Modeling Migration

Michael J. Greenwood
University of Colorado, Boulder, Colorado, USA

Glossary

determinants of migration The personal and regional factors that influence migration decisions.

disequilibrium perspective The idea that wage differentials provide opportunities for utility gains.

equilibrium perspective The idea that wage differentials are compensated and therefore do not offer opportunities for utility gains.

gravity model A place-to-place migration model that assumes that interregional migration is directly related to the population of the origin and of the destination regions and inversely related to the distance between them.

human capital model A model that treats migration as an investment with future expected returns and costs and that discounts these to arrive at a best present discounted value of a move.

migration The change of an individual's usual place of residence from one migration defining area to another during a migration interval.

migration interval The period of time over which a migratory move may occur.

migration propensity The probability of making a migratory move.

model The abstract depictions and simplifications of a complex real-world process, which may or may not be expressed mathematically.

modified gravity model A place-to-place migration model that adds to the basic gravity model additional variables relating to origin and destination characteristics.

Modeling migration refers to ascertaining the factors that underlie decisions to migrate (or not migrate). Such models also address the personal as well as the regional consequences of migration. The models are abstract depictions and simplifications of complex real-world processes that may or may not be expressed mathematically. Empirical models attempt to ascertain the relative importance of various determinants of migration and, in the case of models of the consequences of migration, to ascertain migration's role, along with other factors, in various economic and social outcomes at the personal and the regional level (such as personal earnings change and regional employment, unemployment, or wage change).

Introduction

Migration is inherently a spatial phenomenon. It entails a change in an individual's usual place of residence from one migration-defining area to another over a given period of time (called the migration interval), which may be a month, a year, a 5-year period, a life time, or some other interval. Because it is a spatial phenomenon, modeling migration must refer to at least two regions, even if the second region is "somewhere else." Migration phenomena may be modeled theoretically or empirically. The best empirical models have strong theoretical underpinnings. Migration models attempt to explain either the causes of personal, family, or aggregate interregional migration or the consequences of migration. Thus, the distinguishing feature of a migration model is that it is an attempt to explain, which is to say that it constitutes an effort to answer the question: Why does a certain migration phenomenon occur? Like any good model, migration models may also allow the investigation to predict, but migration models have not frequently been used for predictive purposes.

All migration models begin with the basic assumption (most frequently made implicitly) that migration is a voluntary human act. However, much migration is clearly involuntary. Examples of involuntary migration are the movement of prisoners and the relocation of military personnel. The migration of refugees of war, political strife, and religious differences is also regarded as involuntary in the sense that such individuals move to escape persecution and/or death and therefore seek at least initially to get away; the personal consequences of migration for refugees have been studied. Models of voluntary migration seek to explain the presumably utility-maximizing behavior of free individuals.

Migration models address several broad questions: (1) Who migrates? (2) Why do these people migrate? (3) Where do the migrants come from and where do they go? (4) When do they migrate? and (5) What are the consequences of migration for the migrants themselves as well as for others in the origin and destination areas? Theoretical migration models tend to address the "Why?" and the "Where?" questions, and to a lesser extent the "What are the consequences?" question. Very little theoretical work has ever been done on the "Who?" and the "When?" questions, although numerous descriptive accounts of who migrates go back to the nineteenth century. The question "Who migrates?" refers to various characteristics of the migrants, such as age, education, race, income, marital status, and presence of children in the household.

The relationships between age and education, on the one hand, and migration, on the other, are among the most commonly studied migration relationships. Model migration schedules have been developed to shed light on the exact relationship between age and the propensity to migrate for various population subgroups. These schedules indicate that in the United States annual intercounty migration propensities peak at 23–25 years of age (at approximately 7.5%) and decline steadily until approximately 62 years of age (to approximately 1.5%). These propensities differ greatly for various population subgroups. For example, depending on age, migration propensities are between two and four times higher for those with 5 years or more of college compared to those with 8 years or less of elementary school.

Because migration declines sharply with increased distance, propensities to move over longer distances are considerably lower than for shorter distances. Thus, interstate migration propensities, as indicated in Table I, are somewhat lower than the intercounty propensities. The tendency for annual interstate migration propensities to rise with education and fall with age is clearly evident in Table I. For example, 6.18% of those individuals ages 25–29 with only a bachelor's degree made a U.S. interstate move between 1993 and 1994, but 3.67% of those in the same age class with a high school diploma only made a similar move during the same period. And whereas 6.18% of the 25–29 age group with a bachelor's degree made an interstate move, only 2.58% of those in the same education class but ages 45–64 did so.

Theoretical Perspective

Economic models of migration, as well as models that derive from other disciplines, take one of two possible broad theoretical perspectives: (1) the disequilibrium perspective or (2) the equilibrium perspective. The theoretical perspective taken in almost all migration research conducted by economists and others prior to the late 1970s was that of a disequilibrium system. The perspective is called disequilibrium because migration is assumed to be motivated by the existence of a set of nonmarket clearing regional wages. The idea is that spatial differences in wages, earnings, or income reflect opportunities for utility gains that can be realized through migration. In more recent years, this disequilibrium perspective has been challenged by proponents of the equilibrium hypothesis, which assumes that spatial differences in wages are compensating and therefore do not reflect opportunities for utility gains. Thus, a basic assumption underlying models based on one or the other of these perspectives is considerably different.

The disequilibrium approach does not rely on location-specific amenities, and regional wages and regional rents adjust slowly to exogenous disturbances. In the equilibrium approach, migration is conditional on amenities. Moreover, this approach does not rely on long adjustments of wages and rents to disturbances, especially in the United States where institutional and other impediments to factor mobility are relatively low. Systematic long-term forces, such as rising real income in some or all locations, underlie in an important way consumption

Table I U.S. Interstate Migration Propensities by Age and Education Class, 1993–1994[a]

	Age					
Education	*Under 25*	*25–29*	*30–34*	*35–44*	*45–64*	*65 & over*
All 25 & over	2.38	4.72	3.26	2.50	1.68	1.22
Less than ninth grade	1.33	3.69	2.01	2.08	1.03	1.00
High school graduate	1.98	3.67	2.60	2.12	1.53	1.06
Bachelor's degree	3.41	6.18	4.45	2.91	2.58	1.36
Graduate/profession degree	3.14	8.96	5.10	2.79	2.07	2.61

[a] *Source:* Calculated from data provided in Hansen (1995), Table 4.

amenity demand growth and provide the rationale for migration. Thus, both the disequilibrium and equilibrium approaches assume that spatial variations in utility underlie the motive to migrate, but the differences spring from the source and persistence of the variations. The perspective taken by the analyst not only shapes the precise form of the model that is specified and estimated, it also contributes to the interpretation placed on the estimated coefficients of wage and other variables.

Gravity and Modified Gravity Models

Although they are not often credited with the development of the gravity law of spatial interaction, in 1938 economists Makower, Marschak, and Robinson clearly laid out the basic concept: "Quite a close relationship was found between discrepancies in unemployment rates and migration of labour where allowance was made for the size of the insured population and the distance over which migrants had to travel." The key to the gravity model is the relationship between migration and distance, as well as between migration and origin and destination population sizes.

During the 1940s, Princeton astronomer Stewart noted that the distance to his students' home towns seemed to behave like the Newtonian law of gravitation. Thus, Stewart expressed the gravity law of spatial interaction as $F = GP_iP_j/D_{ij}^2$, where F = Gravitational or demographic force, G = Constant, P_i = Population of origin i, P_j = Population of destination j, and D_{ij} = Distance between i and j. This relationship states that demographic force is directly related to the origin and destination population sizes and inversely related to the square of the distance between them. If the square on the distance term is replaced by α and the relationship is placed in the migration context by substituting migration from i to j, M_{ij}, for F, we get $M_{ij} = GP_iP_j/D_{ij}^{\alpha}j$. If this model is expressed in double-log form, it suggests that the population parameters are equal to 1.0, meaning that a 1% increase in origin or in destination population results in a 1% increase in migration from i to j. This assumption is clearly restrictive, and the population elasticities are subject to empirical tests. Thus, the gravity model can be written as:

$$M_{ij} = \frac{GP_i^{\beta 1}P_j^{\beta 2}}{D_{ij}^{\alpha}}.$$

In this form, the values of β_1 and β_2 may be freely estimated, and the hypothesis that they are equal to 1.0 may be tested. This basic form of the gravity model was tested rarely because little additional effort was required to specify and estimate the more appealing modified gravity model. Moreover, when β_1 and β_2 are estimated freely, their values are rarely equal to 1.0.

During the 1960s, the main thrust of migration research took on a decidedly more formal tone that has continued to the present. Most of the research was not formal in a theoretical sense but rather intuitively generated hypotheses were at first tested formally in an econometric sense with aggregate data, typically (but not always) with place-to-place migration data. These aggregate models of migration frequently were specified in the context of modified gravity models. The models are of the gravity type, in that migration is hypothesized to be directly related to the size of relevant origin and destination populations and to be inversely related to distance. The models are modified in the sense that the variables of the basic gravity model are given behavioral content and additional variables that are expected to influence the decision to migrate are included in the estimated relationship.

The modified gravity models that became common in the migration literature beginning in the 1960s add several additional variables to those of the basic gravity model. Thus, we commonly find studies of place-to-place migration that take the following form:

$$\begin{aligned} \ln M_{ij} &= \ln \beta_0 + \beta_1 \ln D_{ij} + \beta_2 \ln P_i + \beta_3 \ln P_j \\ &\quad + \beta_4 \ln Y_i + \beta_5 \ln Y_j + \sum_{n=1}^{m} \beta_{in} \ln X_{in} \\ &\quad + \sum_{n=1}^{m} \beta_{jn} \ln x_{jn} + e_{ij} \end{aligned} \tag{1}$$

where the Y terms refer to income. Other variables that are commonly included (as reflected in terms containing X) are unemployment rates, degree of urbanization, various climatological amenity variables, various measures of public expenditures and/or taxes, and many other factors. For certain variables, some models contain only origin characteristics, such as median age or median number of years of schooling, which are meant as proxies for the characteristics of the population from which the migrants are drawn. Modified gravity models hold an important place in the migration literature because their formulators tried to incorporate behavioral content into the context of the gravity model approach. These efforts subsequently led to formal models of the migration decision process, such as those reflected in many studies that incorporate microdata. Moreover, such models included a mix of disequilibrium and equilibrium notions that anticipated the later, more rigorous development of the equilibrium hypothesis of migration.

The connection between modified gravity models and the migration decision process has not always been tight. The dependent variable in modified gravity models is meant to proxy the probability of moving from i to j. However, the denominator of the dependent variable frequently has been the population measured at the beginning or end of the migration interval. Such a measure falls short of reflecting the population at risk to make a move from i to j. For example, the beginning-of-period population includes people who die during the period over which migration is measured, as well as those who emigrate from the country and who are thus not available to be counted as migrants. The end-of-period measure includes in-migrants who were not at risk of being out-migrants from the area and also introduces simultaneity between migration and the population measure.

Modified gravity models are frequently estimated in double-logarithmic form, presumably because this functional form yields reasonably good fits and the coefficients obtained from it can be directly interpreted as elasticities of migration's response to changes in the various independent variables of the estimated models. However, common use of the double-logarithmic functional form to estimate modified gravity models has led to arguments for the adoption of nonlinear maximum likelihood logit methods over the double-log form of the model. In part, the argument hinges on the geographic size of the regions for which migration is measured. If all regions had the same population and land area, the migration and nonmigration probabilities would reflect the costs and benefits of the various locational choices. However, the regions of any country differ greatly in population and land area. A larger share of all moves will tend to occur within the boundaries of larger regions. Consequently, more nonmigration will appear to exist for such regions. The result is that nonmigration is spuriously correlated with origin population size and land area.

In the polychotomous logistic model, the migration probabilities are expressed as ratios and the probability of not migrating is used in the denominator of the expression to normalize the flows. That is, the dependent variable is $\ln[m_{ij}/(1 - m_{ij})]$, which is sometimes called the logarithm of the odds ratio. Here m_{ij} refers to the probability of migration from i to j and is thus measured as M_{ij}, or the actual number of movers from i to j divided by the population at risk to migrate from i. The model can be estimated in one of two ways that make sense in the migration context. First, again assuming the double-log form of the model, the log of the ratio of various destination-to-origin characteristics can be used. This approach, referred to as uniform symmetric, implies that coefficients on variables for corresponding origin and destination characteristics are the same except for sign. Second, a two-step decision process can be assumed in which the decision maker first decides to migrate and then decides where to migrate based on destination characteristics and perhaps other variables (such as distance) that link the areas. In this case, origin and destination characteristics are introduced separately. In an analogous fashion, nested decision trees could be constructed for other levels of the decision process (e.g., whether to move, where to move, and what house to select). Some dissatisfaction has been expressed with the notion that individuals can decide whether to move independently of deciding where to move. Thus, whether to move and where to move are seen as joint decisions and not discrete and independent decisions.

The standard gravity approach is seen as inefficient because it fails to incorporate information on the relative frequency of nonmigration $(1 - m_{ij})$. However, as the migration interval grows shorter and shorter, the two specifications approach one another because the population at risk to migrate becomes a better measure of the nonmigrating population when the migration interval is very short. In any case, the logit approach provides a more natural transition from the gravity model to the more behaviorally grounded modified gravity model.

Prior to 1975, virtually all migration research was based on aggregate data. In addition to the problem already noted, modified gravity models were characterized by other problems and shortcomings frequently associated with the use of such data. Aggregate data often were used to proxy the characteristics of the population at risk to move, resulting in empirical estimates that did not reflect accurately the influence of personal characteristics on the decision to migrate. With some notable exceptions, studies of aggregate migration failed to account for different types of moves, such as primary (or new), return, and other repeat migration. Aggregate data also concealed differences in the underlying determinants of migration of various population subgroups, although stratification by age and race was not uncommon. Such data failed to account for the institutional population, of which the military was especially important, and they made the study of family migration decisions difficult. Another problem with modified gravity models was that the variables used to explain migration often were measured at the end of the migration interval and were thus subject to simultaneity bias. During the 1970s, several simultaneous equations models were developed to explain the causes and consequences of migration within the same empirical framework, but for the most part during the last several years these models have not been developed further.

The Human Capital Model

Underlying the disequilibrium perspective, at least implicitly, is the simple income-leisure model of labor economics wherein an optimizing agent maximizes a utility function with two arguments, income and leisure, subject to a full-income constraint. One implication of this model is that the individual will supply labor such that the marginal rate of substitution of consumption for leisure equals the wage rate, which in turn implies that the individual labor supply is a function of the wage rate. If we abstract from mobility costs and accept many other assumptions that underlie this simple yet powerful model, the individual is expected to offer his labor services in the market with the highest wage, which may require migration.

The human capital approach was added to the disequilibrium perspective that was then in vogue. This model provided a paradigm that caught the attention of economists and provided a convenient theoretical framework for their research.

Present Value Comparisons

The potential migrant will select the locality at which the real value of the expected net benefit that accrues to him or her from migration is greatest. The income that the individual expects to earn at each alternative destination enters crucially into his judgment concerning the benefits associated with each location. The relevant income measure for the individual to consider is the present discounted value of his or her expected future stream of net pecuniary returns.

Let the present value of the earnings stream in locality j less that in i be:

$$\sum_{t=1}^{n} \frac{(E_{jt} - E_{it})}{(1+r)^t} \tag{2}$$

where r is the internal rate of discount, which although written as a constant does not have to be constant. Let the present value of net costs associated with residence in this pair of localities be:

$$\sum_{t=1}^{n} \frac{(C_{jt} - C_{it})}{(1+r)^t}. \tag{3}$$

The summation is over the individual's remaining life. Then, the present value of investment in migration from i to j, PV_{ij}, is:

$$PV_{ij} = \sum_{t=1}^{n} \left[\frac{1}{(1+r)^t}\right] [(E_{jt} - C_{jt}) - (E_{it} - C_{it})]. \tag{4}$$

An individual residing in i will presumably select that destination for which PV_{ij} is maximized.

The disequilibrium perspective is clearly evident in this model of migration. In the human capital model, economic opportunity differentials represent potential for household utility gains that can be arbitraged by migration. For all intents and purposes, the human capital model was unrivaled for almost 20 years. Indeed, disequilibrium forces were presumed to be the primary drivers of migration long before the human capital model provided an explanation for migration. Many of the early authors recognized that both disequilibrium and equilibrium forces were at work, but they clearly emphasized the disequilibrium factors.

The human capital model provided an appealing rationale for the presence of income variables in modified gravity models, as well as in other models of migration. Based on the disequilibrium perspective, in modified gravity models the origin wage or income variable is expected to take a negative sign, whereas the destination wage or income variable is expected to take a positive sign because migrants move out of low-income areas and into high-income areas. A number of studies have tested this assertion regarding the importance of wages in explaining migration by examining the factors affecting interregional migration in the United States and in many other countries. Based on aggregate data, the empirical findings associated with income, earnings, and wage variables in modified gravity models have not been uniformly strong, although it is probably fair to conclude that the weight of available evidence favors disequilibrium results, particularly for rural-to-urban migration. Of course, the exact results are sensitive to many factors, such as the precise specification of the model, the country and period studied, the population subgroup under investigation, the type of functional form assumed, and the estimation technique employed. Moreover, income or wage measures have almost never been refined to reflect real consumption wages. This is to say that the role of taxes and public services has not received the attention it deserves.

Spatial Job-Search Models

Another type of migration model that sprang from the human capital approach places migration in the context of a spatial job search. The human capital model suggests that individuals compare their opportunities in alternative localities with their present situation. However, to learn about opportunities in other areas, they must conduct a search that yields a distribution of wage offers in each alternative. Presumably, these individuals have a reservation wage (the lowest wage they would accept in a potentially new region of residence). If the person is unemployed, this reservation wage may fall over time. However, he or she would reject any opportunities that yielded a wage below the reservation wage. The duration of the search extends until the individual receives a wage offer that is at least equal to the reservation wage. The reservation wage itself is dependent on the person's age, education, marital status, family situation, and numerous other personal characteristics, as well as on regional characteristics, such as the unemployment rate.

More on the Equilibrium Perspective

Due in part to the fairly consistent tendency for empirical studies based on aggregate data to fail to confirm the importance of wages or income in migration decisions, the equilibrium approach has been offered as an alternative to the traditional disequilibrium perspective. The equilibrium theorists begin by assuming that households and firms are in proximate equilibrium at any point in time. This assumption means that the marginal household and firm, while maximizing utility and profit, respectively, are spatially arrayed so as to receive zero consumer and producer surplus from their location. Any movement from the general equilibrium configuration cannot improve utility or profit.

According to the equilibrium approach, changes in life-cycle factors or generally rising real incomes continuously change the demand for consumer amenities. Real incomes may rise due, for example, to persistent technical progress. Because amenities are not evenly distributed spatially, migration occurs and quickly re-equilibrates households. Net in-migration to amenity-rich areas tends to drive down wages and drive up the prices of locally produced goods, services, and land, *ceteris paribus*. In amenity-poor areas, opposite patterns of change occur. Wages and local prices diverge across regions until they just compensate households for the differing amenity bundles that the various regions supply.

The equilibrium approach has another important facet. A number of studies use the level of regional wages or rents to measure regional environmental quality, including the quality of the climate. The assumption underlying these studies is that equilibrium prevails so that wage and rent differentials are compensating differentials and thus serve as accurate proxies for differentials in environmental quality. For equilibrium to prevail, regional markets must be efficient so that regional wages and prices quickly realign to clear such markets subsequent to any disequilibrating exogenous disturbances. The equilibrium proponents believe that, at any point in time, it is highly likely that regional wages and prices have adjusted to their equilibrium values.

In the equilibrium approach, regional differentials in wages and prices do not generally reflect utility differences that can be arbitraged through household migration. Only those noncompensating regional differentials that remain after controlling for amenity differentials across regions should represent utility differentials that would induce migration. As previously noted, the implication of this view for migration analysis is that a properly specified migration equation should include both regional amenity and regional wage and rent variables. For this reason, proponents of the equilibrium hypothesis typically include a wide variety of regional amenities in their empirical models. Among the variables frequently included are climatological amenities (e.g., average temperature at some time during the year, average humidity, and degree days) and topological amenities (e.g., the presence or absence of a sea coast, variety of terrain, and national forest lands).

From an economist's perspective, an equilibrium process makes great sense. Without the operation of such forces, economists would be hard pressed to develop a reasonable theory to explain interregional movements and the adjustments that result therefrom. However, for the most part, until recently tests of the equilibrium hypothesis in the context of the migration literature have not been fully convincing.

Modeling Migration with Microdata

As already noted, prior to 1975 virtually all empirical migration models employed aggregate cross-sectional data. However, during the 1970s a number of important advances occurred in the modeling of migration that were contingent on the development of new data sets. Two types of data were critical in the sense that the advances simply could not have occurred without them. The first and certainly the most important involved the development of microdata and micro-based panel data that included a migration component; the second involved the development of aggregate time series migration data.

Aggregate data have a number of limitations for studying migration behavior.

1. Aggregate data, such as census data, have an inherent fixed migration interval, so that, for those who move multiple times, only the first and last location are available for study. Thus, an individual may move from i to j to k, but a move from i to k is recorded for the individual based on original and final places of residence. Moreover, an individual who moves from i to j and back to i during the migration interval appears as a nonmover. Although this objection also applies to certain microdata sets, such as the U.S. Census Public Use Microdata Samples (PUMS), microdata are in many instances better in avoiding this type of problem.
2. Aggregate data conceal the differences in the underlying determinants of migration for various population subgroups that are lumped together. Thus, the motives for migration may differ for the young versus the old, high-income versus low-income people, blacks versus whites, and so on. Some stratification is possible with aggregate data, but it is

generally limited. Thus, in general only variations between groups can be explained with aggregate data, not variations within groups.

3. Aggregate migration studies, with some notable exceptions, fail to account for the different types of moves, namely, new, return, and other repeat migration and long- and short-distance migration. However, in more recent years, with the availability of microdata, these types of distinctions have been more common, and several studies have uncovered significant differences in the magnitudes of the influence of various migration determinants on these different types of moves.
4. Aggregate data frequently fail to account for the movements of institutional populations, for examples, movements of military personnel and college students. Because the motives for such movements may have little or nothing to do with the migration determinants typically studied, these moves obscure the behavioral relationships of interest. Of course, it is often possible to remove institutional populations from the data being studied.
5. Aggregate data, given the heterogeneity that exists within the large regions between which migration is measured, may not represent the average characteristics of many of the moves that actually occur. For example, U.S. migration is frequently measured between states and even between the nine census divisions and between the four census regions. Moreover, many moves over substantial distances are omitted because they occur within the region. To some extent, this type of problem can be avoided by using finely defined regions as the spatial unit, such as counties or metropolitan areas.
6. Aggregate data allow the study of the regionwide consequences of migration, but in no way do they allow the study of the personal consequences of migration. Thus, any assessment of the personal consequences of migration was not possible until the development of microdata sets that included a migration component.

The development and use of several microdata and panel data sets during the 1970s and 1980s led to many important breakthroughs in the social sciences, including the study of migration in economics and sociology. Several data sets particularly stand out. The 1967 Survey of Economic Opportunity was one of the earliest microdata sets used to study migration. The PUMS from various censuses have been invaluable in allowing investigators to introduce personal characteristics into their models. Microdata from the Current Population Surveys have also been used to good advantage. Two panel data sets of particular note are the Panel Study of Income Dynamics (PSID) and the National Longitudinal Surveys (NLS). Canadian microdata, particularly census data, have also been valuable. These are not the only microdata and panel data sets that have allowed major advances in migration modeling, but they are probably the most important.

In general, three types of migration models employ microdata and panel data: (1) those that use a dichotomous or polytomous dependent variable reflecting migration status (e.g., moved, did not move, or moved, changed job; or moved, did not change job; did not move, changed job) or destination choice as a function of various personal, household, and place characteristics; (2) those that use a measure of earnings or wages as a dependent variable and distinguish migrant vs. nonmigrant status and perhaps other characteristics of a move, such as distance—this latter type of study is couched in the human capital framework because it involves the estimation of earnings (or wage) functions; and (3) a small number of studies that use employment status or hours worked as a function of migration status, among other variables. Note that the last two types of studies relate to the personal consequences of migration.

Models of migration based on microdata have had a major impact on four general topics that have received considerable attention from students of migration.

1. They have allowed the development of important insights regarding the relationship between employment status and migration. Models based on aggregate data have frequently found unanticipated signs and/or insignificant coefficients on unemployment rate variables, but the availability of microdata has gone far toward clearing up the earlier puzzles regarding this relationship.
2. They have allowed more proper and more precise estimations of the monetary returns to migration. Studies in this general area have been based on the human capital model and have involved the estimation of earnings functions.
3. They have allowed the study of the relationship between various personal and life-cycle characteristics, on the one hand, and migration, on the other; these relationships are almost impossible to understand in any fundamental way in the absence of microdata and panel data. Although general tendencies are often discernible from aggregate data (e.g., migration propensities decline with age and increase with education), in the absence of microdata investigators are unable to control for many other factors when studying the relationship between any given factor and migration. An associated point is that family migration decisions are difficult to understand in a substantive way without the use of microdata.
4. They have encouraged researchers to examine primary, return, and nonreturn repeat migration more

closely than was possible using aggregate data. Based on microdata, a number of important differences have been uncovered in the responsiveness of different migrant types to the various determinants of migration.

Sample Selection Problems

With the development and use of microdata, a number of new econometric opportunities have become available, but at the same time certain econometric problems have been introduced. One of the most important of these problems in the migration context concerns sample selection.

Sample selection problems arise from situations in which a population subgroup is not representative of the entire population whose behavior is under study. Stated more formally, in the Manski's 1989 formulation, the problem is one "of estimating a regression $E(y \mid x)$ when realizations of (y, x) are sampled randomly but y is observed selectively." Some unobservable variable may distinguish population subgroups. The natural temptation is to analyze only the subgroup for which data are available, but this procedure may result in parameter estimates tainted by selectivity bias.

Sample selection problems in migration studies have many sources; four of the most common are (1) sampling design or population coverage, (2) panel attrition, (3) time-dependent disturbances, and (4) differential behavioral responses. The last of these is the typical cause of sample selection bias.

Sampling Design or Population Coverage

The first possible problem originates in the data used to study migration. Due to sampling design or, if the data are from an administrative source, population coverage, the data may not be representative of the entire population. For example, migration data derived from U.S. Internal Revenue Service files are selective of those with sufficiently high incomes that they are required to file an income tax return, and annual migration data for Canada derived from the Family Allowance System are selective of families with children. Many similar examples are available. This type of data shortcoming does not cause the usual sample selectivity problem, which refers to bias in the estimation of certain parameters. Rather, it leads to an inability to generalize from accurately estimated parameters.

Panel Attrition

Over time, some attrition is almost certain to occur within any panel. Families move and are difficult or impossible to trace. Others do not wish to put up with the effort of being interviewed repeatedly. For others, payments to participants in the panel may become insufficient. Even though the lost panel members may be replaced with seemingly otherwise comparable individuals and families, systematic differences may well exist between those who remain in the panel and those who drop out. This causes bias when the attrition is correlated with the dependent variables in migration studies. Some unobservable differences may distinguish the groups, such as attitudes in general or attitudes toward risk. An advantage of panel data is that fixed-effect estimates may remove this source of bias. Any investigator who uses panel data sets such as the PSID and the NLS should study what is known about panel attrition in the data set and understand how attrition problems might affect the particular study under consideration.

Time-Dependent Selectivity Problems

A third potential source of sample selection problems also arises from the data, but specifically from the time period of the sample. Time-dependent selectivity problems occur when migrants from different periods are compared. The idea is that the model applies over a span of years, but the disturbance term is time-dependent (and perhaps a function of some latent, unobservable variable). Although this type of selectivity problem may occur in many contexts, two are particularly relevant to migration: secular problems and cyclical problems.

Secular Problems The education and training received by individuals during one period may differ from the education and training received during a later period. Thus, estimates of the monetary returns to different cohorts of migrants may be tainted. This type of bias may be especially important in the study of the returns to different cohorts of immigrants, who may differ systematically not only in education, training, and other personal characteristics, but also in the self-selective nature of their decision to migrate.

Cyclical Problems Little or no research has directly addressed the issue of changes in migrant quality over the business cycle, but such changes are expected to occur. Migrants tend to be self-selected in the sense that they are typically of greater innate ability and possess greater motivation for personal achievement than otherwise comparable nonmigrants. The self-selective nature of the migration decision should be more pronounced the greater the costs of migration, including the probability of finding a job and the costs of subsequent adjustment in the new occupational environment. During periods of relatively poor economic conditions, as indicated by slow national growth of job opportunities, the costs associated with migration are higher. These higher costs are due to more intensive job-search activities because access to entry-level jobs, as well as jobs providing specific skill training, is more difficult. In contrast, during periods of more rapid national economic expansion, the probability of gaining access to jobs is increased and, consequently, the costs of migration are lower. Because the costs associated with migration are expected to be lower during a period of economic expansion, a lower degree of self-selection occurs in periods of relatively good economic conditions. In other words, when economic conditions are generally favorable, the average quality of the migrant flow is relatively lower. This lower quality may be manifested in labor force participation patterns or work motivation, as well as by general skill level. Virtually no research has addressed these issues.

Differential Behavioral Responses

The fourth potential source of a selectivity problem is also behavioral and is analogous to the classic selectivity bias. For examples, in 1976 Heckman wrote about the relationship between wage levels and female labor force participation. In the migration context, at least three types of self-selection may occur: favorable opportunities, remigration, and productivity.

Favorable Opportunities Migrants may be selective of those individuals with the most favorable opportunities, as suggested by Marshall in 1948. Rational economic agents select their chosen alternative because they have some basis for believing that it will yield a higher return than their other options. Consequently, those individuals who select a given alternative are not randomly drawn from the population as a whole. The fact that individual A migrates, whereas the otherwise comparable B does not, suggests that an important difference exists between the individuals. These differences may be in the way they view costs or in the way they view future benefits, and therefore could be due to differences in discount rates. Individual A, for example, may be more highly motivated to invest in human capital formation, not only in migration, but in other forms as well. If this is the case, the earnings of the remaining cohort from which the migrant is drawn may not provide an accurate estimate of the earnings the migrant would have received in the absence of migration. The resulting selectivity bias, if not properly taken into account, poses potentially serious problems in econometric attempts to estimate the return to migration. As Lewis pointed out in 1974, due to this type of problem the returns to nonmigrants are also biased.

Remigration Among those who migrate, some stay in the new place, whereas others move back to the original place or move on to a third location. If those who move back or move on are the economically least successful migrants, then the remaining migrants will bias upward any estimate of the returns to migration. The selectivity bias problem as associated with the remigration phenomenon was raised by Yezer and Thurston in 1976: "The departure of unsuccessful migrants from a destination leaves a residual of successful lifetime migrants. Calculation of the returns to migrations based on these individuals alone results in an upward bias." Although remigration selectivity is potentially important in assessing the returns to internal migration, it seems especially relevant in estimating the returns to international migration because the presumably less successful immigrants who later leave are lost completely from any data collection system in the original country of immigration.

Productivity Individuals may sort themselves based on their productivity. In 1951, Roy discussed such self-selection in terms of occupations (hunting and fishing), but the same argument can be made for region of residence as well as for occupation. The sorting could be based on the individual's absolute advantage in a region (and occupation) or on his or her comparative advantage, but the basic idea is that he or she will locate in the region and work in the occupation that yields the highest expected relative earnings.

Controlling for Sample Selection Bias

The effects of sample selection bias are similar to those caused by left-out variables. Controlling for these left-out variables yields consistent estimates. Although a number of econometric procedures are available to accomplish this control, a frequently used approach is to estimate a first-stage (structural) probit in order to form an estimate of the missing expectations in the earnings equation. In the migration context, an example of this probit is to estimate a regression to predict migrant status (i.e., migrant vs. nonmigrant). A practical difficulty is identifying the earnings equation.

See Also the Following Articles

Demography • Dynamic Migration Modeling • Mathematical Demography

Further Reading

Borjas, G. J., Bronars, S. G., and Trejo, S. J. (1992). Self-selection and internal migration in the United States. *J. Urban Econ.* **32,** 159–185.

Charney, A. H. (1993). Migration and the public sector: A survey. *Reg. Stud.* **27,** 313–326.

Greenwood, M. J. (1975). Research on internal migration in the United States: A survey. *J. Econ. Lit.* **13,** 397–433.

Greenwood, M. J. (1985). Human migration: Theory, models, and empirical studies. *J. Reg. Sci.* **25,** 521–544.

Greenwood, M. J. (1997). Internal migration in developed countries. In *Handbook of Population and Family Economics* (M. R. Rosenzweig and O. Stark, eds.), pp. 647–720. Elsevier Science, Amsterdam.

Gronau, R. (1974). Wage comparisons—A selectivity bias. *J. Polit. Econ.* **82,** 1119–1143.

Hansen, K. A. (1995). Geographical mobility: March 1993 to March 1994. In *Current Population Reports* (U.S. Bureau of the Census, ed.), pp. 20–485. U.S. Government Printing Office, Washington, DC.

Hausman, J. McFadden, D. (1984). A specification test for the multinominal logit model. *Econometrica* **52,** 1219–1240.

Hunt, G. L. (1993). Equilibrium and disequilibrium in migration modelling. *Reg. Stud.* **27,** 341–349.

Lewis, H. G. (1963). *Unionism and Relative Wages in the United States.* University of Chicago Press, Chicago, IL.

Lewis, H. G. (1974). Comments on selectivity biases in wage comparisons. *J. Polit. Econ.* **82,** 1145–1155.

Maddala, G. S. (1983). *Limited-Dependent and Qualitative Variables in Econometrics.* Cambridge University Press, Cambridge, U.K.

Manski, D. F. (1989). Anatomy of the selection problem. *J. Hum. Resources* **24,** 343–360.

McFadden, D. (1987). Regression based specification tests for the multinominal logit model. *J. Econometrics* **34,** 63–82.

Mincer, J. (1978). Family migration decisions. *J. Polit. Econ.* **86,** 749–773.

Molho, I. (1986). Theories of migration: A review. *Scott. J. Polit. Econ.* **33,** 396–419.

Ravenstein, E. G. (1885). The laws of migration. *J. Stat. Society* **48,** 167–235.

Roy, A. D. (1951). Some thoughts on the distribution of earnings. *Oxford Econ. Pap.* **3,** 135–146.

Sjaastad, L. A. (1962). The costs and returns of human migration. *J. Polit. Econ.* **70**(Suppl.), 80–89.

Stewart, J. Q. (1941). An inverse distance variation for certain social influences. *Science* **93,** 89–90.

Taylor, J. E., and Martin, P. L. (2001). Human capital: Migration and rural population change. In *Handbook of Agricultural Economics* (B. L. Gardner and G. C. Rausser, eds.). Elsevier Science, New York.

Todaro, M. P. (1976). *Internal Migration is Developing Countries.* International Labour Office, Geneva.

Yezer, A. M. J., and Thurston, L. (1976). Migration patterns and income change: Implications for the human capital approach to migration. *South. Econ. J.* **42,** 693–702.