A RAND NOTE

TECHNIQUES FOR ANALYSIS OF MIGRATION-HISTORY DATA

Julie DaVanzo

May 1982

N-1824-AID/NICHD

Prepared for

The Agency for International Development
The National Institute of Child Health
and Human Development

Rand
SANTA MONICA, CA. 90406

A RAND NOTE

TECHNIQUES FOR ANALYSIS OF MIGRATION-HISTORY DATA

Julie DaVanzo

May 1982

N-1824-AID/NICHD

Prepared for

The Agency for International Development
The National Institute of Child Health
and Human Development

Rand
SANTA MONICA, CA. 90406

PREFACE

Many data sets now available provide retrospective or longitudinal information on human migration and related variables. Such data can provide a richer insight into patterns, determinants, and consequences of migration than is obtainable from data collected in censuses and in most surveys. This Note discusses methods for analyzing migration using life-history or longitudinal data to take advantage of their unique features.

The Note is a revised version of a paper prepared for a technical working group meeting on migration and urbanization organized by the Population Division of the Economic and Social Commission for Asia and the Pacific (ESCAP). The meeting was held at ESCAP offices in Bangkok, December 1-5, 1981. The Population Division of ESCAP, in collaboration with member countries, has developed a set of survey manuals for national migration surveys to be conducted in the ESCAP region in the early 1980s.[1] A key component of the ESCAP survey instrument is a life-history questionnaire that elicits a retrospective accounting of migration and related life events. The purpose of the December 1981 technical working group meeting was to assist ESCAP in formulating a plan for analyzing data from the national migration surveys.

The issues and methods discussed here are applicable not only to the ESCAP life-history questionnaire, but also to other life-history data (for example, the female and male retrospective life histories in Rand's Malaysian Family Life Survey[2] and the Rand-INCAP Guatemalan

[1] ESCAP, National Migration Surveys, Survey Manual II: The Core Questionnaire, United Nations, New York, 1980.

[2] William P. Butz and Julie DaVanzo, The Malaysian Family Life Survey: Summary Report, The Rand Corporation, R-2351-AID, March 1978.

Survey[3]) and to such longitudinal data sets as the University of Michigan Panel Study of Income Dynamics (the core data set being analyzed by Rand's Population Research Center).

This work draws on research supported by Grant No. OTR-G-1822, from the U.S. Agency for International Development, which supports Rand's Family in Economic Development Center; and by Grant P50-HD12639 from the Center for Population Research, National Institute of Child Health and Human Development, U.S. Department of Health and Human Services, which supports Rand's Population Research Center. Partial support was furnished by an honorarium from ESCAP.

---

[3] Henry L. Corona, INCAP-Rand Guatemala Survey: Code Book and User's Manual, The Rand Corporation, P-6181, August 1978.

SUMMARY

Migration-history data have several advantages over other more conventional types of migration data collected in censuses and surveys. First, migration histories record a higher proportion of the moves people make. Second, one can choose the time interval over which migration is measured to best suit the purpose at hand. Third, one can study migration patterns and correlates in different time periods, and assess and analyze changes over time. Finally, when combined with life-history data on variables that may affect or be influenced by migration decisions, such as occupation, industry, marital status, and fertility, migration-history data allow much richer analyses of determinants and consequences of migration than can be done with census or most survey data. These last two features are especially important for policy applications, since they allow examination of the interrelation between migration and social and economic change.

The richness of life-history data is often matched by their complexity. The number of moves, and hence number of records, in the migration history will vary among respondents. For some purposes, moves (perhaps within a specific time frame or age range) may be the appropriate units of analysis; for other purposes, individuals should be the sample units. Furthermore, time periods to which explanatory variables refer can, and should, be linked to the timing of migration.

This Note discusses alternative techniques for analyzing migration and its determinants and consequences using migration-history and life-history data. It discusses methods of processing and analyzing life-history data for four types of studies:

- Descriptions of patterns of migration and how they have changed over time (Sec. II).
- Analyses of determinants of migration (Sec. III).
- Analyses of choice among alternative types of moves (e.g., rural-urban vs. rural-rural; North-to-South vs. North-to-East; return vs. onward)(Sec. IV).
- Studies of individual-level consequences of migration (Sec. V).

For each topic, the Note discusses the general types of data desired or required, how these should be "retrieved" from the life history, and what analytic techniques are most appropriate. The discussions cover some simple, old techniques (such as cross-tabulations and ordinary least squares regression) and some sophisticated new ones (such as regression-switching models and hazard models). The concluding section discusses implications of these recommendations for data processing.

ACKNOWLEDGMENTS

For their helpful comments, the author thanks Rand colleagues Terry Fain, Gus Haggstrom, James Hammitt, Lee Lillard, Mark Menchik, Peter Morrison, William Rogers, and Ross Stolzenberg; conference discussants Daniel Courgeau (Institut National d'Etudes Demographiques, Paris) and Penporn Tirasawat (Institute of Population Studies, Chulalongkorn University, Bangkok); and other conference participants.

INTENTIONALLY
LEFT BLANK

CONTENTS

PREFACE iii

SUMMARY v

ACKNOWLEDGMENTS vii

Section

I. INTRODUCTION 1

II. DESCRIBING PATTERNS OF MIGRATION 4
- Possible Biases in Retrospective Data 4
- Describing Migration Patterns and Trends 6

III. ANALYSES OF DETERMINANTS OF MIGRATION 9
- Conceptual Model 9
- Estimation Techniques 11
  - Cross Tabulations 11
  - Multivariate Analysis with Dichotomous Dependent Variables 12
  - Hazard Models 13
- Defining the Dependent Variable 16
- Explanatory Variables 17
- Time-Subscripting the Explanatory Variables 18
- Subsamples 19
- Summary 20

IV. ANALYSIS OF CHOICE AMONG TYPES OF MOVES 21

V. ANALYSIS OF CONSEQUENCES OF MIGRATION 24
- Types of Comparisons 24
- Methods 25

VI. IMPLICATIONS FOR DATA PROCESSING 29

REFERENCES 33

I. INTRODUCTION

Migration-history data have several advantages over other more conventional types of migration data collected in censuses and most surveys. First, migration histories record a higher proportion of the moves people make. Second, one can choose the time interval over which migration is measured to best suit the purpose at hand. In addition, one can study migration patterns and correlates in different time periods, and can assess and analyze changes over time. This last feature is especially important for policy applications, since it means that the interrelation between migration and social and economic change can be examined.

More elaborate surveys (such as ESCAP's and Rand's) collect not only detailed migration-history data, but also life-history data on variables that may affect or be influenced by migration decisions, such as occupation, industry, wages, marital status, and fertility. Such life-history data are superior to Census data for investigating determinants and consequences of migration. For example, determinants can be measured at or shortly before the time of the migration, not merely afterward. Similarly, with life-history data, consequences can be assessed over a specific period following migration rather than only at the time of interview.

The richness of life-history data is often matched by their complexity. The number of moves, and hence number of records, in the migration-history will vary among respondents. For some purposes, moves (perhaps within a specific time frame or age range) may be the

appropriate units of analysis; for other purposes, individuals or person-year observations should be the sample units. Furthermore, time periods to which explanatory variables refer can, and should, be linked to the timing of migration.

This Note discusses alternative techniques for analyzing migration and its determinants and consequences using migration-history and life-history data, such as those collected in the ESCAP National Migration Surveys or the Rand Mayalsian Family Life Survey.[1] The Note discusses methods of processing and analyzing life-history data for four types of studies:

- Descriptions of patterns of migration and how they have changed over time (Sec. II).
- Analyses of determinants of migration (Sec. III).
- Analyses of choice among alternative types of moves (e.g., rural-urban vs. rural-rural; North-to-South vs. North-to-East; return vs. onward)(Sec. IV).
- Studies of individual-level consequences of migration (Sec. V).

Substantive aspects of these issues (e.g., the pros and cons of alternative definitions of migration, hypotheses regarding particular determinants and consequences) are discussed in other papers prepared for the working group meeting. This paper concentrates on methodological issues common to many of these. For each topic, it discusses the general types of data desired or required, how these should be "retrieved" from the life history, and what analytic

[1] Although the discussion is in terms of retrospective data, most of the points are equally applicable to longitudinal data (e.g., the Panel Study of Income Dynamics).

techniques are most appropriate. The discussions cover some simple, old techniques (such as cross-tabulations and ordinary least squares regression) and some sophisticated, new ones (e.g., regression-switching models and hazard models). The concluding section discusses implications of these recommendations for data processing.[2]

---

[2] For additional discussions of the strengths and weaknesses of migration-history and residential-history data, and for presentations of a number of studies based on such data, see Robin J. Pryor (ed.), Residence History Analysis, Studies in Migration and Urbanization No. 3, Department of Demography, Research School of Social Science, Australian National University, Canberra, 1979.

II. DESCRIBING PATTERNS OF MIGRATION

A key advantage of migration-history data is their superiority for use in describing migration rates and patterns in the past and how these have changed over time.

POSSIBLE BIASES IN RETROSPECTIVE DATA

For such a purpose the data are (potentially) subject to the biases typical of retrospective data:

1. The sample will not be a random one of all persons in a particular birth cohort of interest, because some members of this cohort will have died or emigrated before the date of the survey and their migration experiences will not be recorded. (This corresponds to "panel mortality" or "sample decay" in a prospective study.) The important question regarding the representativeness of the sample is whether the migration experiences of deceased or emigrant members of the cohort differed markedly from those of surviving, resident members.
2. Whenever the sample criteria include an upper age limit, the age range to which the data refer shrinks for dates further in the past. For example, a sample aged 15-64 at the time of the survey will give no information on persons who were older than age 44 twenty years before the survey. However, since most migration activity occurs before age 30, relatively few moves will be missed. Nonetheless, analyses must control for age; for dates many years before the survey, the sample will contain

relatively more people of prime migration ages than it will for dates near to the time of the survey.

3. Retrospective data are subject to recall error. Respondents may forget events that took place many years before the survey or may misplace their dates.[1] Even if there is no underreporting, systematic mistiming of events (e.g., reporting events as occurring more recently than they actually did) can yield spurious trends.[2]

4. A sample that is nationally representative at the time of the survey should, subject to the biases discussed in (1) through (3) above, be representative of the national population ten, twenty, or thirty years earlier. However, if the sample is a stratified one of particular areas, as many samples for the life-history data are, it will be representative of those particular (destination) areas for the time of the survey but will not necessarily provide random samples of the populations in earlier years of the origin areas from which the migrants came. This problem will be most serious for small geographic units (e.g., particular towns) and should become less important as the units become larger or broader (e.g., urban/rural strata).

---

[1] The likelihood of serious mistiming of migrations is reduced when the survey includes questions about related life events (e.g., marriage, births) to which migrations can be related.

[2] Joseph E. Potter, "Problems in Using Birth History Analyses to Estimate Trends in Fertility," Population Studies, Vol. 31, No. 2, July 1977.

DESCRIBING MIGRATION PATTERNS AND TRENDS

A useful way to describe migration trends using migration-history data is to compute, for a particular definition of migration, the triangular matrix showing migration rates for each possible age group in each time period. For example, for a survey done in 1980, one could describe five-year migration rates for all possible time periods and age groups in a matrix like that in Table 1.

Such a matrix enables one to identify age, period, and (birth) cohort effects.[3] The columns of such a matrix show the age patterns of migration rates in different time periods. The rows show how

Table 1

TRIANGULAR MATRIX OF MIGRATION RATES BY AGE AND DATE

Age at Beginning of Migration Interval	Migration Interval								
	1935-1939	1940-1944	1945-1949	1950-1954	1955-1959	1960-1964	1965-1969	1970-1974	1975-1979
15-19	X	X	X	X	X	X	X	X	X
20-24		X	X	X	X	X	X	X	X
25-29			X	X	X	X	X	X	X
30-34				X	X	X	X	X	X
35-39					X	X	X	X	X
40-44						X	X	X	X
45-49							X	X	X
50-54								X	X
55-59									X

[3] Because period = birth year + age, only two of these three effects are identifiable without making particular assumptions about their forms (Stephen E. Fienberg and William M. Mason, "Identification and Estimation of Age-Period-Cohort Models in the Analysis of Discrete Archival Data," in Karl F. Schuessler (ed.), Sociological Methodology, 1979, Jossey-Bass Publishers, San Francisco, Washington, and London, 1978.

migration rates have varied over time, holding age constant. The upper-right to lower-left diagonals trace the experiences of actual birth cohorts. Separate matrices could be calculated for population subgroups, e.g., stratified by sex or ethnicity, to reveal differences in migration propensities by these characteristics.

What migration statistics should go in the body of the table? The answer depends on the particular research or policy question being asked and is complicated by the fact that many migrants move more than once. If concern is with the amount of population redistribution taking place, one could compare place of residence (for a particular type of geographic unit, e.g., district) at the beginning and end of each five-year interval. Dividing (a) the sum of the number of people living in a different place at the end of the five-year period than at the beginning by (b) the number of people in the cohort, will yield a statistic showing the propensity of initial residents to change their area of residence in the five-year period. Alternatively, the numerator could count the number of people who migrated at least once, even if by the end of the five-year period they had returned to the place where they lived at the beginning. Such a statistic measures the propensity of people to migrate. Other types of rates are possible too, e.g., rates of rural-rural and rural-urban migration (each defined with respect to the rural population in the beginning year). These and other possibilities--for example, using information on person-years of

residence in an area--are discussed in papers prepared for the working group meeting by Willekens, Courgeau, and Rogers.[4]

The data can also be used to indicate the proportions of people in the sample who have ever migrated and the shares of these who are repeat or return migrants. Radloff's analysis of the migration-history data in the Malaysian Family Life Survey illustates these possibilities.[5]

---

[4] Frans Willekens, "Identification and Measurement of Spatial Population Movements"; Daniel Courgeau, "Methods of Linking Migration Statistics Collected from National Surveys with Those from Population Censuses"; and Andrei Rogers, "The Migration Component in Subnational Population Projections"; papers presented at ESCAP Technical Working Group Meeting on Migration and Urbanization, Bangkok, December 1981.

[5] Scott Radloff, "Measuring Migration: A Sensitivity Analysis of Traditional Measurement Approaches Based on the Malaysian Family Life Survey," Ph.D. dissertation, Brown University, Providence, R.I., 1982.

III. ANALYSES OF DETERMINANTS OF MIGRATION

An advantage of migration-history data that are combined with other life-history data is their capacity to elucidate why some individuals migrate but others do not. A myriad of factors may affect migration decisions. Some of these are characteristics of the individual; others pertain to his immediate or extended family; still others may exert their influence at the community level.

CONCEPTUAL MODEL

The basic premise underlying many micro-level models of (voluntary) migration decisionmaking is that individuals (or households) migrate in the expectation of being better off by doing so.[1] Alternatively stated, persons choose to migrate if they believe the benefits will outweigh the costs. The other side of the coin is that other individuals do not migrate because, to the extent they have thought about it, the costs of migration appear to outweigh the benefits.

The benefits and costs of migration may accrue over some period of time. They will include both economic considerations, such as obtaining a (better-paying) job, and noneconomic ones, such as being near friends and relatives. The relevant conceptual variable compares expectations about these factors in the future at both origin and alternative destinations. Some complex procedures may enable researchers to come closer to approximating the expected net benefits from migration than has hitherto been possible.[2] A much simpler approach is to view

---

[1] For example, see many of the papers in Gordon F. DeJong and Robert W. Gardner (eds.), Migration Decision Making: Multidisciplinary Approaches to Microlevel Studies in Developed and Developing Countries, Pergamon Press, 1981.

[2] It is unlikely that any survey will ever contain all the

migration decisions as being determined by characteristics of individuals and of their situations before migration. This will avoid the chicken-and-egg dilemma of determining the direction of causation that can arise when post-migration characteristics are considered as possible influences on migration decisions.

Even with such a simplification, a number of potentially confounding issues remain. One is that many people move more than once. Which move should be considered? Are the determinants of repeat migration different from those of primary migration? Are there unobserved differences between "movers" and "stayers?" (This has become known as the problem of "unobserved heterogeneity.") Another potential difficulty is that the variables influencing migration decisions change over time, sometimes with important consequences. This raises the question of when migration determinants should be measured.

These issues have not arisen in many previous analyses of migration. Typically, migration data allow identification of one migration and measure explanatory variables only at one time point. The richness of life-history data allows more. This presents researchers both opportunities and complications. The statistical "technology" for handling these new problems is rapidly developing, but it tends to be complex and expensive.[3] Some of these new methods are discussed

---

information required to construct an appropriate empirical analog to the relevant conceptual variable. For a discussion of these issues, see Julie DaVanzo, "Microeconomic Approaches to Studying Migration Decisions," in DeJong and Gardner, pp. 101-112.

[3] See, e.g., Nancy B. Tuma, Michael T. Hannan, and Lyle P. Groenvald, "Dynamic Analyses of Event Histories," American Journal of Sociology, Vol. 84, No. 4, 1979; and Christopher J. Flinn and James J. Heckman, "New Methods for Analyzing Event History Data," discussion paper, Economics Research Center, National Opinion Research Center, Chicago, 1981.

briefly ahead. Most of the section deals with simpler, often descriptive, techniques. It discusses estimation techniques, measurement of the migration variable, definition and "time subscripting" of explanatory variables, and stratification of the data into subsamples.

ESTIMATION TECHNIQUES

Analyses of the relationships between migration and explanatory variables should ultimately use multivariate estimation techniques, since a variety of factors influence migration decisions and their effects may not be independent of one another.[4]

Cross Tabulations

Cross-tabular analyses are useful for preliminary and complementary analyses. For example, cross-tabs can be used to compare the average values of explanatory variables for migrants and nonmigrants or to compute the proportion of migrants for different values of an explanatory variable. Such analyses should not only examine the values of these means but should also perform the relevant statistical tests (t-tests) to determine whether apparent differences are actually statistically significant. The analyst should keep in mind, however, that bivariate tabulations frequently yield misleading inferences about the relative importance of a particular explanatory variable because other relevant explanatory variables are not held constant. Examination

[4] Most statistical techniques assume that error terms are uncorrelated. If data on different individuals in the same family or on different time periods for a given individual are pooled, this assumption will be violated. The resulting estimates will be unbiased, but their standard errors will be biased downward.

of all possible combinations of explanatory variables can be tedious (and voluminous). Multivariate analysis usually provides a more concise format for assessing the independent influences of explanatory variables. Nonetheless, tabulations can reveal nonlinearities and interactions that may otherwise not be investigated in multivariate analysis. The two forms of analysis can and should be used complementarily.

Multivariate Analysis with Dichotomous Dependent Variables

When the time interval over which migration is being measured is fixed (e.g., whether the person migrated between 1965 and 1969), the dependent variable can be characterized by a 0-1 dummy. Appropriate multivariate techniques for 0-1 dependent variables include logit and probit analysis.[5] These are maximum likelihood, nonlinear techniques that constrain predicted values of the dependent variable to be within the 0-1 range and accommodate several other features of these noncontinuous dependent variables. Nonetheless, even though it does not have all these agreeable statistical properties, ordinary least squares regression analysis (OLS) almost always yields estimates of the significance and direction of relationships similar to those indicated by the more sophisticated techniques.[6] This feature, together with its

[5] Log-linear models are also sometimes used when the dependent variable is qualitative. These models require that all explanatory variables be categorical rather than continuous. This is not always an appropriate representation of many variables hypothesized to influence migration. Where this representation is appropriate, log-linear models are ideal for investigating interactions among variables. (For a relatively nontechnical introduction to log-linear models, see Stephen E. Fienberg, The Analysis of Cross-Classified Categorical Data, The MIT Press, Cambridge, Mass., and London, 1977.)

[6] Gus Haggstrom, "Logistic Regression and Discriminant Analysis by Ordinary Least Squares," The Rand Corporation (forthcoming).

lower computation cost, makes OLS appropriate for preliminary multivariate analyses.

Hazard Models

Another set of statistical techniques, developed fairly recently by biostatisticians, mathematical sociologists, and econometricians, are even more appropriate for the analysis of event history or longitudinal data. The techniques are known by many different labels: survival, renewal, semi-Markov, hazard, time-to-failure, reliability, life-testing, waiting-time, event history, and continuous-time stochastic processes.[7] Their common feature is that they enable investigation of the timing of events. For the analysis of migration, duration of residence becomes a feature of the dependent variable, rather than merely a right-hand-side, explanatory variable.[8]

These models provide an approach to analyzing survival data when the risks (called hazards[9]) vary among individuals.[10] They can be

[7] See, for example, J. D. Kalbfleisch and R. L. Prentice, The Statistical Analysis of Failure Time Data, John Wiley and Sons, New York, 1980; Ralph B. Ginsberg, "Timing and Duration Effects in Residence Histories and Other Longitudinal Data: I. Stochastic and Statistical Models," Regional Science and Urban Economics, Vol. 9, North Holland Press, 1979; Tuma, Hannan, and Groenvald; and Flinn and Heckman.

[8] For an application of these techniques to migration, see Michael C. Keeley, "Migration as Consumption: The Impact of Alternative Negative Income Tax Programs," in J. Simon and J. DaVanzo (eds.), Research in Population Economics, Vol. II, JAI Press, Greenwich, Conn., 1979. For an application to residential mobility, see Mark D. Menchik, "Residential Mobility and Public Policy," in W.A.V. Clark and E. G. Moore, Urban Affairs Annual Reviews, Vol. 19, Sage Publications, Beverly Hills, Calif., 1980.

[9] The hazard is the conditional probability density of occurrence at a particular duration (i.e., given survival to that duration).

[10] E.g., the risk of divorce vis-a-vis survival in a marriage, the risk of conception vis-a-vis survival in the nonpregnant state, the risk of mobility vis-a-vis survival (stay) in a residence.

viewed as a multivariate form of life-table analysis. For migration, one would consider the risk of migration vis-a-vis the duration of stay in a particular location. The researcher can specify the way in which the hazard is expected to vary with duration of time in the state. For example, Menchik concludes that a hazard function based on the duration-dependent logistic distribution best fits his data on residential mobility. (In his analysis of the determinants of length of stay in a residence following the introduction of a housing subsidy program, the risk of mobility first increases and then decreases, peaking at around 2 years duration.)

A particular advantage of hazard models is that they can handle both open and closed intervals. For example, some individuals may have already migrated before the time of the survey. Others may yet migrate but observations on them are "censored" by the date of the survey.

Many applications of hazard models deal only with covariates that are fixed at the beginning of the period.[11] For example, in an application to divorce, this would mean that only those explanatory variables that refer to the time of the marriage (e.g., age at marriage, education, religion, premarital pregnancy) could be considered; factors that changed after that time, such as births of children, would not be considered. For migration, this assumption would limit the analyst who is studying determinants of the decision to leave an area to characteristics of the individual when he arrived in the area (at birth for some) or to whenever the analyst arbitrarily chose to "start the

---

[11] For example, Jane Menken, James Trussell, Debra Stempel, and Ozer Babakol, "Proportional Hazards Life Table Models: An Illustrative Analysis of Socio-Demographic Influences on Marital Dissolution in the United States," Demography, Vol. 18, No. 2, May 1981; and Menchik.

clock." When applied to the life-history data, such a restriction might effectively eliminate consideration of many of the variables documented in the life history. Hazard models can be adapted to allow for time-varying covariates by breaking the time periods into subperiods and treating the exogenous variables as fixed within each of those periods.[12] Allowing for time-varying covariates seems especially appropriate for analyses of migration, since events occurring shortly before the migration may be especially important.

Recently, hazard models have been adapted to handle another feature of stochastic processes--heterogeneity.[13] Heterogeneity occurs when individuals vary in their risks for reasons not included in the model. For example, independent of socioeconomic characteristics, some individuals may be more prone to wanderlust. With such heterogeneity, the migration rate will tend to decrease over time; those most prone to migrate will migrate first, leaving behind an increasingly selected sample of those less and less prone to migrate. Heterogeneity can give the appearance of duration-dependence when none exists. Although migration models are potentially subject to bias because of heterogeneity, the algorithm recently developed by Flinn and Heckman to allow for explicit modelling of heterogeneity depends critically on assumptions about the shape of the distribution of "individual effects." Furthermore, the computer program to implement this algorithm is exceptionally expensive to run.

---

[12] This procedure is employed in Mark D. Menchik, "Intra-Urban Mobility and Family Change," The Rand Corporation (forthcoming).

[13] For example, Flinn and Heckman. This issue was addressed earlier by Ralph B. Ginsberg--e.g., in his "Stochastic Models of Residential and Geographic Mobility for Heterogeneous Populations," Environment and Planning A, Vol. 5, 1973. Ginsberg also discusses duration-dependence and time-varying covariates.

DEFINING THE DEPENDENT VARIABLE

In hazard models, the timing of migration becomes an explicit feature. When logit, probit, OLS, or cross-tabulations are to be used for analyses of determinants of migration, the researcher faces several choices regarding how to define the dependent variable. If each individual moved at most once, the dependent variable could simply be a dummy indicating whether or not he or she ever migrated or whether he or she migrated in a particular time period (i.e., = 1 if migrated, = 0 if did not migrate). If some individuals migrate more than once, there is the question of which migration to choose. Consideration of narrow time periods will reduce the extent of the problem but may not eliminate it altogether.[14] One possibility would be to have the number of migrations in the time period be the dependent variable, but this will cause difficulties for measuring explanatory variables that vary over locations. Another possibility is to arbitrarily choose the multiple migrant's first or last migration in the period. If the last is chosen, the number of other migrations in the period (or ever before) could be included as an explanatory variable. (Section IV discusses repeat migration in more detail.)

[14] One extreme is to have units of observations be person-year observations (this approach was used, for example, in Alden Speare, Jr., Sidney Goldstein, and William H. Frey, Residential Mobility, Migration, and Metropolitan Change, Ballinger, Cambridge, Mass., 1975). However, if different person-year observations on the same individual are pooled, the observations will not be independent (see footnote[4] in this section).

EXPLANATORY VARIABLES

What explanatory variables should be considered in the multivariate analysis of the determinants of migration? These would include information on levels and changes in other relevant contemporaneous variables collected in the life-history, for example, characteristics of pre-migration location (e.g., size of place), employment-related factors, marital status, and fertility and perhaps changes therein. In addition, the analysis should control for age, date, sex, completed education, cultural variables (language, religion, ethnicity), and migration history (e.g., number of previous moves, duration of stay in pre-migration location),[15] all measured as of a time soon before migration.[16]

Analyses of determinants of migration should not control for variables that are only applicable to migrants, such as reasons for migrating or for choosing a particular destination, who was responsible for the decision to migrate, or presence of friends and relatives at destination, since these cannot be defined for nonmigrants.[17] Furthermore, variables pertaining only to the household's situation at or near the time of the interview, e.g., information on current land-

[15] Migration-history is not truly exogenous to the current migration decision process. Hazard and event-history models explicitly recognize this.

[16] If the sample design is stratified (e.g., oversampling geographic areas with a higher concentration of migrants), these strata must be controlled in the analysis. If this is done and the underlying model is correct, maximum likelihood techniques are appropriate even when the data come from a stratified sample design.

[17] Similarly, variables pertaining only to nonmigrants, e.g., reasons for not moving, should not be considered as determinants since they cannot be defined comparably for migrants.

holding or housing characteristics, should not be considered as determinants. To consider these as determinants of migration, it would be necessary to make the unlikely assumption that the current values reflect migrants' situations before moving.

TIME-SUBSCRIPTING THE EXPLANATORY VARIABLES

Once the explanatory variables are chosen, there remains the issue of the time point to which they should refer. There are several possibilities. If migration is being measured over a specific interval, e.g., 1970-74, the explanatory variables for both migrants and nonmigrants can be defined as of the beginning of the interval. That approach is fine for short migration intervals, but becomes problematic for longer intervals because the explanatory variable is measured a variable number of years before the event it is explaining. Hence it will be measured differently for different sample members. The greater the number of years before the move, the likelier the variable has changed since its measurement. For example, if one is explaining 1970-79 migration, a move that took place in 1979 may have had little to do with 1970 levels of explanatory variables.

An alternative approach is to measure the explanatory variables a fixed amount of time before migration. Ideally, that amount of time should be based on information about the migration decisionmaking process. That is, how soon before their actual moves do most migrants decide to move? If the data do not enable one to sort out the ordering of different events that occur in a given year, time intervals averaging less than a year will not be feasible. A reasonable approach, both on conceptual and practical grounds, would be to measure the explanatory variables as of the year immediately preceding the one in which the

migration took place. The explanatory variables could include changes prior to this point also.

With such an approach, the desired time subscript on explanatory variables is clear for migrants. However, since an attempt to understand why the migrants migrated should consider why the nonmigrants chose not to move, to what time period should explanatory variables for nonmigrants refer? This depends in large part on the time period over which migration is being analyzed. If a retrospective survey fielded in 1980 is used to analyze determinants of migrations that mostly took place in the 1960s, it would be inappropriate to measure the explanatory variables for nonmigrants as of the time of the survey. One approach would be to randomly assign time subscripts to nonmigrants based on the distribution of time subscripts for migrants, conditional on their age. The idea is that the conditional distributions of timing of actual and potential moves be similar for migrants and nonmigrants. Otherwise there is the risk that differences in timing of measurement could cause systematic biases. Short of generating a distribution corresponding to that for migrants, or systematically matching migrants with nonmigrants, nonmigrants could be assigned the mean time subscript for broad age groups, or the mean for the overall sample of migrants.

SUBSAMPLES

In addition to controlling for migration determinants by including them as explanatory variables, the analyst may want to stratify the sample by some of these to allow their effects to completely interact with those of the right-hand-side explanatory variables. For example, the samples could be stratified by broad age groups or date groups or both, since the influences on migration decisions may change over time

or vary with age. For example, the determinants and consequences of migration before a particular date (e.g., before independence or prior to the initiation of a particular policy) could be compared with those afterward. This would allow comparisons of the experiences of different migration cohorts. The triangular matrices suggested in the previous section will reveal whether migration rates have changed over time or whether they vary with age. However, even if there is no change or variation, the relative influences of particular explanatory variables may nonetheless vary with age or time. Similarly, the analyst may choose to stratify the sample by sex, ethnicity, broad locational groups (e.g., urban and rural strata), or other sociodemographic variables.

SUMMARY

Migration-history data have great potential for helping us to understand why some individuals migrate but others do not. These data are richer than those typically available to migration analysts and call for methodologies different from those one would apply to, say, census data. A variety of technical procedures are available for extracting the information from migration histories. Perhaps the most promising are hazard models that allow for time-varying covariates. Where these are not feasible, however, several other techniques may be used to take advantage of some of the unique features of life-history data.

IV. ANALYSIS OF CHOICE AMONG TYPES OF MOVES

Another attractive feature of migration history data is their capacity to shed light on different types of moves. Often the policy or research interest is not only in why people migrate, but also some aspect of where--that is, the destination chosen. For example, some migrants from rural areas go to the capital city, but others go elsewhere (e.g., to smaller towns or to other rural areas). The type of destination chosen typically has important implications. Some individuals who have previously migrated return to places where they lived before, while others move on to new places. What affects these choices, and are they subject to policy influence?

These questions can be addressed by dividing the sample into subsamples at risk to a similar set of moves. For example, the analysis of rural outmigration to various possible types of destinations would be based on a sample of rural residents at the beginning of the migration interval. The analysis would model their choices among such alternatives as not migrating, migrating to another rural area, migrating to a small town, or migrating to a metropolitan area. Alternatively, the analysis could be divided into two modelling stages: (1) the decision to migrate, and (2) the choice of destination. For the analyses of return and onward migration, the sample would consist of people who had migrated before, and the analysis would seek to explain the determinants of their choice among the alternatives of staying where they are, returning to a place where they lived before, or moving on to a new place.[1] The complementary subsample of individuals who never

[1] This type of model is presented in Julie DaVanzo and Peter A.

migrated before could be used to analyze the determinants of primary (first-time) migration. Still another possibility would be to model choices among particular geographic areas, e.g., states or broad economic regions.

Both personal characteristics (e.g., age and education) and area characteristics (e.g., differences between origin and destination job opportunities, the distance between origin and destination[2]) will affect migrants' choices among alternative destinations. As in the analysis of determinants of migration, multivariate analysis should ultimately be used to assess the separate influences of the factors that affect choices among alternative destinations. An appropriate multivariate technique for modeling choices among discrete alternatives is polytomous logit analysis, a nonlinear maximum likelihood technique.[3] Log-linear models can be used if all variables are categorical. Discriminant analysis and a recently developed ordinary-least-squares approximation to polytomous logit[4] yield inferences similar to those of polytomous logit and can be used for preliminary analysis. And, as before, one can begin with simple tabulations, for example, comparing the average characteristics of individuals who make different types of choices. Again it is recommended that statistical

---

Morrison, "Return and Other Sequences of Migration in the U.S.," Demography, February 1981; and in Julie DaVanzo, "Repeat Migration in the U.S.: Who Moves Back and Who Moves On?" Working Paper WP-80-158, International Institute for Applied Systems Analysis, Laxenburg, Austria, November 1980.

[2] Such variables are easier to define when the units of choice are discrete areas, e.g., states, than when they are types of areas, e.g., "other rural areas."

[3] See review by Takeshi Amemiya, "Qualitative Response Models: A Survey," Journal of Economic Literature, Vol. 19, December 1981, pp. 1483-1536.

[4] See Haggstrom.

tests (in this case, F tests) be performed to test whether the average characteristics differ significantly among alternatives.

V. ANALYSIS OF CONSEQUENCES OF MIGRATION

Consequences of migration can be assessed at both individual and aggregate levels.[1] At the individual level, are migrants better (or worse) off because they moved? Are areas' average wage rates lower (or higher) after migration because outmigrants earned more (or less) than those they left behind or because inmigrants earn less (or more) than those they joined? Does migration impose externalities on nonmigrants in origin or destination areas (for example, by raising their cost of housing or reducing the wages they receive)? Answers to these questions are needed to design effective migration policies.

This section focuses on the assessment of individual-level consequences of migration for migrants both vis-a-vis what they would have experienced had they not moved and vis-a-vis the experience of nonmigrants. (Ignored here are possible externalities that might affect the experiences of the nonmigrant control group.)

TYPES OF COMPARISONS

To assess whether migrants are better off because they moved, the appropriate conceptual comparison is with what the migrant would have experienced without moving. Since the hypothetical outcome of not

---

[1] The papers prepared for the ESCAP meeting by Hugo, Simmons, Goldstein and Goldstein, and Standing discuss conceptual and substantive issues in assessing consequences of migration (Graeme Hugo, "Methods for Evaluation of the Impact of Migration on Individuals, Households, and Communities"; Allan B. Simmons, "Methods for Evaluation of the Impact of Migration on Individuals, Households, and Communities"; Sidney Goldstein and Alice Goldstein, "Techniques for Analysis of the Interrelations between Migration and Fertility"; and Guy Standing, "Issues in Analyzing Inter-Relationships Between Migration and Employment"; papers presented at ESCAP Technical Working Group Meeting on Migration and Urbanization, Bangkok, December 1981).

moving is not directly observable, most analyses of consequences of migration rely instead on the experiences of the destination residents whom the migrant joined, or of the origin residents from whom the migrant departed. Such comparisons show whether or not the migrants are better off than nonmigrants at either origin or destination, but they do not necessarily reveal whether the migrants themselves are better off than they would have been had they not moved. For example, an unemployed person who migrates and finds a low-paying job has improved his lot; however, he may earn less than nonmigrants at either origin or destination, in which case his improvement appears dubious.

A better way to assess the individual-level consequences of migration is to compare the migrant's own pre- and post-migration situations. However imperfect an indicator of the migrant's hypothetical subsequent experience had he not moved, his own pre-migration experience is in most cases superior to that of other individuals.[2]

METHODS

With a fixed and relatively short migration interval (e.g., no longer than, say, five years), migration consequences can be assessed at the end of the interval or by comparing characteristics at the beginning and end of the interval. Such an approach simplifies definition of the dependent variable for nonmigrants. However, it becomes decreasingly appropriate as the migration interval becomes longer, since the number

[2] These conceptual issues are discussed by John Antel, Returns to Migration: A Literature Review and Critique, The Rand Corporation, N-1480-NICHD, 1980; and by Julie DaVanzo and James R. Hosek, Does Migration Increase Wage Rates?--An Analysis of Alternative Techniques for Measuring Wage Gains to Migration, The Rand Corporation, N-1554-NICHD, 1981.

of years between the migration and the measurement of its consequences becomes more variable among individuals. For some, the "consequence" would be measured one year following the move, for others 10 or 15 years afterward. Alternatively, the after-migration part of the before-and-after comparison can be measured a specific amount of time, say two years, after the move, while the before-migration part is measured a certain amount of time before the move. Whenever migrants are being compared with nonmigrants, either in terms of their after-migration experiences or before-after differences, the time subscripts for nonmigrants should be comparable to those for migrants (as discussed in Sec. III).

Where possible, comparisons of migrants and nonmigrants should control for socioeconomic characteristics (e.g., age, education) that may affect the dependent variables. Since migration tends to be selective along these dimensions, these variables typically differ between migrants and nonmigrants. Even with these controls, however, the comparisons may still be flawed by the existence of other, unobserved differences between migrants and nonmigrants. After-before differences may net out some of these influences, but others may remain. Those particular individuals who chose to migrate did so because they expected to benefit from migration (vis-a-vis what they would have experienced had they not moved); other individuals chose not to migrate, because they felt they would be better off by staying. Where two otherwise identical individuals make opposite decisions--one migrating, the other staying--something unobservable caused their actions to differ, and this same factor may also affect their actual and expected gains to migration. A recently developed statistical technique--the

regression-switching model--appears appropriate for estimating the extent of this unobserved "selectivity bias,"[3] but so far there have been too few empirical applications to judge the practical value of this approach.[4]

Analyses of effects of migration can assess the consequences of particular types of moves, e.g., rural-to-small-town vs. rural-to-metropolitan. Migrants could be compared with nonmigrants at origin (e.g., with rural nonmigrants) or with those at destination (i.e., with nonmigrants in small towns or metropolitan areas). As noted in Sec. II, a stratified random sample of particular areas at the time of the survey will not necessarily yield a random sample of residents of particular origin areas in the past. This should be kept in mind when choosing the geographic units of analysis for assessments of consequences of migration.

One can also assess the influence of characteristics of the move, such as who was responsible for making the decision to move or how the migrant learned about the destination. For example, do individuals who were the main decisionmakers increase their incomes more than those whose spouses or children were mainly responsible for the decision to move? Such an analysis must be restricted to migrants since these explanatory variables cannot be defined for nonmigrants.[5]

---

[3] James J. Heckman, "Sample Selection Bias as a Specification Error," Econometrica, Vol. 47, No. 1, January 1979.

[4] Robert A. Nakosteen, and Michael Zimmer, "Migration and Income: The Question of Self-Selection," Southern Economic Journal, Vol. 46, No. 3, January 1980; DaVanzo and Hosek; Chris Robinson and Nigel Tomes, "Self-Selection and Interprovincial Migration in Canada," Discussion Paper 82-1, Economics Research Center, NORC, Chicago, 1982.

[5] If comparisons are restricted to migrants and consequences are assessed at the end of a fixed interval, the number of years between the migration and the measurement of the consequence can be included as an explanatory variable.

Migration consequences that can be assessed include changes in fertility, marital status, education, employment status, occupation, and wage rates. Some of these can be viewed as continuous (e.g., wage rates, fertility). Others are qualitative (e.g., change in occupation, employment status, or marital status) and could either be converted into continuous measures (e.g., using a Duncan-type scale for occupation) or treated as discrete polytomous variables in the analysis (e.g., remained unemployed, became employed).

VI. IMPLICATIONS FOR DATA PROCESSING

Each of the various types of analyses recommended herein calls for a particular measurement of migration and its determinants and consequences. Each entails a different type of processing of the life-history data. These include:

- Comparisons of areas of residence at the beginning and end of particular migration intervals.
- Counting the number of migrations in each of these intervals.
- Counting person-years of residence in particular locations.
- Measuring the (potential) determinants and consequences for migrants and nonmigrants as of the beginning and end of a fixed migration interval.
- Defining determinants as of a fixed amount of time before the migration; defining consequences a fixed amount of time after the migration; and using a similar procedure (with randomly selected dates with a distribution similar to that for migrants) for defining potential determinants and control measures of consequences for nonmigrants.
- Computations of number of event changes or durations of events that are documented in the life history (for example, number of years in a location or in a job, number of previous migrations or job changes, number of years married, total number of children before migration).

Retrieval of the data required for these various types of analyses is facilitated by computer software with which to structure a hierarchical dataset so that one can (1) convert the variable-length[1] life-history records into fixed-length analysis records (e.g., one per migration interval, or one per migration); and (2) retrieve values of particular variables (e.g., fertility or employment) at fixed dates or a fixed amount of time before or after a migration.

Several computer programs exist for structuring hierarchical datasets. One is SIR, the Scientific Information Retrieval data-handling package.[2] Another is RETRO, a program developed and used at The Rand Corporation to process life-history data from our Malaysian Family Life Survey and INCAP-Rand Guatemala Survey.[3] These programs have a number of retrieval options, most of them keyed to an event (which may be defined as a migration, job change, birth, or a particular age or date). These retrieval options include:

- Value of a variable at (or some specific amount of time before or after) the occurrence of a particular event (e.g., employment status in 1970 or occupation the year before a move).

---

[1] That is, one entry for each new event in the various areas of life covered.

[2] Barry N. Robinson, Gary D. Anderson, Eli D. Cohen, and Wally F. Gazdek, SIR Scientific Information Retrieval Users Manual, SIR, Inc., Evanston, Ill., 1979.

[3] Iva MacLennan, RETRO: A Computer Program for Processing Life History Data, The Rand Corporation, R-2363-AID/RF, March 1978.

- Respondent's age at the time of the event.
- Date of the event.
- Elapsed time between two events (e.g., between two migrations or between a job change and a migration).
- Number of times in a status between two events (e.g., number of migrations or number of children born between two particular dates or ages).

RETRO, SIR, or other software with equivalent capabilities, greatly simplify retrieval of data from life-history data and construction of analytic records.[4]

---

[4] The pros and cons of using RETRO, SIR, or custom programming for processing life-history data are discussed in Terry Fain, Three Methods for Processing Life-History Data, The Rand Corporation, N-1544-AID, July 1980.

INTENTIONALLY

LEFT BLANK

REFERENCES

Amemiya, Takeshi, "Qualitative Response Models: A Survey," Journal of Economic Literature, Vol. 19, December 1981, pp. 1483-1536.

Antel, John, Returns to Migration: A Literature Review and Critique, The Rand Corporation, N-1480-NICHD, 1980.

Butz, William P., and Julie DaVanzo, The Malaysian Family Life Survey: Summary Report, The Rand Corporation, R-2351-AID, March 1978.

Courgeau, Daniel, "Methods of Linking Migration Statistics Collected from National Surveys with Those from Population Censuses," paper presented at ESCAP Technical Working Group Meeting on Migration and Urbanization, Bangkok, December 1981.

DaVanzo, Julie, and James R. Hosek, Does Migration Increase Wage Rates?--An Analysis of Alternative Techniques for Measuring Wage Gains to Migration, The Rand Corporation, N-1554-NICHD, 1981.

DaVanzo, Julie, and Peter A. Morrison, "Return and Other Sequences of Migration in the U.S.," Demography, February 1981.

DaVanzo, Julie, "Microeconomic Approaches to Studying Migration Decisions," in DeJong and Gardner, pp. 90-129.

DaVanzo, Julie, "Repeat Migration in the U.S.: Who Moves Back and Who Moves On?" Working Paper WP-80-158, International Institute for Applied Systems Analysis, Laxenburg, Austria, November 1980.

DeJong, Gordon F., and Robert W. Gardner (eds.), Migration Decision Making: Multidisciplinary Approaches to Microlevel Studies in Developed and Developing Countries, Pergamon Press, 1981.

ESCAP, National Migration Surveys, Survey Manual II: The Core Questionnaire, United Nations, New York, 1980.

Fain, Terry, Three Methods for Processing Life-History Data, The Rand Corporation, N-1544-AID, July 1980.

Fienberg, Stephen E., and William M. Mason, "Identification and Estimation of Age-Period-Cohort Models in the Analysis of Discrete Archival Data," in Karl F. Schuessler (ed.), Sociological Methodology, 1979, Jossey-Bass Publishers, San Francisco, Washington, and London, 1978.

Fienberg, Stephen E., The Analysis of Cross-Classified Categorical Data, The MIT Press, Cambridge, Mass., and London, 1977.

Flinn, Christopher J., and James J. Heckman, "New Methods for Analyzing Event History Data," discussion paper, Economics Research Center, National Opinion Research Center, Chicago, 1981.

Ginsberg, Ralph B., "Timing and Duration Effects in Residence Histories and Other Longitudinal Data: I. Stochastic and Statistical Models," Regional Science and Urban Economics, Vol. 9, North Holland Press, 1979.

Ginsberg, Ralph B., "Stochastic Models of Residential and Geographic Mobility for Heterogeneous Populations," Environment and Planning A, Vol. 5, 1973.

Goldstein, Sidney, and Alice Goldstein, "Techniques for Analysis of the Interrelations between Migration and Fertility," paper presented at ESCAP Technical Working Group Meeting on Migration and Urbanization, Bangkok, December 1981.

Guy Standing, "Issues in Analyzing Inter-Relationships Between Migration and Employment," paper presented at ESCAP Technical Working Group Meeting on Migration and Urbanization, Bangkok, December 1981.

Haggstrom, Gus, "Logistic Regression and Discriminant Analysis by Ordinary Least Squares," The Rand Corporation (forthcoming).

Heckman, James J., "Sample Selection Bias as a Specification Error," Econometrica, Vol. 47, No. 1, January 1979.

Hugo, Graeme, "Methods for Evaluation of the Impact of Migration on Individuals, Households, and Communities," paper presented at ESCAP Technical Working Group Meeting on Migration and Urbanization, Bangkok, December 1981

Joseph E. Potter, "Problems in Using Birth History Analyses to Estimate Trends in Fertility," Population Studies, Vol. 31, No. 2, July 1977.

Kalbfleisch, J. D., and R. L. Prentice, The Statistical Analysis of Failure Time Data, John Wiley and Sons, New York, 1980.

Keeley, Michael C., "Migration as Consumption: The Impact of Alternative Negative Income Tax Programs," in J. Simon and J. DaVanzo (eds.), Research in Population Economics, Vol. II, JAI Press, Greenwich, Conn., 1979.

MacLennan, Iva, RETRO: A Computer Program for Processing Life History Data, The Rand Corporation, R-2363-AID/RF, March 1978.

Menchik, Mark D., "Intra-Urban Mobility and Family Change," The Rand Corporation (forthcoming).

Menchik, Mark D., "Residential Mobility and Public Policy," in W.A.V. Clark and E. G. Moore, Urban Affairs Annual Reviews, Vol. 19, Sage Publications, Beverly Hills, Calif., 1980.

Menken, Jane, James Trussell, Debra Stempel, and Ozer Babakol, "Proportional Hazards Life Table Models: An Illustrative Analysis of Socio-Demographic Influences on Marital Dissolution in the United States," Demography, Vol. 18, No. 2, May 1981.

Nakosteen, Robert A., and Michael Zimmer, "Migration and Income: The Question of Self-Selection," Southern Economic Journal, Vol. 46, No. 3, January 1980.

Radloff, Scott, "Measuring Migration: A Sensitivity Analysis of Traditional Measurement Approaches Based on the Malaysian Family Life Survey," Ph.D. dissertation, Brown University, Providence, R.I., 1982.

Robin J. Pryor (ed.), Residence History Analysis, Studies in Migration and Urbanisation No. 3, Department of Demography, Research School of Social Science, Australian National University, Canberra, 1979.

Robinson, Barry N., Gary D. Anderson, Eli D. Cohen, and Wally F. Gazdek, SIR Scientific Information Retrieval Users Manual, SIR, Inc., Evanston, Ill., 1979.

Robinson, Chris, and Nigel Tomes, "Self-Selection and Interprovincial Migration in Canada," Discussion Paper 82-1, Economics Research Center, National Opinion Research Center, Chicago, 1982.

Rogers, Andrei, "The Migration Component in Subnational Population Projections," paper presented at ESCAP Technical Working Group Meeting on Migration and Urbanization, Bangkok, December 1981.

Simmons, Allan B., "Methods for Evaluation of the Impact of Migration on Individuals, Households, and Communities," paper presented at ESCAP Technical Working Group Meeting on Migration and Urbanization, Bangkok, December 1981.

Speare, Alden, Jr., Sidney Goldstein, and William H. Frey, Residential Mobility, Migration, and Metropolitan Change, Ballinger, Cambridge, Mass., 1975.

Tuma, Nancy B., Michael T. Hannan, and Lyle P. Groenvald, "Dynamic Analyses of Event Histories," American Journal of Sociology, Vol. 84, No. 4, 1979.

Willekens, Frans, "Identification and Measurement of Spatial Population Movements," paper presented at ESCAP Technical Working Group Meeting on Migration and Urbanization, Bangkok, December 1981.