = 1$ (dropping for the time-being the problem of census enumeration errors). All the three formulae of Siegel and Hamilton assume that only one cause of change operated at a time. For example, the numerator of the Forward Formula represents the number of net migrants only if it is assumed that (a) mortality alone operates as a cause of change throughout the census decade and (b) at the end of the census decade for an instant of time, mortality ceases to operate and only net migration as a cause of change operates. This is the extreme kind of assumption underlying the Forward Formula. By a complete swing of the pendulum, as it were, the Reverse Formula is derived on the assumptions that net migration as a cause of change operates only for a brief instant at the start of the decade; thereafter throughout the decade, mortality alone acts as a cause of change affecting the group. Similarly the Average Formula is valid on the assumptions that (a) mortality alone operates as the cause of change from the start of the decade to the mid-point of the decade from $t = 0$ to $t = 1/2$, (b) precisely at $t = 1/2$ mortality ceases to operate and net migration as a cause of change operates for that instant and (c) mortality again takes over and operates as the sole cause of change from $t = 1/2$ to $t = 1$.

5. Some further comments are necessary on the Average Formula. In this formula, the numerator is simply the average of the numerators of the Forward and Reverse Formulae and so is the denominator. In fact, we could have an infinite number of Average Formulae by taking any weighted average of the numerators of the Forward and Reverse Formulae and corresponding weighted average of the denominators of the two formulae. If W_f is the weight attached to the numerator and the denominator of the Forward Formula and $W_r = 1 - W_f$ is the weight attached to the Reverse Formula, any weighted average formula obtained by the use of these weights will still give the same net migration rate. If the Forward Formula for the net migration rate is N_f/D_f and the Reverse Formula is N_r/D_r, then the generalized weighted Average Formula will be

$$\frac{W_f \cdot N_f + W_r \cdot N_r}{W_f \cdot D_f + W_r \cdot D_r} \quad \ldots\ldots\ldots\ldots \quad (1.7.1)$$

and will be equal to N_f/D_f or N_r/D_r since we have N_f/D_f equal to N_r/D_r. The generalized formula will hold for all arbitrary values of W_f or W_r lying in the closed interval 0 to 1. Thus, we can use any value for the numerator of the expression for the net migration rate lying between N_f and N_r provided we adjust the denominator accordingly. In fact, one may look at the Forward and Reverse Formulae as the special cases of the generalized formula when $W_f = 1$ and $W_r = 1$ respectively.

6. What passes as three formulae for the net migration rate is really a single formula valid for a distorted structure of the model in which mortality operates as the only cause of decrement throughout the closed interval of time $0 \le t \le 1$ except at a point of time $t = t^{**}$ and in which net migration as a cause affecting the group does not operate at all throughout the decade except at the brief instant of time $t = t^{**}$. As the generalized weighted average formula shows the choice of t^{**} is absolutely arbitrary and does not affect the value of the ratio viz. the net migration rate.

7. To summarize, the main objection to the procedure underlying the derivation of the three formulae of Siegel and Hamilton is that it is valid only when the group of lives is subject to one cause of change at a time and not when the structural situation is one in which the group of lives is continuously and simultaneously acted upon by two causes of change. Secondly, in terms of the structure of the model, neither the numerator nor the denominator of any of the formulae, possesses real interpretation or significance. Thirdly, none of the expressions for the number of net migrants given by Siegel and Hamilton is free from the effect of census enumeration errors. Similar comments apply to formulae (4 and 5).

8. Notation:

Let p^d and p_i^d denote the independent survival rate against mortality for the specified CSA category during the decade, for U. S. and state i respectively. Hence $q^d = 1 - p^d$ and $q_i^d = 1 - p_i^d$ where q^d and q_i^d are independent mortality rates.

Let p_i^w denote the independent survival rate against net migration for state i for the specified CSA category during the decade. Hence $q_i^w = 1 - p_i^w$ where q_i^w is the independent rate of net migration. Taking net in-migration as negative and net out-migration as positive,[1] we have when there is net in-migration $p_i^w > 1$ since q_i^w is negative, and when there is net out-migration $p_i^w < 1$ since q_i^w is positive.

9. It will be assumed that: (a) the U.S. population (numbering A at the start of the decade) is a random sample from an infinitely large population for which probability of dying in the unit time interval (defined as an inter-censal decade) is $Q = 1 - P$ where P is the Life Table (L.T.) 10-year Survival Rate for the particular CSA category

in U. S. population for the relevant intercensal period; and (b) state i population (number A_i at the start of the decade) is a random sample from an infinitely large population for which probability of dying in the unit time interval (defined as an intercensal decade) is $Q_i = 1 - P_i$ where P_i is the Life Table (L.T.) 10-year Survival Rate for the particular CSA category in state i population for the relevant intercensal period.

10. It is important at this stage to comment upon the manner in which the significance of the impact of census enumeration errors is proposed to be viewed. For U. S., $A = (1 - e_o)A'$ and $B = (1 - e_1)B'$ so that $R = R' \times (1 - e_o) / (1 - e_1) = KR'$ where $K = (1 - e_o) / (1 - e_1)$, and e_o and e_1 represent the extent of under-enumeration in U. S. at census at t = 0 and t = 1 respectively. Similarly for state i, $R_i = K_i R_i'$ where $K_i = (1 - e^i_o) / (1 - e^i_1)$ and e^i_o and e^i_1 represent extent of underenumeration in state i at t = 0 and t = 1 respectively.

It may be observed that K (or K_i) is always positive. Whenever there is underenumeration e's are positive fractions $0 \leq e < 1$, so that the factors (1 - e) are always positive. When there is over-enumeration, corresponding e is negative so that $1 - e > 1$ is always positive.

11. Let us define η_t by the relationship

$$K = \exp\left[-\int_0^1 \eta_t \, dt\right] = e^{-\eta} \text{.............(1)}$$

$$\text{where } \eta = \int_0^1 \eta_t \, dt.$$

$$\text{or } \eta = -\log_e K \text{.................(2)}$$

$$\eta \gtreqless 0 \text{ according as } K \lesseqgtr 1 \text{ ....(3)}$$

---

[1]Usually the present formulas treat net out-migration as negative and net in-migration as positive. In this paper, net out-migration has been handled as positive and net in-migration as negative. The reason for doing so is that if net out-migration rate (say, denoted by q) is treated as positive, $p = 1 - q$ acquires a real physical significance, and represents the survival rate against the cause of "net migration" in the same way as when q is the life table mortality rate, $p = 1 - q$ represents the life table survival rate. If net out-migration is treated as negative as is done at present by Hamilton & Siegel and others, $p = 1 - q > 1$ will have no significance whatsoever. Thus, the conventions set up in this paper of treating net out-migration as positive and net in-migration as negative, give real significance to the complementary quantity $p = 1 - q$ and enable the use of binomial distribution when $q > 0$ (net out-migration) and negative binomial when $q < 0$ (net in-migration), $p < 1$ in the case of net out-migration and $p > 1$ in the case of net in-migration.

When $K > 1$, $\eta < 0$ so that $-\eta > 0$. In that case,
let $\delta = -\eta > 0$,
so that $K (>1) = e^{\delta}$ where $\delta > 0$.

When K = 1, the differential errors of underenumeration have no effect on the census survival ratio and the observed CSR is the same as true CSR. When K < 1, the effect of the errors of differential underenumeration is equivalent to the operation of a rate of decrement whose average force of exit over the intercensal decade (interval 0, 1) is $\eta = -\log_e K$. When K > 1, the effect is equivalent to the operation of a rate of increment whose average force of entry, is $\delta = \log_e K$. In general

$$R = KR' = R'e^{-\int_0^1 \eta t \, dt} = R'e^{-\eta} \text{...(4)}$$

Similar remarks apply to K_i for state i. We have

$$\frac{B_i}{A_i} = K_i \frac{B_i'}{A_i'}$$

$$\text{and } K_i = e^{-\int_0^1 \eta_{it} \, dt} = e^{-\eta_i}$$

12. We may thus view the situation for each CSA category in U.S. population as enumerated as a case of a population subject to the operation of two causes, (i) mortality and (ii) a cause of exit (or entry) whose average force of exit (or entry) is η so that the independent 'survival' rate against this cause is $e^{-\eta} = K$. Similarly we will view the situation for each CSA category in a state i population as enumerated as a case of a population subject to the operation of three causes, (i) mortality (ii) net migration (in or out) and (iii) a cause of exit (or entry) whose average force of exit (or entry) is η_i so that the independent 'survival' rate against this cause is $e^{-\eta_i} = K_i$.

13. The proposed method is based on a fundamental theorem in the Theory of Life Contingencies which states that "when a group of lives is simultaneously acted upon by two or more independent causes of change[2], the over-all rate of survival against all causes acting together is equal to the product of the various rates of survival against individual causes acting separately. Thus if a group of lives aged 'x' is subject to the simultaneous operation of three independent causes α, β and γ, then (in actuarial notation):

$$ap_x = p_x^{\alpha} \cdot p_x^{\beta} \cdot p_x^{\gamma}$$

where ap_x denotes the over-all rate of survival against all causes combined and p_x^{α} denotes the independent rate of survival against cause α alone.

14. Assuming that the effect of census enumeration errors on the observed census survival ratio is the same for U. S. as for each state i for a specified CSA category, i.e., $K = K_i$, (This assumption is the same as has been made by Zachariah (1962) and others.) We have:

For U. S. $R = K \cdot p^d$ ....(5)
For state i $R_i = K\, p_i^d\, p_i^w$ ....(6)

on the basis of the fundamental formula

$$ap_x = p_x^\alpha\, p_x^\beta\, p_x^\gamma.$$

15. In terms of the assumptions of para 9., we may regard p^d as a binomial variable with mean P and variance PQ/A[3] where A is the initial population in the specified CSA category in U. S. at the start of the decade. Similarly p_i^d may be regarded as a binomial variable with mean P_i and variance P_iQ_i/A_i

$$p^d \sim \text{Bin}\ (P,\ PQ/A) \ \ldots\ldots (7)$$

$$p^d \sim \text{Bin}\ (P_i,\ P_iQ_i/A_i) \ \ldots\ldots (8)$$

Since A and A_i are generally very large, p^d and p_i^d may be regarded as distributed normally.

16. Let us write:

For U.S.: $R = Kp^d = KP\varepsilon_o$ ....(9)
where ε_o is a random variable.

For state i: $R_i = Kp_i^d\, p_i^w = Kp_i^w\, P_i\, \varepsilon_i$ (10)
where ε_i is a random variable.

From (9) we have

$$p^d = P\,\varepsilon_o$$

$$E\,(p^d) = PE\,(\varepsilon_o)$$

But $E\,(p^d) = P$, so that

$$E\,(\varepsilon_o) = 1 \ \ldots\ldots (11)$$

$$V\,(p^d) = V\,(P\,\varepsilon_o)$$

$$\frac{P(1-P)}{A} = P^2\, V\,(\varepsilon_o)$$

$$\text{or } V\,(\varepsilon_o) = \frac{1-P}{A\,P} \ \ldots\ldots (12)$$

---

[2]The theorem is generally proved in any text book on "Life Contingencies" in relation to causes of decrement, but it can easily be established in the more general case when the causes operating are the causes of change, some or all of which may be of incremental type. The proof is briefly as follows:

$$ap_x = \exp\left[-\int_0^1 \mu_{x+t}\,dt\right]$$

where μ_{x+t} is the over-all force of decrement (or change). $p_x^\alpha = \exp\left[-\int_0^1 \mu_{x+t}\,dt\right]$ etc.

If the forces of decrement (or change) due to causes α, β and γ are independent $\mu_{x+t}^\alpha + \mu_{x+t}^\beta + \mu_{x+t}^\gamma$. Hence the result.

[3]Strictly speaking, the denominators in the expressions for variances are not A or A_i but the true exposed to the risk of death.

Similarly,

$$E\,(\varepsilon_i) = 1 \ \ldots\ldots (13)$$

$$V\,(\varepsilon_i) = \frac{1-P_i}{A_iP_i} \ \ldots\ldots (14)$$

17. Note that p^d and p_i^d are independent and hence ε_o and ε_i are independent. Further the variability of $\varepsilon_o = \frac{\sigma(\varepsilon_o)}{E(\varepsilon_o)} = \sqrt{\frac{1-P}{AP}}$ is generally near to 1 and A is very large. We may therefore, use the following results for the mean value and variance of a ratio:

$$E\left(\frac{X_1}{X_2}\right) = \frac{E(X_1)}{E(X_2)} \ \ldots\ldots (15)$$

$$V\left(\frac{X_1}{X_2}\right) = \frac{\xi_1^2}{\xi_2^2}\left[\frac{\sigma_1^2}{\varepsilon_1^2} + \frac{\sigma_2^2}{\varepsilon_2^2} + \frac{3\sigma_1^2\,\sigma_2^2}{\varepsilon_1^2\,\varepsilon_2^2}\right] \ldots\ldots (16)$$

where $E(X_1) = \xi_1$; $E(X_2) = \xi_2$; $V(X_1) = \sigma_1^2$, $V(X_2) = \sigma_2^2$

and (i) X_1 and X_2 are independent; and

(ii) $\frac{\sigma_2}{\xi_2}$ is small so that its higher than second powers can be ignored. We can therefore take:

$$E\left(\frac{\varepsilon_i}{\varepsilon_o}\right) = \frac{E(\varepsilon_i)}{E(\varepsilon_o)} = 1 \ \ldots\ldots (17)$$

Now $p_i^w = \frac{R_i}{R}\,\frac{P}{P_i}\,\frac{\varepsilon_i}{\varepsilon_o}$

$$E(p_i^w) = \frac{R_i}{R}\,\frac{P}{P_i}\,E\left(\frac{\varepsilon_i}{\varepsilon_o}\right) = \frac{R_i}{R}\,\frac{P}{P_i} \ \ldots\ldots (18)$$

We may take $\frac{R_i}{R}\,\frac{P}{P_i}$ as an unbiased estimate of p_i^w.

$$\hat{p}_i^w = \frac{R_i}{R}\cdot\frac{P}{P_i}$$

$$\hat{q}_i^w = 1 - \hat{p}_i^w = 1 - \frac{R_i}{R}\cdot\frac{P}{P_i} \ \ldots\ldots (19)$$

Estimate of the independent rate of migration is given by $1 - \frac{R_i}{R}\cdot\frac{P}{P_i}$.

18. We have already shown that,

$E(\varepsilon_o) = 1$, $V(\varepsilon_o) = \frac{1-P}{AP}$. Denote by σ^2.

$E(\varepsilon_i) = 1$, $V(\varepsilon_i) = \frac{1-P_i}{A_iP_i}$. Denote by σ_i^2.

As in our problem $\frac{\sigma(\epsilon_o)}{E(\epsilon_o)}$ and $\frac{\sigma(\epsilon_i)}{E(\epsilon_i)}$ are both small, we apply the approximate relationship (16) and omit the last term in the bracket viz.

$$\frac{3\sigma^2\sigma_1^2}{\epsilon_1^2\epsilon_2^2}.$$

We have

$$V\left(\frac{\epsilon_i}{\epsilon_o}\right) = V(\epsilon_o) + V(\epsilon_i) + 3V(\epsilon_o) \cdot V(\epsilon_i). \quad (20)$$

$$= \sigma^2 + \sigma_i^2 \quad \ldots\ldots (21)$$

ignoring the last term.

$$V(p_i^w) = V\left(\frac{R_i}{R}\frac{P}{P_i}\frac{\epsilon_i}{\epsilon_o}\right)$$

$$= \frac{R_i^2}{R^2}\frac{P^2}{P_i^2}\left[\frac{Q}{AP} + \frac{Q_i}{A_iP_i}\right] \quad \ldots\ldots (22)$$

Where $Q - 1 - P$ and $Q_i = 1 - P_i$.

19. $q_i^w = 1 - p_i^w$ and $\hat{q}_i^w = 1 - \hat{p}_i^w$

$\hat{q}_i^w > 0$ when $\hat{p}_i^w > 1$ or there is out-migration.

$\hat{q}_i^w < 0$ when $\hat{p}_i^w > 1$ or there is in-migration.

$$E(q_i^w) = E(1 - p_i^w) = 1 - \frac{R_i}{R}\frac{P}{P_i} \quad \ldots\ldots (23)$$

$$V(q_i^w) = V(1 - p_i^w) = V(p_i^w)$$

$$= \frac{R_i^2P^2}{R^2P_i^2}\left[\frac{Q}{AP} + \frac{Q_i}{A_iP_i}\right] \quad \ldots\ldots (24)$$

20. To estimate the expected number of net migrants ad_i^w, we will use the following formula, from the Theory of Life Contingencies:

$$ad_i^w = A_iq_i^w\left[1 - 1/2\,(q_i^d + q_i^k) + 1/3\, q_i^d \cdot q_i^k\right] \quad (25)$$

where

$q_i^d = 1 - p_i^d$ and $q_i^k = 1 - K$

Substituting for p_i^d and K, we have:

$$ad_i^w = 1/6\, A_i\left[2 + P_i\epsilon_i + \frac{R}{P}\frac{1}{\epsilon_o} + 2\,\frac{P_iR}{P}\frac{\epsilon_i}{\epsilon_o} - 2\,\frac{R_i}{R}\frac{P}{P_i}\frac{\epsilon_o}{\epsilon_i} - \frac{R_iP}{R}\cdot\epsilon_o - \frac{R_i}{P_i}\frac{1}{\epsilon_i} - 2\,R_i\right] \quad \ldots\ldots (26)$$

Using the results:

$$E(\epsilon_i) = E(\epsilon_o) = 1 \qquad E\left(\frac{\epsilon_i}{\epsilon_o}\right) = \frac{E(\epsilon_i)}{E(\epsilon_o)} = 1$$

$E\left(\frac{1}{\epsilon_o}\right) = 1 \quad E\left(\frac{1}{\epsilon_i}\right) = 1$, it can be shown that

$$E(ad_i^w) = \frac{A_i}{6}\left[1 - \frac{R_i}{R}\frac{P}{P_i}\right]\left[2 + P_i + \frac{R}{P} + 2\,R\cdot\frac{P_i}{P}\right] \quad \ldots\ldots (27)$$

21. Let us rearrange the terms in ad_i^w from (26)

$$ad_i^w = \frac{A_i}{6}\left[2 - 2\,R_i + P_i\epsilon_i - \frac{R_iP}{R}\epsilon_o - \frac{R_i}{P_i}\frac{1}{\epsilon_i} + \frac{R}{P}\frac{1}{\epsilon_o} + 2\,\frac{P_iR}{P}\frac{\epsilon_i}{\epsilon_o} - 2\,\frac{R_iP}{RP_i}\frac{\epsilon_o}{\epsilon_i}\right]$$

$$= \frac{A_i}{6}\left[2 - 2\,R_i + C_1\epsilon_i + C_2\epsilon_o + C_3\frac{1}{\epsilon_i} + C_4\frac{1}{\epsilon_o} + C_5\frac{\epsilon_i}{\epsilon_o} + C_6\frac{\epsilon_o}{\epsilon_i}\right] \quad \ldots\ldots (28)$$

$$V(ad_i^w) = \frac{A_i^2}{36}\left[V(C_1\epsilon_i + C_2\epsilon_o + C_3\frac{1}{\epsilon_i} + C_4\frac{1}{\epsilon_o} + C_5\frac{\epsilon_i}{\epsilon_o} + C_6\frac{\epsilon_o}{\epsilon_i})\right] = \frac{A_i^2}{36}\left[C_1^2\,V(\epsilon_i) + C_2^2\,V(\epsilon_o) + C_3^2\,V\left(\frac{1}{\epsilon_i}\right) + C_4^2\,V\left(\frac{1}{\epsilon_o}\right) + C_5^2\,V\left(\frac{\epsilon_i}{\epsilon_o}\right) + C_6^2\,V\left(\frac{\epsilon_o}{\epsilon_i}\right) + \text{Covariance Terms}\right] \quad \ldots\ldots (29)$$

There will be $C_2^6 = 15$ convariance terms whose coefficients will be $2C_iC_j$ $i \neq j$. ($i = 1, 2, \ldots 6$, $j = 1, 2, \ldots 6$). Of these, 11 of the covariance terms are all zero when we use the independence of ϵ_o and ϵ_i and use the results $E\left(\frac{1}{\epsilon_i}\right) = \frac{1}{E(\epsilon_i)}$ etc. These terms correspond to following C_iC_j coefficients.

C_1C_2; C_1C_3; C_1C_4; C_1C_6; C_2C_3; C_2C_4; C_2C_5; C_3C_4; C_3C_5; C_4C_6; C_5C_6

The non zero convariance terms are:

$Cov\left(\epsilon_i, \frac{\epsilon_i}{\epsilon_o}\right) = V(\epsilon_i)$ Coefficient $2\,C_1C_5$

$Cov\left(\epsilon_o, \frac{\epsilon_o}{\epsilon_i}\right) = V(\epsilon_o)$ Coefficient $2\,C_2C_6$

$Cov\left(\frac{1}{\epsilon_i}, \frac{\epsilon_o}{\epsilon_i}\right) = \frac{-Q_i}{Q_i + A_iP_i} = \frac{-V(\epsilon_i)}{1 + V(\epsilon_i)}$, coefficient $2\,C_3C_6$ ……(30)

$Cov\left(\frac{1}{\epsilon_o}, \frac{\epsilon_i}{\epsilon_o}\right) = \frac{-Q}{Q + AP} = \frac{-V(\epsilon_o)}{1 + V(\epsilon_o)}$, coefficient $2\,C_4C_5$

Dealing with the variance terms, we have by use of (16)

$$V\left(\frac{1}{\epsilon_i}\right) = V(\epsilon_i) = \sigma_i^2$$

$$V\left(\frac{1}{\epsilon_o}\right) = V(\epsilon_o) = \sigma^2 \quad \ldots\ldots (31)$$

$$V\left(\frac{\epsilon_o}{\epsilon_i}\right) = V\left(\frac{\epsilon_i}{\epsilon_o}\right) = V(\epsilon_o) + V(\epsilon_i) = \sigma^2 + \sigma_i^2$$

Going back to (29) and noting that σ^2 and σ_i^2 are very small compared to 1 so that we can use the approximate formula $(1 + X)^{-1} = 1 - X$ by ignoring second and higher powers of X. On this assumption $X(1 + X)^{-1} = X$. Therefore if we ignore σ^4 and σ_i^4 as being negligible, it can be shown that

$$V(ad_i^{\bar{W}}) = \frac{A_i^2}{36} \left[\sigma_i^2 \left\{ P_i^2 + \frac{R_i^2}{P_i^2} + 4P_i^2 \frac{R^2}{P^2} + \frac{4R_i^2P^2}{R^2P_i^2} + 4P_i^2 \frac{R}{P} - 4 \frac{R_i^2}{P_i^2} \frac{P}{R} \right\} + \sigma^2 \left\{ \frac{R_i^2}{R^2} P^2 + \frac{R^2}{P^2} + 4P_i^2 \frac{R^2}{P^2} + \frac{4R_i^2P^2}{R^2P_i^2} + \frac{4R_i^2P^2}{R^2P_i} - \frac{4R^2}{P^2} P_i \right\} \right]$$

Let us write:

$$\bar{W}_i = \frac{R_i}{R} \cdot \frac{P}{P_i} \text{ and } \hat{K} = \frac{R}{P} \text{ and } \lambda_i = \frac{R_i}{P_i}$$

so that $\bar{W}_i = \lambda_i/\hat{K}$.

In our expression we retain P, P_i, $\hat{K}$ and $\bar{W}_i$.

$$V(ad_i^W) = \frac{A_i^2}{36} \left[\sigma_i^2 \left\{ P_i^2 (1 + 2\hat{K})^2 + \bar{W}_i^2 (2 - \hat{K})^2 \right\} + \sigma^2 \left\{ \hat{K}^2 (2 P_i - 1)^2 + \bar{W}_i^2 (2 + P_i)^2 \right\} \right] \quad ..(32)$$

22. By way of illustration, necessary calculations have been done for the state of North Carolina for 1950-60 decade in respect of white males (Appendices A and B). By way of comparison, the estimates of the number of net migrants and of net migration rates given by the proposed method and by U.S. Department of Agriculture for 1950-60 decade for the state of North Carolina for white male and white female categories are shown in Appendices C and D.

LIST OF REFERENCES

Bailey, W. G., and Haycocks, H. W., "Some Theoretical Aspects of Multiple Decrement Tables," (1946) Institute of Actuaries, London.

Hamilton, C. Horace, "Educational Selectivity of Net Migration from the South," Social Forces 38 (1959) 33-42.

Hamilton, C. Horace, and Henderson, F. M., "Use of the Survival Rate Method in Measuring Net Migration," Journal of the American Statistical Association, 39 (1944) 197-206.

Lee, Everett S., and Lee, Anne S., "Internal Migration Statistics for the United States," Journal of the American Statistical Association, 55 (1960) 664-697.

Lee, Everett S., et al., "Population Redistribution and Economic Growth, United States, 1870-1950," Volume 1. Philadelphia: The American Philosophical Society, 1957.

Price, Daniel O., "Examination of Two Sources of Error in the Estimation of Net Internal Migration," Journal of the American Statistical Association, 50 (1955) 689-700.

Siegel, Jacob S., and Hamilton, C. Horace, "Some Considerations in the Use of the Residual Method of Estimating Net Migration," Journal of the American Statistical Association, 47 (1952) 475-500.

Spuregeon, "Life Contingencies," Institute of Actuaries, London.

Wolfenden, Hugh H., "Population Statistics and Their Compilation, (1954)," published for the Society of Actuaries by the University of Chicago Press.

Zachariah, K. C., "A Note on the Census Survival Method of Estimating Net Migration," Journal of the American Statistical Association, 57 (1962) 175-183.

APPENDIX A

Main results of Calculations pertaining to:

(a) Expected Value and
(b) Variance of net out-migration rate

(i) North Carolina--White Male--1950-60 Decade

Age in 1950 X	Formula 5@	PROPOSED METHOD	
	$(MR)_i^u$	$E(q_i^W)$	$V(q_i^W)$
(1)	(2)	(3)	(4)
			10^{-5}
0-4	.03420	.03312	.00409
5-9	-.02867	-.02964	.00550
10-14	.02670	.02814	.00889
15-19	.13318	.13536	.00953
20-24	.10822	.10580	.01182
25-29	.06193	.05758	.01696
30-34	.03960	.03511	.02907
35-39	.04019	.03361	.04810
40-44	.02158	.01311	.09142
45-49	.01373	.00883'	.17300
50-54	.01997	.01916	.30940
55-59	-.01575	-.01644	.62533
60-64	-.02620	-.01503	1.19502
65+	-.00089	.01890	1.93990

@$(MR)_i^u = (A_iR - B_i) / A_i = R - R_i$.

(Note: A negative value signifies net in-migration rate)