# A MODIFIED CENSUS SURVIVAL RATIO METHOD FOR THE ESTIMATION OF INTERCENSAL NET INTERNAL MIGRATION

Alexander Mazurkewycz, Brandon University

As commonly applied, the forward census survival ratio method is defined as follows:

$$(1)\quad M_i = P_i^{(x+t,\,t)} - S_c P_i^{(x,o)}$$

where $M_i$ is the net migration for the ith region, $P_i^{(x+t,\,t)}$ is the enumerated population in region i at age x + t at time t, $P_i^{(x,o)}$ is the enumerated population in region i aged x at time o, and $S_c$ is the national survival ratio defined as

$$S_c = \frac{P_{\cdot}^{(x+t,\,t)}}{P_{\cdot}^{(x,\,o)}}$$

where P refers to summation over all i. The term $S_c P_i^{(x,o)}$ can be thought of as the "expected" population, expected in the sense that if there were no migration and the mortality conditions of the nation were evenly distributed, then this would be the "aged" population that we would expect.

The POBCSR method of Eldridge and Kim introduces a place of birth component into the method. Thus Eq. (1) becomes

$$(1a)\quad M_{ij} = P_{ij}^{(x+t,\,t)} - S_i P_{ij}^{(x,o)}$$

where $M_{ij}$ refers to net migration into region j of the population born in i, $P_{ij}$ refers to population residing at j and born in i, and

$$S_i = \frac{P_{i.}^{(x+t,\,t)}}{P_{i.}^{(x,\,o)}}$$

where $P_{i.}$ is population summed over all j, or in other words, total population born at i. It is clear that this technique is amenable to matrix manipulation so that we may now define the following:

$$P_{ij} = \begin{bmatrix} P_{11} & P_{12} & \cdot & \cdot & \cdot & P_{1n} \\ P_{21} & P_{22} & \cdot & \cdot & \cdot & P_{2n} \\ \cdot & \cdot & & & & \cdot \\ \cdot & \cdot & & & & \cdot \\ \cdot & \cdot & & & & \cdot \\ P_{n1} & P_{n2} & \cdot & \cdot & \cdot & P_{nn} \end{bmatrix}$$

and

$$M_{ij} = \begin{bmatrix} m_{11} & m_{12} & \cdot & \cdot & \cdot & m_{1n} \\ m_{21} & m_{22} & \cdot & \cdot & \cdot & m_{2n} \\ \cdot & \cdot & & & & \cdot \\ \cdot & \cdot & & & & \cdot \\ \cdot & \cdot & & & & \cdot \\ m_{n1} & m_{n2} & & & & m_{nn} \end{bmatrix}$$

and Q a diagonal matrix with the main diagonal cells $q_{ii}$ corresponding to the place-of-birth CSR's, $S_i$'s, $S_i$'s, where i = j. Equation (1a) can now be rewritten in matrix form:

$$(1b)\quad M_{ij} = P_{ij}^{(x+t,\,t)} - QP_{ij}^{(x,o)}$$

Let us further define a constant of proportionality C as follows:

$$(2)\quad C = \frac{P_{..}^{(x+t,\,t)}}{\sum_{i=1}^{n}\sum_{j=1}^{n} \frac{P_{j.}^{(x+t,\,t)}}{P_{j.}^{(x,o)}} P_{ij}^{(x,o)}}$$

where $P^{(x+t,\,t)}$ refers to the total population aged x + t at time t, and $P_{j.}$ refers to the total population born in region j irrespective of current place-of-residence. It is now possible to introduce a place-of-residence component into $M_{ij}$ as follows:

$$(3)\quad M_{ij} = P_{ij}^{(x+t,t)} - \tfrac{1}{2}(QP_{ij}^{(x,o)} + CP_{ij}^{(x,o)}Q)$$

The essence of Eq. 3 is that net migration is the difference between the enumerated population at the time of the second census, and an expected population at that place, calculated on the basis of expected mortality. Thus, the expected population in Eq. 3 is equivalent to the term $\tfrac{1}{2}(QP_{ij}^{(x,o)} + CP_{ij}^{(x,o)} Q)$. amd consists of a place-of-birth census survival ratio, i.e. $QP_{ij}^{(x,o)}$ a place-of-residence survival ratio, i.e. $P_{ij}^{(x,o)}$ Q.

## Discussion of Method

It is important to remember, first of all, that Equation 3 is written in matrix form. Thus, the term $QP_{ij}^{(x,o)}$ has the effect of multiplying each row in $P_{ij}^{(x,o)}$ with the corresponding element in the diagonal matrix Q. This is the procedure advocated by Eldridge and Kim in the POBCSR method. The postmultiplication term $P_{ij}^{(x,o)}$ Q is equivalent to multiplying every column by the corresponding survival ratio in Q. The net effect of these multiplications is that every cell in the $P_{ij}^{(x,o)}$ matrix gets a unique combination of survival ratios applied to it, i.e. a combined survival ratio of all those born at i and j. Since the CSR's are usually heavily weighted with these residing at place-of-birth, the combination of CSR's or Equation 3 can be justified as follows. The mortality experiences of a place-of-birth cohort may be combined with the mortality experiences of these born at place-of-residence because the place-of-residence CSR is heavily influenced by the population still residing at place-of-birth. The place-of-birth CSR is similarly weighted with those still residing at place-of-birth. Thus the method of Eq. 3 is a mean of the CSR's at place-of-birth and place-of-residence. This combination of mortality experiences should reflect actual mortality somewhat better than merely taking a place-of-birth CSR and applying it across a cohort.

It may appear initially that a more appropriate combination of survival ratios might include place-of-residence by using some formulation such as this: $S_j = P_{.j}^{(x + t,\ t)}/P_{.j}^{(x,o)}$. Although such a formulation might increase the mathematical elegance of the model, it does not help in interpreting any estimates since $P_{.j}$ does not consist of any identifiable cohort. Individuals may move into or out of a particular place at any time, therefore it makes little sense to construct a survival ratio for a group that is not closed in any way. Such a survival ratio already would have migration confounded within it, and there is no meaningful way of extracting the migration component from the mortality component. As a result, the constant of proportionality C must be introduced into the method.

The C coefficient is not as cumbersome and difficult to calculate as Eq. 2 may imply. It is merely the quotient of the total enumerated population at the second census divided by the overall total of all the cells of the matrix resulting from post-multiplication of the $P_{ij}^{(x,o)}$ matrix by the matrix operator Q. It is constant for every cohort, that is, for every $P_{ij}^{(x,o)}$ matrix.

A loss in elegance occurs when the POBCSR method is modified into the method of Eq. 3, or Combined CSR method for short; principally, no longer is the total expected population of a given place-of-birth equal to the total enumerated population of a place-of-birth. This is one of the strengths of the POBCSR method. The row totals Computed by the combined CSR method are off by a factor equal to (1 - C). However, this error is not greater than that incurred when the regular CSR method is employed, and most likely is less. Furthermore, this "error" is distributed throughout the population matrix in proportion to the size of each cell. The "error" involving each cell is not only contingent on the value of C, but is also dependent on the overall size of the $P_{ij}$ matrix, and the relative frequency in each cell.

One further point regarding the constant of proportionality C requires elaboration. The closer that the value of C is to 1, the less "error" is involved in the migration estimates. This is largely a function of the number of cells in the overall population matrix $P_{ij}$, as well as the distribution of population in the matrix, and number of zero cells. Population matrices with disperse populations will produce C values that are quite close to 1. Because the value of C for a cohort is easily derived, it is apparent that a quick check is available to the researcher. If computed values of C deviate considerably from 1, say, by more than .05 or so, then an alternate method may be called for, perhaps the POBCSR method. If the C values lie close to 1, then some degree of confidence may be placed in migration estimates derived from the combined CSR method.