The Changing Geographic Distribution of the Elderly

Estimating Net Migration Rates with Social Security Data

Kevin F. McCarthy, Allan Abrahamse, Charles Hubay
The research reported herein was performed pursuant to Grant No. 5 R01 AG-01767-02 from the National Institute on Aging, U.S. Department of Health and Human Services.

The Rand Publication Series: The Report is the principal publication documenting and transmitting Rand's major research findings and final research results. The Rand Note reports other outputs of sponsored research for general distribution. Publications of The Rand Corporation do not necessarily reflect the opinions or policies of the sponsors of Rand research.

Published by The Rand Corporation
The Changing Geographic Distribution of the Elderly

Estimating Net Migration Rates with Social Security Data

Kevin F. McCarthy, Allan Abrahamse, Charles Hubay

March 1982

Prepared under a grant from the National Institute on Aging, U.S. Department of Health and Human Services
PREFACE

This report was prepared under a grant from the National Institute on Aging, U.S. Department of Health and Human Services. It is one product of a larger Rand inquiry into the elderly population's changing geographic distribution. The report describes a procedure designed to estimate a county's elderly net migration rate using administrative data from the Social Security Administration. This procedure, which is relatively straightforward, is intended for use by state and local planning agencies.
SUMMARY

Although the aging of America's population represents a fundamental demographic shift with far-reaching implications, its pressures will be felt unevenly by the nation's localities. This report presents a straightforward procedure that state and local planning agencies can use to monitor the movement of elderly persons into and out of individual U.S. counties. The procedure uses readily available Social Security data to produce accurate estimates of elderly net migration rates. The report describes the procedure's method and rationale, documents the formal estimation model, explains how to apply the model, and furnishes an illustration to guide the user.
ACKNOWLEDGMENTS

The authors acknowledge the considerable assistance provided by their Rand colleagues. Thomas Blaschke and Michael Hirsch performed much of the computer work. Peter Morrison's review of an earlier draft contributed significantly to the organization and presentation of this report.

We also gratefully acknowledge the comments of Norfleet Rives, University of Delaware, and William Serow, Florida State University, who reviewed the final draft.
CONTENTS

PREFACE ................................................................. iii

SUMMARY ............................................................... v

ACKNOWLEDGMENTS ................................................... vii

LIST OF TABLES ...................................................... xi

Section

I. INTRODUCTION ..................................................... 1

II. THE ESTIMATION METHOD AND ITS RATIONALE ...................... 4
   Rationale for Design .............................................. 4
   Choice of Symptomatic Indicators ................................. 6
   Linking the RWB Indicator to Measured Migration ............... 8
   The Model .......................................................... 8
   Choosing Data to Calibrate the Model ............................ 12

III. CALIBRATING THE MODEL ......................................... 15
    Parameter Estimates for All Counties .......................... 15
    Parameter Estimates by Strata ................................ 17
    End-Period Parameters by Period ............................... 19

IV. APPLYING THE MODEL ............................................... 22
    Adjusting the Model to the Estimation Interval ............... 22
    Choice of Rates ................................................ 27
    An Illustrative Application .................................... 28
    Step 1--Assembling the Raw Data ............................... 28
    Step 2--Computing Input Variables for the Model ............ 29
    Step 3--Determining the Appropriate and Model .............. 30
    Step 4--Making Assumptions About Migration Patterns ....... 30
    Step 5--Calculating a Ten-Year Net Migration Estimate ....... 31
    Step 6--Adjusting to the Estimation Interval ................ 33

Appendix

A. DERIVATION OF THE MODEL ...................................... 35
B. PREDICTION EQUATIONS FOR BEGINNING AND END-PERIOD MIGRATION RATES ........................................... 39

BIBLIOGRAPHY ........................................................ 41
TABLES

1. Comparison of Regression Coefficients of Basic Estimation Model for End-Period and Begin-Period Migration Rates: All Counties, 1960-1970 ................................. 16

2. Comparison of $R^2$ of Basic Estimation Model for End-Period and Begin-Period Migration Rates, By Size of County: 1960-1970. 18

3. Comparison of $R^2$ of End-Period Model in 1960-60 and 1965-70 Data Sets: By Size of County ................................. 20

4. Comparison of Volume and Rate of Net Migration Under Linear and Geometric Assumptions ................................. 25
I. INTRODUCTION

The nation's elderly (people over 65) now number over 25 million, or 11 percent of the total U.S. population. In the years ahead the ranks of the elderly will continue to swell; by the year 2035 they will number over 55 million—between 18 and 23 percent of the total population. (Florida, by comparison, currently leads the nation with approximately 18 percent of its population 65 and older.) Moreover, the elderly population is itself growing older; more and more people are living past the age of 64 (Uhlenberg, 1977). Although the full force of these shifts will not be felt until after the year 2010 (when the large baby-boom generation begins to turn 65), the intervening decades will nonetheless witness a pronounced increase of the elderly.

This so-called "graying" of the population is one of the fundamental and far-reaching demographic shifts of our time, and its effects will be pervasive. As the structure of our national population shifts toward the older ages, health-care needs will rise, not only absolutely but also per capita. The elderly, who are increasingly inclined to live alone, will need a variety of support services, such as homemaking assistance, visiting nurses and, perhaps eventually, nursing homes, to cope with the difficulties of being alone in declining health. The availability of housing and transportation facilities suited to the special needs of the elderly may become problematic.

Of paramount importance, however, is where the pressures of these developments will be felt. The elderly are unevenly distributed across the nation's states and cities, and the demographic processes giving
rise to this unevenness can be expected to continue. Accordingly, the overall national outlook will understate the experience in store for some localities: a more abrupt or extreme "graying" of their population that will come about sooner than the next century.

As responsibility for human welfare and public services is shifted downward from federal to lower levels of government, local jurisdictions will have to learn to contend with the local manifestations of an aging population. Clearly, they will need to monitor where older citizens are settling if they are to plan for meeting their future needs.

This report is one product of a larger inquiry into the elderly population's changing distribution. It deals with the most elusive component of change in a locality's elderly population: the influx of elderly newcomers through migration. Elderly migration is elusive because it is inadequately and often poorly measured and, over the long term, difficult to predict.

Currently, age-specific net migration data on a local scale are systematically gathered only every ten years through the decennial census. Furthermore, census migration data are collected on a sample basis and are unstable for many of the nation's more sparsely populated counties.

In addition, the decision process underlying the locational choices of elderly persons is far less straightforward than that of the working-age population. Whereas younger migrants typically pursue employment opportunities and favor economically prosperous labor markets, elderly migrants are not "opportunity-oriented" in that sense. With the expansion of private pension plans and the indexing of Social
Security benefits for inflation, older persons are assured a steady income regardless of where they choose to locate; consequently, their migration decisions are more "consumption-oriented."

This report focuses on the problem of monitoring the yearly flow of elderly persons into and out of individual U.S. counties. We report on a procedure designed to estimate a county's elderly net migration rate using administrative data from the Social Security Administration. The procedure, which is relatively straightforward, is intended for use by state and local planning agencies, which otherwise would be forced to rely on outdated or less objective "back of the envelope" estimates of elderly migration.

Section II describes our method and its rationale. Section III documents the formal estimation model and its calibration. Section IV explains how to apply the model and furnishes an illustration to guide the user. The appendixes document several technical aspects of our procedure and the parameters to use in making elderly net migration estimates.
II. THE ESTIMATION METHOD AND ITS RATIONALE

Our estimation method is driven by an administrative data series that reflects change in a county's elderly population: annual counts of the number of elderly Social Security beneficiaries in each of the nation's roughly 3000 counties. The method transforms year-to-year changes in this series into estimates of change in the county's elderly population, and the net migration implied by such change.

RATIONALE FOR DESIGN

Virtually all estimation techniques operate with reference to the following balance equation:

\[ P_e = P_c + B - D + I - O \]  

(1)

where \( P_e \) = population to be estimated

\( P_c \) = population at last available census

\( B \) = number of births between census and estimation dates

\( D \) = number of deaths between census and estimation dates

\( I \) = number of in-migrants between census and estimation dates

\( O \) = number of out-migrants between census and estimation dates

This equation states a mathematical identity: The population to be estimated equals the enumerated population at an initial date (typically the date of the last Census) plus the two components of change in that population: natural increase (the difference between births and deaths), and net migration (the difference between in- and out-migration).
Estimation techniques vary in how these components of change are derived. Ideally, both components would be directly measured from official registrations, as are births and deaths. Direct measurement, while intrinsically more accurate, is often infeasible for handling the net migration component, since migration is not registered in this way. Analysts must therefore rely instead on symptomatic indicators that mirror this component. The range of possible symptomatic measures is limited only by the analyst's ingenuity. They may consist of any population-related objects and services, e.g., new residential utility hookups (which would mirror residential change) or new motor vehicle licenses (reflecting the arrival of interstate migrants). Although recorded for other purposes, such records can furnish useful demographic intelligence where direct information is lacking.

One standard demographic procedure for estimating migration is known as the "forward survival method." It entails comparing an area's recorded population at two successive censuses, subtracting estimated natural increase, and inferring net migration as the residual. [1] This method generally yields more accurate estimates of net migration than do symptomatic measures (especially for small areas). However, it requires data on the population's age structure at both the beginning and end of the estimation interval—a obvious limitation for postcensal estimation.

[1] Natural increase is estimated by applying life table or "census" survival rates to the age structure of the population at an earlier census to calculate expected deaths, and adding in births between censuses. These adjustments produce an expected population, adjusted for natural increase, in the absence of any net migration. The difference between the expected population and the population actually recorded is then attributed to net migration, allowing for the effects of undercounting (see Shryock and Siegel, 1973).
(i.e., estimation beyond the latest census). Lacking such data, the forward survival approach relies instead on indirect estimates (see Tordella, 1980), which themselves have been generated by symptomatic data.

CHOICE OF SYMPTOMATIC INDICATORS

The appropriateness of a given symptomatic indicator hinges on its possessing several characteristics suggested by Shryock and Siegel (1973):

- Availability on a timely and continuing basis;
- Substantial coverage of the population to be estimated;
- Internal consistency from year to year;
- Fluctuation principally in response to changes in the population they are intended to monitor.

We added the following criteria for the method to satisfy the requirements of a user population whose primary interest lies in applying the method and whose individual needs will vary:

- Ease of computation: The method should be computationally straightforward and use readily available data to facilitate widespread use at the state and local level.
- Flexibility: The method should be adaptable to varying circumstances and user needs.
- Proven accuracy: The method should be validated by means of test runs and its level of accuracy gauged.
We have chosen Social Security Administration data on retired worker beneficiaries (RWB) as our symptomatic indicator. Annual counts of RWB by county are available on a timely and continuing basis. Coverage is excellent because Social Security covers approximately nine-tenths of all Americans 65 and older (Social Security Administration, 1979). In addition, an internally consistent year-to-year record of RWB can be derived from Social Security administrative data.\(^2\) Finally, the size of a county's elderly population appears to be the major determinant of changes in its number of RWB.

RWB constitute the core of prime beneficiaries of the Social Security Program. This group is, by administrative definition, at least 62 years of age, and hence excludes nonelderly beneficiaries who qualify on other grounds (e.g., as survivors of prime beneficiaries or on the basis of disability). However, the RWB group excludes certain elderly beneficiaries: those who receive payments as survivors of prime beneficiaries, those who do not qualify for benefits because they have not worked sufficient quarters in covered employment, and those who are currently employed and whose earnings exceed the allowable maximum. Although these exclusions may affect the internal consistency of the RWB indicator, the problem should not be serious, in view of the substantial proportion of the elderly who currently receive benefits, combined with the fact that the working elderly appear highly sensitive to the allowable earnings maximum in determining their work effort (Munnell, 1977). One possible exception may occur in the handful of counties with

\(^2\)Although categories of assistance and types of beneficiaries within the various Social Security programs have changed over time, few recent changes have affected the retired worker program.
a large number of retired federal workers, since federal employees are not enrolled in the Social Security system and will not qualify as RWBs unless they also have sufficient covered employment. Thus, we can reasonably assume that changes in the number of RWB in a county over time faithfully mirror concurrent changes in the size of its elderly population.

LINKING THE RWB INDICATOR TO MEASURED MIGRATION

Having chosen the RWB as our measure, we must resolve two more issues before we can estimate elderly net migration rates. First, the relationship between changes in counts of RWB and elderly net migration rates must be specified in terms of a model that can be estimated. Second, that model must be estimated for a benchmark period for which both Social Security and net migration data are available. The model can then be used to calculate the unknown net migration rates for postcensal years using known RWB data.

THE MODEL

Our estimating model derives from Eq. (1), rewritten to express the net migration component to be estimated in terms of the other variables:[3]

\[ I - O = (P_e - P_c) - (B - D) \quad \text{(2)} \]

From Eq. (2) a net migration rate can be defined as the net change in the number of migrants divided by an appropriate denominator.

[3]A formal derivation of this model is presented in App. A.
Traditionally, two such denominators are used: the population at either the beginning or the end of the migration interval. [4] Use of the former yields a beginning-period rate, the latter an end-period rate. Dividing both sides of Eq. (2) by $P_c$, the initial population, gives the expression for the beginning-period net migration rate:

$$\frac{I - O}{P_c} = \left(\frac{P_e}{P_c}\right) - 1 - \left(\frac{B - D}{P_c}\right).$$

The two parenthetical terms on the right-hand side of Eq. (3) correspond to the net growth rate and the rate of natural increase (births minus deaths), respectively. Solving this equation requires estimating both terms.

In the population as a whole, natural increase equals births minus deaths. Natural increase among the elderly does not consist of "births," of course, but is rather the result of local residents reaching the age of 65 and staying where they are—a process we term "aging in place." Their numbers can be estimated from available census data as some number, $Q_c$, of people aged 55 to 64 who live in the county at the beginning of the migration interval. To estimate the number of deaths, we must account for mortality among persons over 65 and among those turning 65 during the interval. We assume that some proportion, $a$, of those already 65 or older at the beginning of the interval, and

[4] This tradition reflects the availability of data more than it does the theoretically appropriate choice. Indeed, a true demographic rate should express the incidence of a particular phenomenon (in this case migration) in the population at risk. Neither the initial population nor the final population satisfies this criterion (Hamilton, 1965).
some different proportion, b, of those who will "age in place," will die during the interval. We then rewrite B - D as follows:

\[ B - D = aP_c + bQ_c, \]  

where a and b are unknown parameters. Dividing through by \(P_c\) thus yields:

\[ \frac{B - D}{P_c} = a + b\frac{Q_c}{P_c}. \]  

(5)

We have no direct measure of the elderly population's net growth rate, of course, since \(P_e\) is unknown. However, changes in RWB between the census date and the date of our estimate \(n\) years later should reflect changes in the size of that population. Thus, the ratio of recipients \(R_i\) at those two periods, \((R_c + n^/-R_c)\), can serve as an estimate of \(P_e/P_c\).

The relationship between \((R_c + n^/-R_c)\) and \(P_e/P_c\) will be conditioned, of course, by any changes in the recipiency rate (i.e., the rate at which the elderly receive Social Security).[5] This rate is defined as the number of recipients divided by the number of elderly at a given time. The net growth rate, therefore, can be estimated as:

\[ \frac{P_e}{P_c} = c\left(\frac{R_c + n}{R_c}\right)^{-d}\left(\frac{Q_{c+n}}{Q_c}\right), \]  

(6)

[5] If the recipiency rate is increasing, \((R_c + n^/-R_c)\) will over-predict changes in the elderly population, whereas it will underpredict if the rate is decreasing.
where c and d are unknown parameters and $\theta_{c+n}$ and $\theta_c$ are recipiency rates at $c + n$ and c, respectively.[6]

The beginning-of-interval recipiency rate, $\theta_c$, is calculated by dividing the number of recorded recipients, $R_c$, on the date the census was taken by the number of census enumerated elderly, $P_c$, on that date.[7] The end-of-interval recipiency rate, $\theta_{c+n}$, can only be estimated since we do not know the number of elderly residents in year $c + n$ and must instead use an estimate of that population. Since the Census Bureau issues annual estimates of county populations by age, $\theta_{c+n}$ was estimated by dividing the number of recipients in year $c + n$ by the Bureau's estimate of the elderly population in that year. Although the Bureau's age-structure estimates often have high variances--particularly for sparsely populated counties--they are unbiased and should not introduce systematic error into the estimation procedure.[8]

[6] We note, by the way, that for $c = 1$ and $d = P_{e} / P_c$, Eq. (6) is an identity. In App. A we show that when $c = 1$ and $d$ is the average growth rate, the right-hand side of Eq. (6) may be a much more accurate estimator of $P_{e} / P_c$ than the average growth rate. When we estimate both $c$ and $d$, we can expect still greater accuracy.

[7] In fact, recipient counts are for December 31 of the preceding year, while the census records population as of April 1, three months later. In calibrating our model, these two counts are treated as though they were recorded simultaneously.

[8] The high variance of these age-structure estimates is one of several reasons for preferring the estimation technique proposed here to the forward survival approach for intercensal migration estimates. With the forward survival approach, the annual age-structure estimates serve as the central control total for estimating net migration, since net migration is estimated by subtracting the initial population and estimated natural increase from these annual estimates. To the extent that there is significant error in the annual age-structure estimates--as indeed there appears to be, especially in counties with less than 10,000 residents (see Table 2)--that error will be directly reflected in the migration estimate. Our technique, on the other hand, uses these age-structure estimates as only one component of the estimation model and thus should be less sensitive to significant errors in the estimate. In addition, the forward survival approach assumes the appropriate and fixed survival rates, whereas our approach essentially calculates those
Combining the expressions for the natural increase and net growth components yields the model:

\[
\frac{I - O}{P_c} = a + c\left(\frac{R_c + n}{R_c}\right) - d\left(\frac{\theta_c + n}{\theta_c}\right) - b\left(\frac{Q_c}{P_c}\right)
\]  

(7)

where \(a\), \(b\), \(c\), and \(d\) are parameters that can be estimated by ordinary least squares (OLS) regression. Because \(a\) is a yet-to-be-determined parameter, we can ignore the "minus one" term in Eq. (3) for this equation.

The same approach can be used to estimate an end-period migration rate by substituting \(P_{c+n}\) as the denominator. However, since the population at the end of the interval reflects net growth during the interval, this approach is equivalent to multiplying each of the model's four parameters by the elderly population's average growth rate over the interval. County-by-county variation around that average growth rate will introduce some error into specific county estimates.

**CHOOSING DATA TO CALIBRATE THE MODEL**

To estimate the model, we need data on each of the terms in Eq. (7). Counts of prime worker beneficiaries at the county level are

rates based on the actual data and allows them to vary by strata. Indeed, our technique affords users the flexibility to tailor their estimates to specific local circumstances by altering assumptions about the pattern of local growth and providing a choice of rates. Finally, once the parameters of our model are estimated, calculating the desired estimate is accomplished by simply plugging local data into the appropriate model.
available for each year since 1959;[9] aging in place and coverage change by county can be estimated with census data.[10] The principal constraint, then, was the availability of benchmark net migration data.

Two alternative sources on net migration were available: (1) 1960-1970 age-specific net migration estimates prepared by Bowles et al. (1975); (2) 1965-1970 age-specific gross migration as measured in the 1970 Census (U.S. Bureau of Census, 1977). These sources differ in important ways. The former estimates were made by applying the forward survival procedure to the 1960 and 1970 population totals enumerated in those censuses to estimate net migration during the interval between Census dates. The 1965-1970 migration measures were calculated from a 1970 Census question on prior residence in 1965, asked of a 15 percent sample of the enumerated population. These differences affect the suitability of the alternatives for our purposes; we chose the 1960-1970 estimates as our preferred alternative.[11]

[9] We are grateful to Mr. Philip Lerner of the Social Security Administration for providing these data.

[10] The number of people aging in place can be estimated from the county's age structure at the prior census. Estimates of coverage change by county require annual counts of recipients and of the elderly population. The initial size of the elderly population is available in the prior census; estimates of the elderly population for any year between censuses can be made by interpolation.

[11] Our choice was based on several considerations. First, the forward survival method generally yields more accurate estimates when applied to successive census data. Second, since the 1960-1970 estimates were calculated using enumerated population totals rather than sample data, those estimates will be more reliable than the 1965-1970 estimates, particularly for thinly settled counties. Third, considerable error was introduced into the 1965-1970 estimates by misreporting of residence in 1965. Finally, the 1960-1970 interval offers greater flexibility than the 1965-1970 interval, since the availability of Census data at the beginning and end of the migration interval facilitates the calculation of both beginning- and end-period migration rates. To estimate beginning-period rates with the 1965-1970 migration estimates, prior estimates of the 1965 elderly population subject to the risk of migrating are required. By their very nature, such estimates would add additional error to the estimation procedure.
Still, by using both data sets, we can not only test the robustness of our model but also compare the accuracy and reliability of the estimates themselves. Indeed, as we demonstrate in Sec. III, the reliability of these alternative net migration estimates appears to vary significantly.

Although our model is designed to estimate migration for all counties, certain data problems forced us to delete a few counties from our analysis.[12] These deletions should not seriously affect the generality of our model, though, since they account for only some 50 of the nation's approximately 3100 counties.

[12] In Alaska, for example, counties do not exist as such; boroughs and census divisions constitute the equivalent units. The Social Security Administration reports on Alaskan beneficiaries for different geographic units from those in either the 1960-1970 or 1965-1970 estimates. Accordingly, we excluded Alaska from our analysis. (Less than 5 percent of Alaska's population is over 65.) Similar problems of geographic comparability arose for selected parts of Virginia, where data for independent cities are reported separately from counties. Although we attempted to link data for independent cities and the counties in which they were located, this was not always possible. Consequently, selected independent cities in Virginia were dropped from the analysis. Finally, a small number of comparability problems elsewhere resulted in a few other counties being dropped from the analysis.
III. CALIBRATING THE MODEL

Two sets of parameters were estimated for the model: one for beginning-period and another for end-period elderly migration rates. Several samples were used in preparing these estimates. First, separate estimates were made for all counties contained in the 1960-1970 file. Second, the two sets of parameters were estimated separately for each of eight size-strata of counties, since sample stratification generally, although not always, improves the accuracy of estimation models (see Rosenberg, 1968; Pursell, 1970; Pittenger, 1976; Martin and Serow, 1978). (Of the several stratification schemes examined, population size proved to be the most effective.) Stratification not only improves the overall accuracy of the model, but also allows us to compare the model's accuracy among different county strata. Finally, to evaluate the robustness of our model, we estimated end-period parameters with the Census Bureau's 1965-1970 elderly net migration rates for all counties, both unstratified and stratified by size.

PARAMETER ESTIMATES FOR ALL COUNTIES

Our first step was to estimate regression coefficients for the beginning- and end-period models for all counties together. These coefficients and the summary statistics for the two equations are listed in Table 1. The high $R^2$ for both of these equations indicates that the estimation model predicts beginning- and end-period elderly net migration rates quite accurately. All four parameters in both equations are highly significant and have the correct sign. Moreover, the
Table 1

COMPARISON OF REGRESSION COEFFICIENTS OF BASIC ESTIMATION MODEL FOR END-PERIOD AND BEGIN-PERIOD MIGRATION RATES: ALL COUNTIES, 1960-1970

<table>
<thead>
<tr>
<th>Independent Variable Parameter</th>
<th>End Period</th>
<th>Begin Period</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coefficient</td>
<td>t</td>
</tr>
<tr>
<td>Constant (a)</td>
<td>.465</td>
<td>70.2</td>
</tr>
<tr>
<td>Recipient ratio (c)</td>
<td>.367</td>
<td>121.9</td>
</tr>
<tr>
<td>Change in coverage (d)</td>
<td>-.418</td>
<td>93.5</td>
</tr>
<tr>
<td>Aging in place (b)</td>
<td>-.544</td>
<td>93.6</td>
</tr>
<tr>
<td>R^2</td>
<td>.842</td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>5,423.3</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>3053</td>
<td></td>
</tr>
<tr>
<td>Standard error of estimate</td>
<td>.046</td>
<td></td>
</tr>
</tbody>
</table>

The coefficient of the recipient ratio term, (R_c - n)/R_c, has the highest t-value in both equations, suggesting, as we hypothesized, that changes in the number of RWB is, in fact, an excellent indicator of changes in the size of a county's elderly population and hence serves as an extremely good predictor of its elderly net migration rate.

The higher R^2 for the beginning-period equation confirms that, as expected, the model predicts beginning-period net migration rates more accurately than it does end-period rates. The coefficients in the end-period equation are, in effect, multiplied by the elderly population's
average growth rate during the interval. Since not all counties experience the same rate of growth, this procedure increases the error component in the end-period model.

PARAMETER ESTIMATES BY STRATA

Calibrating our estimation model on all counties is equivalent to attaching the same weight to Los Angeles County's 7,000,000 residents and Hinsdale County, Colorado's population of scarcely 500. The model's predictive capacity can be improved by estimating parameters separately for similar size counties. One factor embedded in the end-period parameters, for example, is the county's elderly population growth rate. That rate may differ considerably among different types of counties. Similarly, the degree of elderly "aging in place" or change in recipiency rates also may differ among counties. For these reasons, we stratified counties by size and estimated the beginning- and end-period parameters for each stratum separately. The predictive power, $R^2$, of these separate equations, as well as the size of each stratum, are reported in Table 2. (The complete equations are reported in App. B.)

These results indicate that stratification by population size improves the model's predictive power. With only two exceptions, the $R^2$s of the stratum-specific equations surpass the $R^2$s of the equations for all counties. The exceptions, not surprisingly, are in the smallest population size stratum. As noted earlier, the estimates of end-period elderly populations (used to compute the change in coverage term) frequently have very high variances in such thinly settled counties. Overall, however, these comparisons underscore the predictive power of
Table 2

COMPARISON OF $R^2$ OF BASIC ESTIMATION MODEL FOR END-PERIOD AND BEGIN-PERIOD MIGRATION RATES, BY SIZE OF COUNTY: 1960-1970

<table>
<thead>
<tr>
<th>Size of County, 1960</th>
<th>End-Period Rate</th>
<th>Begin-Period Rate</th>
<th>Number of Counties</th>
</tr>
</thead>
<tbody>
<tr>
<td>1,000,000+</td>
<td>.973</td>
<td>.999</td>
<td>22</td>
</tr>
<tr>
<td>5000,000-999,999</td>
<td>.913</td>
<td>.991</td>
<td>53</td>
</tr>
<tr>
<td>250,000-499,999</td>
<td>.961</td>
<td>.957</td>
<td>71</td>
</tr>
<tr>
<td>100,000-249,999</td>
<td>.904</td>
<td>.988</td>
<td>190</td>
</tr>
<tr>
<td>50,000-99,000</td>
<td>.896</td>
<td>.957</td>
<td>332</td>
</tr>
<tr>
<td>25,000-49,999</td>
<td>.821</td>
<td>.988</td>
<td>555</td>
</tr>
<tr>
<td>10,000-24,999</td>
<td>.875</td>
<td>.955</td>
<td>989</td>
</tr>
<tr>
<td>&lt; 10,000</td>
<td>.859</td>
<td>.900</td>
<td>841</td>
</tr>
<tr>
<td>All counties</td>
<td>.842</td>
<td>.937</td>
<td>3053</td>
</tr>
</tbody>
</table>

the basic model. For example, although the model performs somewhat better in more than in less populous counties, in no stratum does it explain less than 82 percent of the variance and, in all but the smallest strata, it uniformly explains at least 90 percent of the variance. Similarly, the standard errors of the estimates are uniformly low, ranging from .012 to .052 in the begin-period equations and from .026 to .054 in the end-period equations. Finally, this comparison also documents the consistency with which the model predicts beginning-period rates more accurately than end-period rates.
END-PERIOD PARAMETERS BY PERIOD

Although the 1960-1970 elderly net migration estimates should be more accurate than the Census Bureau's 1965-1970 estimates, we have calibrated our end-period model separately for both sets of estimates in order to test our model's robustness on two very different types of estimates. (We have not estimated the beginning period rate on the 1965-1970 migration estimates since we lack an estimate of the 1965 elderly population. Calculating such an estimate would increase the error of the estimate.)

Overall, the model performs much better with the 1960-1970 than with the 1965-1970 data. The $R^2$ for the all-counties, end-period equation is .842 with the 1960-1970 data versus .312 with the 1965-1970 data. (See Table 3.) Similarly, the standard error of the estimate for the 1965-1970 equation, 7.52, is far larger than that for the 1960-1970 equation, .046. We cannot tell, a priori, whether these differences reflect the model or the relative quality of these two data sets. However, a comparison of the model's predictive power across the eight strata in Table 3 suggests that the difference can be attributed to the large sampling error associated with the Census Bureau's 1965-1970 migration estimates, particularly those for small counties. Specifically, the model's performance falls off sharply for the less populous counties in the Census Bureau's data set, but only moderately for the 1960-1970 data. The model's failure to predict elderly net migration more accurately in the 1965-1970 data is not surprising, given that the Census Bureau's estimates are based on a 15-percent sample of
Table 3

COMPARISON OF R² OF END PERIOD MODEL IN 1960-70 AND 1965-70 DATA SETS: BY SIZE OF COUNTY

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1,000,000+</td>
<td>.973</td>
<td>.915</td>
</tr>
<tr>
<td>500-999,999</td>
<td>.913</td>
<td>.948</td>
</tr>
<tr>
<td>250-499,999</td>
<td>.961</td>
<td>.854</td>
</tr>
<tr>
<td>100-249,999</td>
<td>.904</td>
<td>.850</td>
</tr>
<tr>
<td>50-99,999</td>
<td>.896</td>
<td>.708</td>
</tr>
<tr>
<td>25-49,999</td>
<td>.821</td>
<td>.620</td>
</tr>
<tr>
<td>10-24,999</td>
<td>.875</td>
<td>.356</td>
</tr>
<tr>
<td>&lt; 10,000</td>
<td>.859</td>
<td>.244</td>
</tr>
<tr>
<td>All Counties</td>
<td>.842</td>
<td>.312</td>
</tr>
</tbody>
</table>

the counties' 1970 residents. [1] Since approximately 60 percent of the nation's counties contain fewer than 25,000 inhabitants, this pattern produces the substantial R² discrepancy between the all-county equations. Although the 1965-1970 estimates have been allocated for nonresponse, it is also possible that substantial error has been introduced by misstatement of place of residence in 1965. Regardless of

[1] As an illustration, assume that the average population of counties in the two least populous strata is 15,000, of whom 10 percent (the national average, U.S. Bureau of Census, 1979) are at least 65. Each of these counties would then contain 1500 elderly residents, 225 of whom would have been included in the Census's 15 percent sample. If the elderly in these counties moved between counties at the same rate, 8.3 percent, as did all elderly Americans between 1965 and 1970 (U.S. Bureau of Census, 1972), then the net migration estimates for these counties would be based, on average, on a gross migration (both in-migrants and out-migrants) of fewer than 20 cases. Estimates based on such small samples are inherently unstable and should be used cautiously.
the reasons, however, this comparison illustrates the instability of the Census Bureau's 1965-1970 net migration rates for the nation's many small counties and suggests they be used with considerable care.

In summary, our estimation model has only four parameters and is computationally straightforward. Nonetheless, it predicts elderly net migration rates—especially beginning-period rates—at the county level very accurately. Although it performs better when applied to larger rather than smaller counties, it provides a remarkably good estimate even in the smallest places. Finally, stratifying counties by population size and calibrating the model separately for each stratum enhances the model's accuracy.
IV. APPLYING THE MODEL

An estimation model's utility depends not only on its accuracy but also on its ease of implementation. Our procedure is designed to be computationally straightforward, enabling the user to simply plug the necessary data into the appropriate formula. However, there are definite pitfalls in mechanically applying any estimation procedure without an appreciation of the strengths and weaknesses of the approach. This section considers two issues confronting the user: tailoring the model to the user's desired estimation interval, and choosing the appropriate net migration estimate for the user's purposes. Finally, we carry though a concrete application of the procedure with data for a specific county.

ADJUSTING THE MODEL TO THE ESTIMATION INTERVAL

Since our model was calibrated on data for a ten-year migration interval, the parameters are designed to produce ten-year migration estimates. The choice of the ten-year interval was dictated only by the availability of the necessary net migration data, not any substantive considerations. To produce migration estimates for shorter intervals (e.g., less than ten years as in intercensal applications), the basic estimating model must be adjusted accordingly. This adjustment entails extrapolating from the interval for which information is available to produce a ten-year rate, and determining the rate in year 10-X that would yield this ten-year rate. These adjustments necessarily require assumptions about the pattern of net migration during the ten-year
interval. The following discussion suggests two alternative assumptions that can be used to make those adjustments.

The basic estimating model Eq. (7) was:

\[
\frac{I-O}{P_c} = a + c \left( \frac{R_{c+n}}{R_c} \right) - d \left( \frac{\theta_{c+n}}{\theta_c} \right) - b \left( \frac{Q_c}{P_c} \right).
\]

This model was calibrated using data in which \( n = 10 \). For \( n < 10 \) (e.g., to produce intercensal estimates) we must adjust the parameters by extrapolating

\[
\frac{R_{c+n}}{R_c} \text{ and } \frac{\theta_{c+n}}{\theta_c} \text{ to } \frac{R_{c+10}}{R_c} \text{ and } \frac{\theta_{c+10}}{\theta_c},
\]

respectively.

While any number of approaches to this adjustment might be adopted, we suggest either linear or geometric extrapolation. Users may prefer some other approach, and are advised to consider alternatives where a clear rationale exists.

Linear extrapolation is equivalent to assuming a constant volume of migration per unit of time. Specifically:

\[
\frac{R_{c+10}}{R_c} = \frac{10}{n} \left( \frac{R_{c+n}}{R_c} \right) - \frac{10-n}{n}
\]

The corresponding assumption for \( \theta_c \) implies:

\[
\frac{\theta_{c+10}}{\theta_c} = \frac{10}{n} \left( \frac{\theta_{c+n}}{\theta_c} \right) - \frac{10-n}{n}
\]
Plugging these terms into Eq. (7) gives:

\[
\frac{I-O}{P_c} = a - b \left( \frac{Q_c}{P_c} \right) + \frac{10}{n} c \left( \frac{R_{c+n}}{R_c} \right) - \frac{10}{n} d \left( \frac{\theta_{c+n}}{\theta_c} \right) - (c-d) \left( \frac{10-n}{n} \right)
\]  

as the ten-year estimate of net migration using linear extrapolation.

Geometric extrapolation, by contrast, is equivalent to assuming a constant annual rate of change over the ten-year interval. Specifically:

\[
\frac{R_{c+10}}{R_c} = \left( \frac{R_{c+n}}{R_c} \right)^{10/n}
\]  

and

\[
\frac{\theta_{c+10}}{\theta_c} = \left( \frac{\theta_{c+n}}{\theta_c} \right)^{10/n}
\]

Thus, the geometric extrapolation to a ten-year net migration rate becomes:

\[
\frac{I-O}{P_c} = a - b \left( \frac{Q_c}{P_c} \right) + c \left( \frac{R_{c+n}}{R_c} \right)^{10/n} - d \left( \frac{\theta_{c+n}}{\theta_c} \right)^{10/n}
\]

To illustrate the difference between these two assumptions, consider a county with 10,000 residents in year \( c \), which we estimate has experienced a net migration of 5 percent (500 elderly residents) by year \( c + 5 \). Table 4 shows the different patterns of net migration between
years $c$ and $c + 5$ assumed by the linear and geometric approaches, as well as their implications for the net migration pattern by year $c + 10$.

Under both assumptions the five-year average annual volume and rate of net migration are identical; however, these averages are reached in different ways and imply a different end result. Under the linear assumption, 100 net migrants are added to the population every year, implying a ten-year increase of 1000 migrants. As new migrants are added to the population base each year, however, the annual rate of net migration declines monotonically in succeeding years. Under the

<table>
<thead>
<tr>
<th>Year</th>
<th>Linear</th>
<th>Geometric</th>
<th>Linear</th>
<th>Geometric</th>
</tr>
</thead>
<tbody>
<tr>
<td>c+1</td>
<td>100</td>
<td>98</td>
<td>10.00</td>
<td>9.81</td>
</tr>
<tr>
<td>c+2</td>
<td>200</td>
<td>197</td>
<td>9.90</td>
<td>9.81</td>
</tr>
<tr>
<td>c+3</td>
<td>300</td>
<td>297</td>
<td>9.80</td>
<td>9.81</td>
</tr>
<tr>
<td>c+4</td>
<td>400</td>
<td>398</td>
<td>9.71</td>
<td>9.81</td>
</tr>
<tr>
<td>c+5</td>
<td>500</td>
<td>500</td>
<td>9.62</td>
<td>9.81</td>
</tr>
<tr>
<td>c+6</td>
<td>600</td>
<td>603</td>
<td>9.52</td>
<td>9.81</td>
</tr>
<tr>
<td>c+7</td>
<td>700</td>
<td>707</td>
<td>9.43</td>
<td>9.81</td>
</tr>
<tr>
<td>c+8</td>
<td>800</td>
<td>812</td>
<td>9.35</td>
<td>9.81</td>
</tr>
<tr>
<td>c+9</td>
<td>900</td>
<td>918</td>
<td>9.26</td>
<td>9.81</td>
</tr>
<tr>
<td>c+10</td>
<td>1000</td>
<td>1025</td>
<td>9.17</td>
<td>9.81</td>
</tr>
</tbody>
</table>

5-yr average: 100 100 9.81 9.81
10-yr average: 100 102.5 9.58 9.81
geometric assumption, the annual rate is constant but the number of annual migrants increases continuously, until after ten years 1025 migrants have been added to the population. Thus, after ten years the geometric assumption implies both a greater volume and a higher average annual rate of net migration.

Neither of these assumptions is a priori correct: The choice between them has to be based on local circumstances. If the amount of yearly net migration is roughly constant, then linear extrapolation will more clearly fit the actual circumstances. On the other hand, if annual net migration is known to be accelerating, then geometric extrapolation would be more appropriate. In some areas, of course, the pattern of net migration will be so irregular that neither the linear nor the geometric approach will approximate the true pattern.[1] In such cases, an alternative approach would be preferable.

Having made the appropriate extrapolation to a ten-year rate, we must estimate the migration rate for the interval $c$ to $c + n$. Again, the approach used depends upon the assumption made as to the patterns of migration during the interval. Lacking information to the contrary, the most straightforward assumption is that net migration operates uniformly over the decade. Thus, the $n$-year net migration rate would equal $n/10$ times the ten-year rates listed in Eqs. (10) and (13).

[1] For example, migration to and from local areas is often tied to specific events, such as the closing of a military base or the opening of a new retirement community, which would concentrate movement into certain specific years. While for some purposes (e.g., long-term planning) users may want to smooth out such irregularities, for other purposes it may be important to estimate the actual patterns.
CHOICE OF RATES

One strength of our approach is that it affords flexibility to choose the type of migration rate that suits the user's purposes. For example, our approach can be used to generate both beginning- and end-period net migration rates under alternative assumptions as to the pattern of net migration during the decade. Thus, it is up to the user to select the appropriate rate before applying the model.

The beginning-period rate will generally be the preferred alternative since the model predicts it more accurately than the end-period rate. However, accuracy should not be the sole criterion. Although the model produces very accurate estimates in general, there is no assurance that the estimate for any particular area will be as accurate as the average estimate. Moreover, the less accurate estimate may be preferable as long as it is still reasonably close to the true value. Indeed, situations can occur, e.g., a comparison of elderly and non-elderly migration rates, in which the choice of the appropriate rate is determined by the availability of other data. Finally, the user should be aware that the beginning- and end-period rates are not directly comparable, since for a given volume of net migration, end-period rates will always be lower than beginning-period rates.[2]

It is also important to note that this model has been developed for county-level estimates. In principle a similar model could be used to

---

[2] This difference results from the fact that end-period rates incorporate population change during the interval into the denominator. Hence, when net migration is positive, \( m/p > n/p+x \); and when net migration is negative: \( -m/p > -m/p-x \), where \( m \) = number of net migrants, \( p \) = population at beginning of interval, \( x \) = population increase (decrease) during the interval.
estimate net migration for smaller or larger areas. However, those estimates would require an alternative set of parameter estimates and a different set of benchmark data, since the definition of migration used here, intercounty moves, would not be applicable to other geographic units.

AN ILLUSTRATIVE APPLICATION

To demonstrate the computational straightforwardness of our approach and the ease with which it can be implemented, we present here a concrete application. Producing elderly net migration rates with this procedure involves six steps: (1) assembling the raw data; (2) computing the variables to be used in the model from the raw data; (3) determining the appropriate rate and model to use; (4) making an assumption about the pattern of migration during the estimation interval; (5) calculating the ten-year estimate; and (6) adjusting the ten-year figure for the estimation interval. The data used in this illustration cover the 1970 to 1975 period for Denver County, Colorado.

Step 1--Assembling the Raw Data

The six data items required for the procedure, their values for Denver County, and the source from which the data were taken are listed below.
<table>
<thead>
<tr>
<th>Data Item</th>
<th>Value</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Total population of county, 1970</td>
<td>514,678</td>
<td>1970 Census</td>
</tr>
<tr>
<td>2. Population 65+ in 1970 (P_c)</td>
<td>58,786</td>
<td>1970 Census</td>
</tr>
<tr>
<td>3. Population 65+ in 1975 (P_{c+n})</td>
<td>60,232</td>
<td>Census Bureau estimates</td>
</tr>
<tr>
<td>4. Retired Worker Beneficiaries in 1970 (R_{c})</td>
<td>41,132</td>
<td>Social Security Administration</td>
</tr>
<tr>
<td>5. Retired Worker Beneficiaries in 1975 (R_{c+n})</td>
<td>45,695</td>
<td>Social Security Administration</td>
</tr>
<tr>
<td>6. Population aged 55-64 in 1970 (Q_{c})</td>
<td>58,277</td>
<td>1970 Census</td>
</tr>
</tbody>
</table>

All but one of these items are readily available from published sources. Counts of total residents, and of elderly and near-elderly, are routinely published in decennial census reports. Counts of retired worker beneficiaries are contained in the Social Security Administration's annual report on the number of recipients by county of residence. The only item not routinely published is the intercensal estimates of elderly residents by county, but these are available from the Census Bureau and may also be obtained from state and local sources.

Step 2—Computing Input Variables for the Model

Once the raw data are assembled, they must then be used to compute the three measures used in the model: the recipient ratio and the change-in-coverage and aging-in-place variables. Two of these measures are computed directly from raw data; the third, the change-in-coverage variable, can be computed after first constructing coverage ratios for years c and c + n. The computation steps are listed below.
<table>
<thead>
<tr>
<th>Input Measure</th>
<th>Formula</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Recipient Ratio</td>
<td>( \frac{\frac{R_{c+n}}{R_c}}{Q_c} )</td>
<td>( \frac{45.695}{41.132} = 1.111 )</td>
</tr>
<tr>
<td>2. Aging in Place</td>
<td>( \frac{Q_c}{P_c} )</td>
<td>( 58.277 )</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( \frac{58.786}{58.786} = 0.991 )</td>
</tr>
<tr>
<td>3. Coverage Change</td>
<td>( \frac{\eta_{c+n}}{\eta_c} )</td>
<td>( \frac{.759}{.700} = 1.084 )</td>
</tr>
<tr>
<td>4. Recipient Rate-1970 (( \theta_c ))</td>
<td>( \frac{R_c}{P_c} )</td>
<td>( \frac{41.132}{58.786} = .700 )</td>
</tr>
<tr>
<td>5. Recipient Rate-1975 (( \theta_{c+n} ))</td>
<td>( \frac{R_{c+n}}{P_{c+n}} )</td>
<td>( \frac{45.695}{60.232} = .759 )</td>
</tr>
</tbody>
</table>

Step 3--Determining the Appropriate Rate and Model

At this step the user must decide whether to estimate a beginning- and/or end-period net migration rate and whether to use the parameters estimated for all counties or those estimated for specific strata. In general, the beginning-period rate estimated with the stratum-specific parameters will be the most accurate estimate and hence the preferred choice. However, as we have already noted, accuracy is not the sole criterion here and other considerations might dictate a more appropriate alternative.

Step 4--Making Assumptions About Migration Patterns

Once the appropriate model has been selected, the user must determine whether the actual pattern of migration between \( c \) and \( c + n \) is best approximated with the linear, the geometric, or some other extrapolation assumption. Barring information about local circumstances to the contrary, the choice will generally be between the linear and geometric approaches. The annual counts of RWB available from the
Social Security Administration can be used to inform this choice.
Listed below, for example, are the annual changes in RWB for Denver
County between 1970 and 1975.

<table>
<thead>
<tr>
<th>Year</th>
<th>Change in Number of RWB</th>
</tr>
</thead>
<tbody>
<tr>
<td>1970-71</td>
<td>+ 625</td>
</tr>
<tr>
<td>1971-72</td>
<td>+ 693</td>
</tr>
<tr>
<td>1972-73</td>
<td>+ 862</td>
</tr>
<tr>
<td>1973-74</td>
<td>+ 1268</td>
</tr>
<tr>
<td>1974-75</td>
<td>+ 1115</td>
</tr>
</tbody>
</table>

With the exception of the 1974-1975 period, the change in
beneficiaries appears to be increasing—consistent with the geometric
approach's assumption of an increasing volume of net migrants and a
constant rate. However, the pattern is not monotonic, so the evidence
is ambiguous.

Step 5—Calculating a Ten-Year Net Migration Estimate

Having chosen the rate to be estimated, the parameters to use, and
the appropriate extrapolation assumption, the user now needs only to
plug the computed input variables and estimated parameters into the
appropriate model to calculate a ten-year elderly net migration
estimate. To illustrate, we calculate below the ten-year beginning-
period rate for Denver County using the stratum-specific parameters[3]
under both the linear and geometric extrapolation assumptions.

[3]Since Denver's population in 1970 was 514,678, the parameters
are for the 500,000-999,999 stratum.
Under the linear assumption, the ten-year rate is estimated with Eq. (10):

\[
\frac{I-0}{P} = a - b \left( \frac{Q}{P} \right) + \frac{10}{n} c \left( \frac{R_c}{R} \right)^n - d \left( \frac{\theta_c + n}{\theta} \right)^n (c-d) \left( \frac{10-n}{n} \right)
\]

Substituting the input measures with \( n = 5 \), \( a = 1.120 \), \( b = .849 \), \( c = .726 \), and \( d = 1.102 \) yields a ten-year migration rate of -12.1 percent:

\[
\frac{I-0}{P} = 1.120 - .849(.991) + 2(.726)(1.111) - 2(1.102)(1.084) - 1(.726-1.102)
\]

\[
= 1.120 - .841 + 1.613 - 2.389 + .376
\]

\[
= -.121
\]

Under the geometric assumption, the ten-year rate is estimated with Eq. (13):

\[
\frac{I-0}{P} = a - b \frac{Q}{P} + c \left( \frac{R_c}{R} \right)^n - d \left( \frac{\theta_c + n}{\theta} \right)^n \frac{10}{n}
\]

Substituting the same input measures and parameter values also yields a ten-year migration rate of -12.0 percent:

\[
\frac{I-0}{P} = 1.120 - .849(.991) + .726(1.111^2) - 1.102(1.084^2)
\]

\[
= 1.120 - .841 + .896 - 1.295
\]

\[
= -.120
\]
Step 6--Adjusting to the Estimation Interval

The final step is to adjust the estimated ten-year rate to the actual estimation interval, in this case 5 years. Lacking local information to the contrary, we assume that elderly net migration operates uniformly over the decade. Thus, the rate in year $c + n$ equals $n/10$ times the ten-year rate, or 6.05 percent under the linear assumption, and 6 percent under the geometric consumption.
Appendix A

DERIVATION OF THE MODEL

BEGIN-PERIOD RATES

Consider n counties and a fixed interval of time. For each county we can define a number of elements where the subscript "$i" is used to denote the specific county and the superscript: "b" and "e" denote beginning and end period elements respectively. We define:

- $P_i^b$ Number of elderly at the beginning of the time interval in county $i$.
- $P_i^e$ Number of elderly at the end of the time interval in county $i$.
- $R_i^b$ Number of recipients at the beginning of the time interval in county $i$.
- $R_i^e$ Number of recipients at the end of the time interval in county $i$.
- $P_i$ Number of people who "age in place" during the interval.
- $N_i$ Net number of elderly migrating into county $i$, and present at the end of the interval.

Of those $P_i^b$ people who were present at the beginning of the interval, a certain fraction, call it $d_i^b$ will survive and be present at the end. Similarly, of those $P_i$ people who aged in place, a certain fraction, call it $d_i$ will survive and be present at the end. By definition of $N_i$, we can write the balance equation

$$P_i^e = d_i^b P_i^b + d_i P_i + N_i.$$
Dividing through and rearranging terms, we obtain the relation

\[
\frac{N^i}{p^b_i} = \frac{p^c_i}{p^b_i} - d_i \frac{p^c_i}{p^b_i} - d_i.
\] (1)

We now define two "recipiency rates" \( \vartheta^b_i \), \( \vartheta^c_i \) as follows:

\[
\vartheta^b_i = \frac{R_i^b}{p^b_i}, \quad \vartheta^c_i = \frac{R_i^c}{p^c_i}
\]

and we denote the change in recipiency rate thus:

\[
\Delta \vartheta_i = \vartheta^b_i - \vartheta^c_i.
\]

Some simple algebra yields the identity

\[
\frac{p^c_i}{p^b_i} = \frac{R_i^c}{R_i^b} - \frac{\Delta \vartheta_i}{\vartheta^b_i} \frac{p^c_i}{p^b_i}.
\] (2)

We now define \( D \), the average elderly population growth rate for all counties, as follows:

\[
D = \frac{1}{n} \sum_{i=1}^{n} \frac{p^c_i}{p^b_i}.
\]

If all we knew were the numbers of recipients, the recipiency rates at the beginning and end of the interval, and \( D \), Eq. (2) suggests that we estimate the county specific growth rate, \( E_i \), as

\[
E_i = \frac{R_i^c}{R_i^b} - \frac{\Delta \vartheta_i}{\vartheta^b_i} D.
\] (3)

From Eq. (2) and (3), we get

\[
\frac{p^c_i}{p^b_i} - E_i = \frac{\Delta \vartheta_i}{\vartheta^b_i} \left( \frac{p^c_i}{p^b_i} - D \right).
\]
Now, suppose \( \Delta \Theta \) is small relative to \( \Theta \) everywhere, i.e., \( |\Delta \Theta| < \Theta \) for some small \( \varepsilon > 0 \), then

\[
\frac{1}{n} \sum \left( \frac{p^e \hat{b}}{p^b \hat{b}} - E_i \right)^2 \leq \frac{\varepsilon}{n} \sum \left( \frac{p^e \hat{b}}{p^b \hat{b}} - D \right).
\]

This argument suggests that for appropriate value of \( D \), the expression

\[
\frac{R^e \hat{b}}{R^b \hat{b}} - D \frac{\Delta \Theta \hat{b}}{\Theta \hat{b}}
\]

is an accurate estimate of \( \frac{p^e \hat{b}}{p^b \hat{b}} \).

Now assume the survival rates \( d^b \hat{b} \) and \( d \hat{b} \) are nearly constant from county to county. Combining this assumption with (1), we obtain an estimator

\[
\frac{N^e \hat{b}}{P^b \hat{b}} \approx \frac{R^e \hat{b}}{R^b \hat{b}} - D \frac{\Delta \Theta \hat{b}}{\Theta \hat{b}} - d \frac{p^e \hat{b}}{p^b \hat{b}} - d^b
\]

or equivalently,

\[
\frac{N^e \hat{b}}{P^b \hat{b}} \approx \frac{R^e \hat{b}}{R^b \hat{b}} - D \frac{\Delta \Theta \hat{b}}{\Theta \hat{b}} - d \frac{p^e \hat{b}}{p^b \hat{b}} - (d-D)
\]

whose parameters we estimate by O.L.S. In practice, we also estimate a coefficient for the leading term \( R^e \hat{b}/R^b \hat{b} \), and our estimate is usually different from 1, probably a result of correlated errors.

END-PERIOD RATES

Suppose \( \Phi \) is an estimator of \( N^e \hat{b}/P^b \hat{b} \), the begin period net migration rate, and we want to specify an estimator of \( N \hat{b}/P^e \hat{b} \), the end-period net migration rate. Take any constant, \( a \), and write
\[
\frac{N_i}{p_i^e} - a \phi_i = \frac{p_i^b}{p_i^e} \frac{N_i}{p_i^b} - \frac{p_i^b}{p_i^e} \phi_i + \frac{p_i^b}{p_i^e} \phi_i - a \phi_i \\
= \frac{p_i^b}{p_i^e} \left( \frac{N_i}{p_i^b} - \phi_i \right) + \left( \frac{p_i^b}{p_i^e} - a \right) \phi_i.
\]

Now, if \( a \) is "close to" \( \frac{p_i^b}{p_i^e} \) for most counties, then since \( \phi_i \) is "close to" \( N_i/p_i^b \) for most counties already, the term \( a \phi_i \) will be "close to" \( N_i/p_i^e \) for most counties as well. (These statements can be made rigorous, but we will omit the details here.) Thus, we can transform the estimator of the begin-period net migration rate into an estimator of the end-period net migration rate by multiplying the former by a suitable constant. In our case, this shows that we can specify the same linear model for the end period, and re-estimate the parameters using the end-period rate as the dependent variable.
## Appendix B

**Prediction Equations for Beginning and End-Peiod Migration Rates**

<table>
<thead>
<tr>
<th>Size of County</th>
<th>Constant</th>
<th>Reciprocality Ratio</th>
<th>Change in Coverage</th>
<th>Aging in Place</th>
<th>Summary Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coefficient</td>
<td>Coefficient</td>
<td>Coefficient</td>
<td>Coefficient</td>
<td>Coefficient</td>
</tr>
<tr>
<td>1,000,000+</td>
<td>1.137</td>
<td>.720</td>
<td>99.85</td>
<td>-1.144</td>
<td>-794</td>
</tr>
<tr>
<td>500-999,999</td>
<td>1.120</td>
<td>.726</td>
<td>72.23</td>
<td>-1.102</td>
<td>-849</td>
</tr>
<tr>
<td>250-499,999</td>
<td>1.312</td>
<td>.754</td>
<td>37.37</td>
<td>-1.288</td>
<td>-807</td>
</tr>
<tr>
<td>100-249,999</td>
<td>.984</td>
<td>.684</td>
<td>121.35</td>
<td>-1.042</td>
<td>-786</td>
</tr>
<tr>
<td>50-99,999</td>
<td>1.030</td>
<td>.715</td>
<td>85.50</td>
<td>-1.059</td>
<td>-781</td>
</tr>
<tr>
<td>25-49,999</td>
<td>.807</td>
<td>.701</td>
<td>208.97</td>
<td>-1.043</td>
<td>-857</td>
</tr>
<tr>
<td>10-24,999</td>
<td>.763</td>
<td>.603</td>
<td>142.87</td>
<td>1.25</td>
<td>-724</td>
</tr>
<tr>
<td>&lt; 10,000</td>
<td>.567</td>
<td>.531</td>
<td>79.40</td>
<td>-1.563</td>
<td>-736</td>
</tr>
</tbody>
</table>

### Beginning Period Rates

### End Period Rates

<table>
<thead>
<tr>
<th>Size of County</th>
<th>Coefficient</th>
<th>Coefficient</th>
<th>Coefficient</th>
<th>Coefficient</th>
<th>R²</th>
<th>F</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>1,000,000+</td>
<td>.277</td>
<td>1.92</td>
<td>.411</td>
<td>22.63</td>
<td>-3.07</td>
<td>8.12</td>
<td>.973</td>
</tr>
<tr>
<td>500-999,999</td>
<td>.331</td>
<td>3.76</td>
<td>.349</td>
<td>21.73</td>
<td>-3.53</td>
<td>5.08</td>
<td>.976</td>
</tr>
<tr>
<td>250-499,999</td>
<td>.631</td>
<td>18.80</td>
<td>.400</td>
<td>4014</td>
<td>-5.87</td>
<td>26.70</td>
<td>.961</td>
</tr>
<tr>
<td>100-249,999</td>
<td>.386</td>
<td>12.62</td>
<td>.324</td>
<td>40.68</td>
<td>-3.83</td>
<td>16.63</td>
<td>.825</td>
</tr>
<tr>
<td>50-99,999</td>
<td>.480</td>
<td>26.91</td>
<td>.369</td>
<td>53.02</td>
<td>-4.63</td>
<td>34.05</td>
<td>.842</td>
</tr>
<tr>
<td>25-49,999</td>
<td>.393</td>
<td>23.71</td>
<td>.277</td>
<td>49.28</td>
<td>-3.32</td>
<td>31.11</td>
<td>.821</td>
</tr>
<tr>
<td>10-24,999</td>
<td>.509</td>
<td>52.56</td>
<td>.374</td>
<td>81.04</td>
<td>-4.56</td>
<td>63.68</td>
<td>.875</td>
</tr>
<tr>
<td>&lt; 10,000</td>
<td>.693</td>
<td>42.69</td>
<td>.463</td>
<td>61.97</td>
<td>-4.95</td>
<td>54.41</td>
<td>.859</td>
</tr>
</tbody>
</table>
BIBLIOGRAPHY


