

TIME AND AREAL AGGREGATION PROBLEMS IN ESTIMATING MIGRATION*

İlhan TEKELİ and Leila ERDER

Received June 3, 1977.

* The authors wrote this article as part of a project on "Population Distribution Policies for the Planning of National Spatial Organization in Turkey". The project was supported by an award from the Ford and Rockefeller Foundations in their Program on Population Policy. Other parts of the authors' research are to be published as a book by Hacettepe University, Institute for Population Studies.

Migration is a current topic of scholarly discussion and general interest as more and more of the world's population moves. In spite of the growing number of migration studies, the observation and measurement of migration as a social process in time and space have escaped rigorous review. This paper explores the theoretical problems of observing and measuring internal migration between administratively defined area units and over time periods specified by observation dates. Emphasis is on understanding the way the observed units, whether as demographic or areal components, may make estimated patterns diverge from actual movement when aggregated in time and space.

After reviewing both theoretical and empirical work on migration published over the past decade, time and area aggregation questions appeared to us of central importance. As Lowry has noted, mobility rates for a nation may appear strikingly stable at the highly aggregated level; the introduction of a geographical dimension for sub-national units destroys, however, these apparent regularities.¹ If the researcher knows at the outset the effect dividing a large area into spatial units may have on the estimates, then it may be possible to standardize accordingly. This paper investigates how the selection of geographic boundaries and time periods determines to a large degree the estimate of movement between areas and over time. Since most studies relating migration to any set of explanatory variables have failed to give correlation coefficients at the level of significance obtained in other areas of demography, an improved definition of space and time appear as two promising vantage points for raising the explanatory power of current models.

As a general frame we have taken internal migration as it relates to national population distribution and policy; some questions are explored in more detail because of the nature of the migration phenomenon and data in developing countries.² Our current research in Turkey has set this frame. All aspects are generalized, however, under the assumption that with few exceptions other study areas pose similar dilemmas.

The investigation is in three parts - an overview of migration as a social process and its identification from a policy standpoint; the effect of time on estimates whether as age groupings or as observation points; a dissection of the peculiarities of spatial groupings as they affect migration estimates.

1. I.S. LOWRY, *Migration and Metropolitan Growth: Two Analytical Models*, San Francisco: Rand Corporation, 1966, p.5.

2. The authors encountered these questions at the outset of their study of population distribution and policy in Turkey.

MIGRATION AS A SOCIAL PROCESS

J. J. WOLPERT, Behavioral Aspects of the Decision to Migrate, *Regional Science Association Papers*, Vol. 15, 1965, pp. 159-173; S. L. EVERT, A Theory of Migration, *Demography*, v. 3, 1966, pp. 511-511.

Migration is usually identified as a count of moves during a specified time period out of and into a particular residential area.³ This definition means that migration is partially a function of the time span and area of observation which are selected. Thus, the selection of the time period and units of areal observation is of central concern.

It is conceivable that the statistical problems related to migration could be solved by removing all restrictions related to the expanse of time and space, thus registering all changes of place as migration. Two doubts arise about the adequacy of this approach. First, the necessary information is unlikely to be found. Second, it is doubtful that all changes in place arise from similar motives or have the same impact on spatial organization. Of these two weaknesses the second is more important, for even when detailed moves can be traced these must be grouped to yield a pattern for analysis.

Most groupings in migration analysis try to delimit community in various terms, although these serve only as a crude approximation of community in most instances. Crudeness arises from the fact that movement where no change in permanent residence takes place comprises the bulk of movement in most social systems. To trace only permanent change, a small fraction of the system disturbance, is difficult. Various types of sources and analytical approaches have been engineered to sift out permanent change from the normal disturbance patterns which have no lasting impact on the system.

From a sociological standpoint a migrant is any individual who has changed his community of residence. The community forms the definitional base because departure involves a change in the pattern of social relationships maintained by the individual. For practical purposes this definition must be reduced to the more manageable description of a migrant as any individual who is recorded as crossing some specified boundary when he changes residence, the boundary presumably separating one community from another. In many instances, however, sources or the analytical techniques available do not reveal whether a move actually involved residential change. Thus, the researcher may be forced to assume that estimated changes in population location patterns between two periods stem from residential shift and that non-residential changes form a small portion, reinforcing or disturbing only slightly the estimated location patterns.

4. For a major work in this area see: KUZNETS, ELDRIDGE and THOMAS, *Population Redistribution and Economic Growth in the United States, 1870-1950*. Philadelphia, American Philosophical Society, 1964, 3 vols.

Detecting these changes in spatial organization are the aim of macro-level studies of internal migration;⁴ these examine the way people respond to changing economic opportunities, on the one hand, and the way population distribution is affected by migration on the other.

5. I. TEKELİ, "Mekan Organizasyonlarına Makro Yaklaşım ve Türkiye Üzerinde Bir Deneme" Ankara, 1972, (limited circulation).

Migration from this standpoint is not spatial interaction but spatial process. As a uni-directional change of place, migration is irreversible both in time and space. Like every process it can be understood only in reference to changes in structure.⁵

TIME AND DEMOGRAPHIC PARAMETERS IN MIGRATION ESTIMATES

The population of an area changes continuously over time through the interaction of births, deaths and migration. Coale and Lopez have shown the way the first two components interact with the age distribution to determine the growth and structure of populations where no migration takes place.⁶ The birth rate under these conditions is shown to be a function of the annual rate of bearing a female child at age a and time t ($m(a,t)$) and the proportion of the female population between age a and $a+da$ ($c(a,t)da$).

$$(1.1) \quad b(t) = \int_0^{\omega} c(a,t) m(a,t) da$$

The death rate is determined by the annual death rate of person at age a and time t ($\mu(a,t)$) and the proportion of the population between age a and $a+da$ ($c(a,t)da$).

$$(1.2) \quad d(t) = \int_0^{\omega} c(a,t) \mu(a,t) da$$

Because of these relationships and their application to composing sets of regional model life tables,⁷ we can analyse and predict demographic change in areas with little or no migration even in the absence of complete vital statistics.⁸

In areas open to migration, however, the third component is an important one. It often plays as important a role in areal population growth as natural increase, even under conditions of high rates of natural increase.⁹ Age and sex specific schedules of migration have not been systematically collected and analysed and the present state of our knowledge about the socio-economic context of migration patterns have prevented the proposal of migration-inclusive models for analysing the demographic components of population change in areas open to movement.

Although the temporal distribution of migration is likely to differ considerably from the curves of natural events, it is nevertheless a continuous phenomenon in most cases. From our unit of observation we witness departures, out-migration, and arrivals, in-migration, with the net effect over time on the area of observation being the net change due to the flows (v) of migrants at each age a over time t between our area of observation, j , and all other areas, k .¹⁰

$$(1.3) \quad \begin{aligned} O_j(t) &= \sum_{k=1}^m \int_0^{\omega} c_j(a,t) v_{jk}(a,t) da \\ i_j(t) &= \sum_{k=1}^m \int_0^{\omega} c_j(a,t) v_{kj}(a,t) da \end{aligned}$$

The discrete nature of our demographic sources means, however, that we must guess at the features of continuous change within time intervals by using temporal signposts. The measures we shall examine represent only approximations of the foregoing

6. A.J.COALE, *The Growth and Structure of Human Populations: A Mathematical Investigation*, Princeton: Princeton University Press, 1972.

7. A.J.COALE and P.DENENY, *Regional Model Life Tables and Stable Populations*, Princeton: Princeton University Press 1966.

8. United Nations, *Methods of Estimating Basic Demographic Measures from Incomplete Data*, New York: United Nations, 1967.

9. For estimates of the growth of individual cities with populations of 10,000 or greater in Turkey from natural increase and net change due to migration see: B.TEKÇE, *Migration to Cities in Turkey: Estimates for 1955-60 and 1960-65*, Ankara: State Institute of Statistics, 1967.

10. A.ROGERS and F.WILLEKENS, *Spatial Population Dynamics, Papers of the Regional Science Association*, vol.36, 1976, p.4. Rogers's work, still at its initial stage, promises to give us regional migration modelling tools similar to those developed for the natural increase of closed populations.

relationships and have been dictated by the nature of the information available.

A set of individuals and points to satisfy the foregoing relationships, could be drawn only from a continuous population register where places of residence are recorded on a fine grid. Although an ideal source from a demographic point of view, the areas where such registers are available are limited primarily to Scandinavia. The two other major types of data sources, census counts and sample surveys, are widely available. Thus methods developed to exploit the latter sources are considered below in an examination of the way migration and the estimating techniques for identification involve time.

All counts which are less than continuous occur at specific points in time and usually group the individuals counted according to sex and some definition of age. The mode of the count, the time of year when the count is taken, the interval between each count, and the legal definition of the count affect migration estimates in a number of ways. In order to minimize the counting of individuals in places other than their permanent residence the census may occur on a date when seasonal workers, for example, have returned to their permanent residence.

The second consideration, the interval between the count, has less obvious effects on estimates. Longer intervals do not necessarily ensure the observation of significant system change because lengthening the interval may, in fact, reduce the volume of permanent moves that are estimated. Migrants may die, for example, in the interval and never be identified as changing place of residence or if return migration is widespread these moves may also pass unnoticed if completed in the interval. If one adheres to the strict definition of system change, return moves might be ignored. When retirement or returns forced by unemployment form a large portion of movement in a system, this may become of central importance, however. Finally the legal definition of a census as *de jure* or *de facto* simply distinguishes between individuals as allocated to their legal residence or as counted where found on the day of the census. The bias of the latter is unimportant if most individuals are at their place of permanent residence when counted. A question about permanent residence can also make results comparable to those from a *de jure* count.

Sample surveys are a strong source of additional information. Where censuses are of poor quality, this additional information can strengthen the estimating techniques by contributing parameters which can be used in conjunction with measures derived from the census. The estimating techniques below draw primarily on the census with additional information inserted from sample surveys. In each case emphasis is on the interaction of time and system change.

Most censuses give tables for estimating *lifetime migrants*. In these tables anyone reporting himself in an areal unit other than his unit of birth is classified as a migrant. No time of movement is specified and multiple moves between birth and the most recent census are lost. The statistical picture is further complicated by the fact that one usually does not know the age distribution of those who report themselves as migrants in each census. Moreover, as with other methods, those who die in the interval escape analysis. There is also a reference problem where the names and boundaries of areal units have

changed. Finally, those who report themselves as residing in their place of birth may very well have moved during their lifetime. The age pattern of return moves obviously interacts with the age distribution to make this problem more or less important in the analysis given the volume and age structure of return migrants and the resident population.

When first examining broad change in population distribution these lifetime migrants are useful for constructing a crude index of net gain or loss due to migration. This does not obviate the conceptual weakness of the measure. If the population is closed to international migration, the sums of net gain and net loss necessarily equal zero. The positive or negative sum may then be used with the national population to give a redistribution index.

(2)

$$(+)\text{Net } M_t / P_{Nt} = \text{RI (Redistribution Index)}$$

(+)Net M_t = absolute sum of net migration for each unit time t .

P_{Nt} = national population time t .

11. This takes the form $(\text{Net } M_t / 2) / \text{Mid-period } P_{Nt}$

This can also be revised slightly to give an average measure based on an assumed mid-period population.¹¹ With an index as crude as this a refinement for the population at risk during the interval is probably of little significance. When comparing two different census years the index may be used to give the percentage of the population (P_t) which would have to be redistributed to obtain the preceding (P_{t-n}) distribution. This index can be significantly affected by both the areal units employed and the time which elapsed between $t-n$ and t . Here both time and area explicitly interact to produce the index.

Crude estimates of intercensal net migration may also be obtained from these tables. First, the areal units must be comparable. The lifetime in-migrant (I_t) and out-migrants (O_t) are listed for each (a) and period.

(2.1)

$$\text{Net } M_a = (I_{at} - O_{at}) - (S_I I_{t-n} - S_O O_{t-n})$$

Net M_a = Migration for area A, time interval $t-n$ to t .

S_I = Survival ratio for in-migrants

S_O = Survival ratio for out-migrants

I_a = Lifetime in-migrants to area a.

O_a = Lifetime out-migrants from area a.

This method is unlikely to be used because any set of censuses which would supply the information necessary for calculating intercensal survival ratios by regions would give data that would make possible the estimate of far more robust indices. Moreover, time is so highly aggregated here that movement over half a century or more may affect the indices even when calculated for a shorter intercensal interval.

Another method, the *census survival method*, has the advantage of reducing the aggregation of time to that immediately surrounding the census interval itself if age groups of widths similar to the length of the census interval are used. Here the problem depends on having the population by sex and age groups to give an indirect estimate of change due to migration by projecting the expected population from the observed census count to the next census. Involved in this method are two different time dimensions.

The first time dimension is introduced by the length of the census interval itself. If the time from one census to the next is a decade, in the absence of further information movement must be assumed to occur evenly over the interval. For studying the labor market, for example, this is highly suspect and weakens any analysis where measures of wage levels or employment opportunities are employed as explanatory variables. A lagged model would usually be inappropriate when the interval is so long. If the study focuses only on broad system changes the length of the interval, if it varies from five to ten years as is usually the case, is not so serious.

The second dimension interacts with the length of the census interval and arises from the way the population is grouped by age. Although the time between each census may be five years, if the population is aggregated into five-year age groups or cohorts the actual ages experienced by the cohort for which migration is measured are ten years for all cohorts except those born within the interval.¹² Similarly, if the census period is expanded to cover ten years and five-year cohorts are maintained the width becomes fifteen years for all cohorts except the first, and this is ten. Many countries with decennial censuses would give results with a twenty-year period transited if ten-year cohorts were used. In all instances, the reference period this allows is much longer than that desirable for medium-range planning purposes.

This peculiarity is not immediately apparent in the usual descriptions of the technique, where both the length of the interval a and the time between the census counts, t to $t+a$, appear to make the reference period of length a , falling presumably at the mid-point of the period if migration occurs evenly over the period. In fact, movement has been aggregated into a longer period as shown above and the age groups referred to are much broader than is immediately obvious.

There is one other temporal problem that deserves mention here although fortunately in most instances of less importance. The census survival method requires the respondent to place himself in time - assign himself a correct age. Less error may be involved here than in recalling place of birth or last residence. Nevertheless, in areas where age misreporting is serious this deserves mention. The census survival method reduces the problem of age heaping, particularly preference for digits ending in zero or five, in two ways. First, the age groups themselves, if the interval is five years, have the effect of reducing migration patterns that would appear solely as a result of misreporting by staggering the digits ending in zero and five, although this is complicated by overlapping in both the under and over five-year group in the manner described above. For this reason, a ten-year migration period will tend to make five-year cohorts work out better than a five-year migration period. Secondly, if survival factors are calculated

12. We are indebted to F.C. Shorter of the Population Council for an understanding of this problem first described to us in the manuscript version of his *Computational Methods for Population Projections*, Section 5.19 of Chapter 5 "Urban Population and Migration". This work was subsequently published as: F.C. SHORTER and P. PASTA, *Computational Methods for Development Planning* New York: The Population Council, 1974.

directly from a national census this incorporates national patterns of age misreporting.

(2.2)

$$S(a) = P_{(a)}^{o+t} / P_{(a)}^o$$

$P_{(a)}^o$ = National five-year cohort a, at base year o

$P_{(a)}^{o+t}$ = National five-year cohort a, at time o+t.

where (a) is the five-year cohort.

Variation in age misreporting within a nation certainly exist, and could significantly affect the estimate of internal migration in some cases. Of perhaps more importance in developing nations is the variation in mortality that makes the application of national survival factors unacceptable at a regional level. Survival ratios for each region r may be obtained by scaling the national mortality pattern upwards or downwards if the requisite information is available.

(2.3)

$$S_{ra} = (P_{(a)}^{o+t} / P_{(a)}^o) * [(M_r p(a)) / (M_n p(a))]$$

S_{ra} = Regional survival ratio five-year cohort a

$P_{(a)}^{o+t}$ = National population cohort a at time o+t

$P_{(a)}^o$ = National population cohort a, base year o

$M_r p(a)$ = Survivorship in region r, cohort a

$M_n p(a)$ = Survivorship in the nation, cohort a.

The second term gives the relationship between the survivorship values for each age group in region (r) to those for the same group in the nation.

Thus net change due to migration for any age group a, any region j is found as:

(2.4)

$$m_{j,(a)} = P_{j,(a)}^{t+n} - P_{j,(a)}^{e t+n}$$

$m_{j,(a)}$ = net change due to migration, region j, cohort a

$P_{j,(a)}^{t+n}$ = Population enumerated in region j, time t+n, cohort a

$P_{j,(a)}^{e t+n}$ = Population expected in region j, cohort a, time t+n.

When estimating population change due to migration excluding migration by persons born during the migration period by filling in the first cell for births, the question of the direction of the estimate is of some interest. While applying essentially the same method, the direction of the projection has an effect on the volume of migration that is estimated. This is of concern when comparing results obtained from reverse and forward projections.

If estimates from the forward method are denoted as $M_{r(x)}^a$ FOR and from the reverse as $M_{r(x)}^a$ REV, then the relationship between the forward and reverse methods may be expressed as follows:

(3.1)

$$M_{r(x)}^a \text{ REV} = (M_{r(x)}^a \text{ FOR}) / S_{r(x-n)}^a$$

or conversely

(3.2)

$$M_{r(x)}^a \text{ FOR} = M_{r(x)}^a \text{ REV} * S_{r(x-n)}^a$$

It is clear that the forward estimates will always be the reverse estimates deflated by the census survival rates. Where the rates are high the differences are of less consequence but where they are low they have considerable impact on the results. One way to avoid the bias introduced by the selection of one method over the other is to use both and find the mean of the two estimates of net migration, thus assuming that migration occurs evenly over the period.

All the measures considered so far have only reduced the span into which movement is aggregated and given net gain and loss estimates for subnational units. Flow information, specifying place of origin and place of destination during an interval requires different tables and methods. This is not treated here in any detail because the temporal aggregation problems are the same. A brief review is in order, however, because the areal problems in the following section employ the concept. Flow information is of far more interest for evaluating changes in population distribution as a broad system because it enables comparison between place of origin and destination. This identifies not only areas of loss and gain but also the areas drawn on for each unit's increase. Cross tabulations by age and sex of the population enumerated residing in one unit and born in another can be used to make census change estimates less sensitive to errors in estimating demographic parameters as well as yield flow information.¹³

13. T.K. BURCH, "Internal Migration in Venezuela," Unpublished doctoral dissertation, Princeton University, Princeton, 1962. Burch demonstrates how these tabulations may reduce error.

This classification allows only the outborn population to be treated. The expected population is found by surviving the outborn population from one census and one age group to the next where the number of people born in unit (i), counted in (j) at age (x) for time (t) is $P_{ij,x}^t$.

Place of birth survival ratios ($B_{i,x}$) may be obtained by finding the sum of all the out-born population born in place (i) and the in-born population in a place (i) for a specific age group over a particular period.¹⁴

14. This involves the perhaps questionable assumption that lifetime mortality characteristics are determined by an individual's place of birth.

Thus;

(4)

$$P_{ij,(x+n)}^e = P_{ij,(x)}^t * B_{i,(x)} \quad (\text{for } i \neq j)$$

This gives net change by age and place-of-birth groups of population.

When two consecutive censuses with similar tables cannot be obtained the error reducing advantages are lost. With one census tabulation an important piece of information remains, however, and this alone may prove sufficient for estimating flows. Children born during the interval may prove to be good proxies for the flows of all age groups. Since the first cell of the census tables by definition contains those children who moved during the period, the out-born population in each areal unit of the 0-4 cohort for a five-year period or 0-9 for a ten-year census may be extracted as cross-classified by place of birth to yield a general flow pattern.¹⁵

15. This method was successfully applied to such cross-tabulations from the 1965 Census of Population in Turkey by F.C.Shorter. It rests, naturally, on the demonstration that child migrants are reliable proxies for the flows of all age groups.

16. This question was inserted into Turkey's 1970 Census of Population. Analysis based on the resulting tables may be found in: L.ERDER, "The Making of Industrial Bursa," Unpublished doctoral dissertation, Princeton University, 1976; A.GEDIK, "A Causal Analysis of Internal Migration in Turkey 1965-1970," (doctoral dissertation in progress) University of Washington, 1977; S.YENER, 1965-1970 Döneminde İller Arası Göçler ve Göç Edenlerin Nitelikleri, Ankara: Devlet Planlama Teşkilatı, 1977.

17. A thoroughgoing comparison of the net migration results from this and other techniques needs to be undertaken.

By inserting a question into the census on the respondent's place of residence, or enumeration, at the preceding census it is possible to obtain flow information on the intercensal period for those who were alive at the preceding census and have survived to the next census.¹⁶ This cell may then be crosstabulated with other responses to obtain a detailed view of the socio-economic characteristics and geographic origins of movers. In theory, the time dimension would appear to have been reduced to the census interval by using only two, very distinct reference dates. In fact, the recall problems associated with this method may be so serious even with a short intercensal interval as to offer little improvement over more indirect, less time-specific methods.¹⁷

Throughout this review no mention has been made of areal changes or standardization, leaving this for examination in the third section. One area of interaction with the age structure and timing of movement is of relevance here. Since areal definitions change with time, either as the same units are renamed or as they are divided into new units, some preserving an older name whose areal reference differs, both flow and change estimates become less reliable as the reference period expands. Few censuses control for the way the respondent interprets questions of place and these, perhaps as much as age, deserve the demographer's attention.

AREAL AGGREGATION PROBLEMS IN MIGRATION ANALYSIS

In migration analysis the researcher is confronted by two different types of problems caused by the aggregation or grouping of information on migration into areal units.

The first problem involves the extent to which there may be an underestimate of actual migration caused by areal aggregation. Areal aggregation in itself systematically causes the observed changes in place to fall short of the actual moves. If the researcher is examining a single areal unit or units of equal size he needs to know the percentage of actual moves he has observed. If, however, migration information is compiled for areal units of different sizes, then the question becomes what corrections must be made to achieve comparability. This problem's solution first requires a satisfactory solution to the one for units of equal size.

Although this paper approaches the question from a purely theoretical stance, it is a constant problem in all internal migration estimates where actual observations are involved. Every nation is divided into a set of areal units which are called here observation units. Each nation's individual history and natural geography have made these units differ greatly in size, shape and composition. Migration estimates, with rare exceptions, must be made for these units and thus they implicitly contain the effect of the unit variation on the estimates. By beginning with a solution for units of equal size, we shall proceed to introduce increasingly complicated areal features until we find a solution for units that approximate the very mixed areal characteristics that exist in the real world.

The second type of aggregation problem involves the extent to which the migration we observe through data is heterogenous in its response to different socio-economic and environmental factors. If multiple motives and adaptational processes are involved in migration, then control for areal size may eliminate or conceal the estimate of migration from particular motives.

These two questions take on importance because migration in general declines with distance and migration from different causes responds differently to the friction of distance. If migration increased rather than declined with distance the question of the configuration of observation units would not be a minor one for small areal units. As observed in actual cases, however, migration declines with distance and thus the question is a central one for small units. Furthermore, if migration from different motives behaved similarly with respect to distance then the second question would be unnecessary. Thus a solution to the aggregation problem requires first an examination of the central relationship between migration and distance.

THE VOLUME OF MIGRATION DISTANCE

We shall first identify the process which produces the change in migration with distance and then use a mathematical expression to represent this process. First, let us examine the process of an individual's adaptation to different conditions in order to see why the volume of migration changes with distance.

An individual may change his location within a local labor market. Such movement involves only a change in the location of housing. This type of decision is the simplest one for an individual to take since it requires a minimum of change in his social relationships. This is not, however, an important question for national spatial organization, but may be central to the increase in intra-labor market efficiency.

The distance of this type of change in place of residence is bounded by the size of the local labor market, which in turn is a function of the transportation technology employed in the country under study. While commuting distance by automobile of one hour may be reasonable for developed countries, the city's physical limits are more relevant in developing countries where automobile ownership is negligible and rail links to neighboring settlements are non-existent.

A second type of change in place of residence involves an individual's or family's move to another local labor market. While the individual's occupation remains the same, his social relationship and working environment change entirely. This may occur as the result of a variety of adaptational responses. If a region has specialized in certain economic activities, an individual can change his residence without changing his type of employment. This type of migration represents adaptation which may cause intraregional efficiency to increase. On the other hand, a region may lose its ability to compete within a former area of production with migration operating in the direction of those regions which have gained advantage. In this case an interregional efficiency increase occurs. This type of migration requires more severe adjustment on the part of the individual, and therefore involves fewer individuals.

A third type of migration involves much more radical experience on the part of the individual or family. The mover changes both his social environment and type of job. Just as this type of migration may occur between sectors of the urban employment system, it may also operate between rural and urban areas.

This three-fold division, covering both individual adjustments and system change, is useful for establishing general relationships between the volume of migration and distance. Changes in place of residence of the first group take place within a small area and are the most frequent. Changes of the second group, however, operate at a regional or interregional level and therefore are fewer in number. The third type may be assumed to be even fewer as they involve more radical change. *A priori* one can claim that this last group will cover longer distances because the very radical nature of its causes will reduce the importance of distance relative to other factors.

If all types of migration are observed as is usually done, without separating them according to the type of adaptational response, then the three-fold classification given above would appear sufficient to explain the decline of migration with distance. In fact, however, this proves insufficient if each different type is studied separately. It has been shown empirically that although migration declines with distance regardless of the adaptational response, migration in response to different factors behaves differently with respect to distance.¹⁸ Thus, additional mechanisms are needed to explain the decline with distance for each type of migration mentioned above.

These mechanisms may be summarized in two groups. The first covers the frictional influence of distance. Just as the cost of moving increases with distance, the possibility of obtaining information on distant places diminishes. The second group is comprised of "intervening opportunities" which affect the mover's decision by confronting him with alternatives along the way. Since opportunities near the place of origin are the first to be encountered, there is increasing probability with distance from the origin that the migrant will have selected one of these, removing him from the continuing stream of migrants. Although various studies have examined the influence of distance and "intervening opportunities", little attention has been given to the interrelationship of these two variables.¹⁹ This relationship is important when information on migration is aggregated for areal units.

18. R.L. MORRILL, The Distribution of Migration Distances, *Regional Science Association Papers*, vol. 11, 1963, pp. 75-84.

19. E. MILLER, A Note on the Role of Distance in Migration Costs of Mobility Versus Intervening Opportunities, *Journal of Regional Science*, vol 12, no. 3, 1972, pp. 475-479.

Opportunities that are scattered between origin and destination points must be distributed homogeneously if the frictional explanation is to be of importance in migration analysis. If opportunities are distributed homogeneously in space, then the volume of observed migration will decline evenly with distance and the aggregation question becomes important.

MIGRATION AS A DECLINING FUNCTION OF DISTANCE

The knowledge that migration declines with distance is not sufficient to enable the selection of areal units of appropriate size. The shape of this function must also be known. Let us think of migration as departing from any single point. If there is no directional bias, then the migration function may have the probabilistic distribution of a random walk and tend to exhibit a normal distribution. Various empirical studies, however, have shown that migration distributions are not normal.²⁰

20. G. KULLDORF, *Migration Probabilities*, *Lund Studies in Human Geography*. No. 14, Lund University, 1955.

21. J. AITCHISON and J. A. BROWN, *The Lognormal Distribution*, Cambridge: Cambridge University Press, 1957.

Lognormal distributions, in fact, generally give a better fit.²¹ In this case distance is expressed in logarithmic form. In other words, the marginal friction of distance declines. Lognormal distributions like this are known to arise as a result of "genesis" processes. Thus migration must represent a continuous process which is influenced by preceding and following moves. The use of family or neighbor multipliers as independent variables in migration analysis stems from the individual's behavior in search of opportunities in space and are a part of such a genesis process.

Kulldorf has proposed the following lognormal distribution for migration:

(5.1)

$$P(r) = 2r/v \times e^{-(\log r)^2}$$

In this expression

r = distance

v = scale factor, constant

$P(r)$ = the percentage of total migrants who migrate to distance r .

In another form,

(5.2)

$$\int_{r=0}^{\infty} P(r) dr = 1.$$

In many migration and spatial interaction analyses the following type of exponential functions are used to show decline with distance.

(6)

$$P(r) = ce^{-\alpha r}$$

In this expression

r = distance

α, c = constant scale factors

$P(r)$ = the percentage of migrants to distance r .

22. A.G. WILSON, *Entropy in Urban and Regional Modelling*. London: Pion Ltd., 1970.

Wilson has obtained this type of distance decay function by using entropy maximization subject to the constraint that the total cost of movement remains constant.²²

The third type of expression current in migration analyses takes the classical form of the gravity model where migration decreases with a specified power of distance.

(7)

$$P(r) = ar^{-\alpha}$$

In this expression

r = distance

a, α = constant scale factors

This expression may be used for values of α greater than 1.

23. CAVALLI-SFORZA, *The Distribution of Migration Distances: Models and Applications to Genetics*, *Annales de Monaco*. May, 1962.

Although these are the three major types of mathematical distributions used in migration analysis, others of interest have also been suggested such as the gamma distribution proposed by Cavalli-Sforza to emphasize the search for opportunities involved in migration.²³

SOLUTIONS TO AREAL AGGREGATION QUESTIONS USING SPECIAL DISTRIBUTIONS AND REGULAR SHAPES

In the preceding sections mention has been made only of the size of the areal units used for the statistical observation of migration. Size, however, is not the only factor which affects the amount of migration which is estimated. The amount of migration is affected for each statistical areal unit by its (a) size, (b) shape, (c) and the extent to which its borders are closed to migration; its internal structure also plays a part by the extent to which its own population distribution is (a) homogeneously distributed or concentrated, (b) concentrated at a geometric center or eccentric from this center.

Selected population distributions and regular areal shapes will be used in this section to investigate the influence of these variables on the aggregation problem as formulated in the preceding section.

24. This is particularly characteristic of Turkey and also, the authors presume, of many other areas, especially those in developing countries. The following arguments depend on this general frame.

In general one encounters a nodal center, or city and a rural population homogeneously distributed around this center.²⁴ Therefore, two different population distributions may be employed for the aggregation problems.

- (1) Population concentrated at a point within an area;
- (2) Population homogeneously distributed throughout an area.

If the results obtained for these two peculiar distributions are superimposed on each other, they should approximate actual situations. By taking different shapes, - a circle, square, hexagon, triangle, and rectangle, the results may be generalized for areal units of different shape. Migration information as total and net flows will be treated separately.²⁵

25. The authors have completed such a study to be published as a chapter in their book by the Hacettepe Institute for Population Studies.

Problem One: Migration From a Settlement at the Center of a Circle as a Function of the Radius of the Circle:

The solution to this problem offers an answer to the first aggregation question posed at the outset of this section.

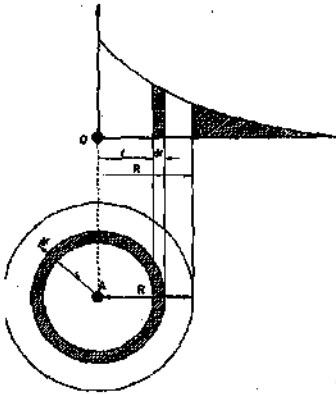


fig. 1

Earlier we examined the distribution of migration from a point A as it declines with distance. This is seen in the upper curve of Fig. 1.

Observed migration remains in the area below the curve from the origin to distance R. In this curve the area described by distance R from the center and thus the space is uni-dimensional. If the same question is considered for two-dimensional space as in the lower part of Fig.1, migration from the center over distance r will spread in the circle within the area $2\pi r.dr$. The way the proportion of total out migration falls outside a circle with radius R may be investigated with different assumptions about the curve of migration with distance.

THE CASE OF GRAVITY FORMULA

The gravity formula for movement over distance is used here. The volume of total migration which can leave the circle of radius R is:

$$\int_{r=R}^{\infty} P(r) dr = \int_{r=R}^{\infty} ar^{-\alpha} dr$$

This integral takes on the value when $\alpha = 1$. Thus cases where it equals 1 are not appropriate for migration analysis.

If $\alpha > 1$, the integral takes on the value:

$$(8.2) \quad \int_{r=R}^{\infty} ar^{-\alpha} dr = \left| \frac{ar^{-(\alpha+1)}}{1-\alpha} \right|_R^{\infty} = \frac{a}{\alpha-1} R^{1-\alpha}$$

Since the curve $P(r)$ is a probability curve, the integral from $R=0$ to ∞ must take on the total value of (1). However, as R approaches 0, the value of the integral approaches infinity. Therefore, when migration analyses employ this variety of decay function the origin for the curve R is arbitrarily set at some value \bar{b} which is not equal to zero.

$$\text{When } \frac{a}{\alpha-1} \bar{b}^{1-\alpha} = 1; \quad \bar{b} = \frac{1-\alpha}{\sqrt{(\alpha-1)}} a.$$

This analysis assumes that the distance \bar{b} defines the extent of the local labor market. Observed migration as a proportion of total migration is

$$\mu = (a/\alpha-1) R^{1-\alpha}$$

For the special case where $\alpha = 2$, then $\mu = a/R$. As R approaches \bar{b} , the value of μ approaches 1. The proportion may be calculated as it changes with R in the form $d\mu/dR = a \ln R$. While the value of this expression changes rapidly for small values of R, it changes slowly for large values of R.

These simple findings make it possible to suggest some solutions to the first aggregation problem. If one wants to find the proportion of total migration that has been observed, then the value of μ gives the answer (when working with areal units of different sizes whose μ values can be calculated.) For large values of R, $d\mu/dR$ does not change very rapidly. Even if large units do not equal each other exactly, the differences between their μ values are nevertheless small. Small units, however, have μ values which are very sensitive

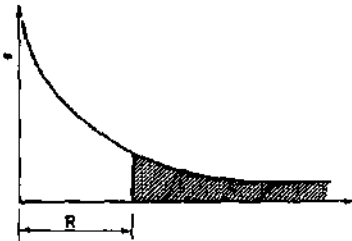


fig. 2

to change in R and thus there may be large variations in between them. Thus, there is an advantage to be gained by aggregating small areal units into larger ones.

THE CALCULATION OF μ FOR THE EXPONENTIAL FUNCTION :

For this second function the volume of migration outside a circle of radius R is calculated as:

(8.3)

$$\int_{r=R}^{\infty} P(r) dr = \int_{r=R}^{\infty} c e^{-\alpha r} dr = \frac{c}{\alpha} e^{-\alpha R}$$

This expression is defined for $r=0$. Total migration is

$$\int_{r=0}^{\infty} c e^{-\alpha r} dr = \frac{c}{\alpha} . \text{ Then } \frac{c}{\alpha} = 1 \text{ and as a result } \mu = e^{-\alpha R} .$$

To find the change in μ with an increase in R, it is necessary

to find the derivative of μ according to R, $\frac{d\mu}{dR} = -\alpha e^{-\alpha R}$. As

R takes on large values, the change in μ diminishes. In this case the solution to the aggregation problem is the same as that found for the gravity formula.

THE CASE OF THE LOGNORMAL DECAY FUNCTION

The volume of migration which leaves the circle of radius R is

(8.4)

$$\int_{r=R}^{\infty} P(r) dr = \int_{r=R}^{\infty} \frac{2r}{v} e^{-(\log r)^2} dr$$

After performing the necessary steps this expression takes the form

(8.5)

$$\int_{r=R}^{\infty} P(r) dr = \frac{2e\sqrt{\pi}}{v} \int_{\sqrt{2} \ln R/e}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du$$

The expression within the integral is a cumulative normal distribution. If this is called Φ_R then this takes the form

$$\int_{r=R}^{\infty} P(r) dr = \frac{2e\sqrt{\pi}}{v} \Phi_R . \text{ The values of } \Phi_R \text{ may be calculated}$$

from tables for the normal distribution. In the special case

where $R=e$, $\int_{r=e}^{\infty} P(r) dr = \frac{2e\sqrt{\pi}}{v}$. Since the function

$P(r) = \frac{2r}{v} e^{-(\log r)^2}$ is not defined for $r=0$, total

migration is calculated for a local labor market with a radius that commences at an arbitrary point at distance \bar{b} . As required

$\frac{2e\sqrt{\pi}}{v} \Phi_{\bar{b}} = 1$. This equality may be used to calculate the \bar{b}

distance. In this case $\mu = \frac{\Phi_R}{\Phi_{\bar{b}}}$. The general recommendations

for using this type of decay function correspond to those found for the preceding two functions.

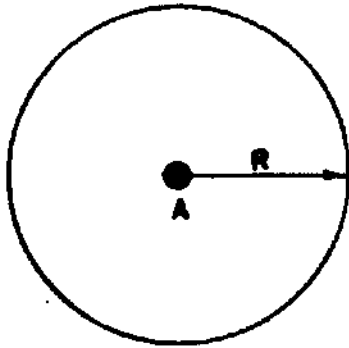


fig. 3

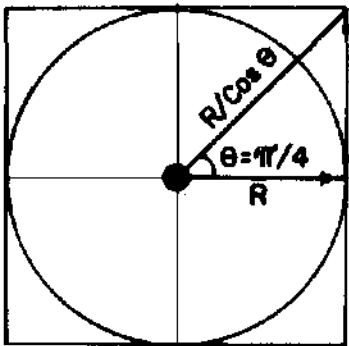


fig. 4

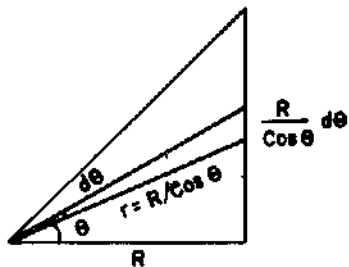


fig. 5

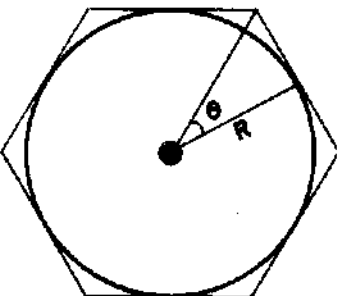


fig. 6

Problem Two: The Influence of Shape on Out Migration from Regular Geometric Units with a Settlement at the Center.

Let us first calculate out migration for a set of areas where the basic shape is composed of a circle of radius R. Around this circle we describe a hexagon, square, equilateral triangle, and a rectangle whose angles between its diagonals are 60° and 120°. Then, in order to separate out the influence of shape on these areas, we estimate the volume of migration for different shapes of equal area.

The Circle

(9)

Observed total migration for the particular gravity function where $\alpha=2$, expressed here as (P_{OM})

$$cP_{OM} = \int_{r=R}^{\infty} P(r) dr = \frac{a}{R}$$

was shown in the preceding section.

The Square

(10)

For a square described around the same circle of radius R, total migration may be found by using the polar coordinates. If migration for section $\theta = \pi/4$ is known, then total migration will be 8 times greater.

$$\begin{aligned} kP_{(OM)} &= 2.4 \int_{\theta=0}^{\pi/4} \frac{cP_{(OM)}}{2\pi r} r \cdot d\theta \\ &= 2.4 \int_{\theta=0}^{\pi/4} \frac{a \cos \theta}{2\pi r} d\theta \\ &= \frac{a}{R} \frac{2\sqrt{2}}{\pi} \end{aligned}$$

The quantity of out migration from the square is less than that from the circle of equal radius. Since the two figures are not of equal area, this result cannot be said to arise purely from differences in their shape.

The Hexagon

Out migration from a hexagon described around the same circle may be calculated in the same way.

(11)

$$hP_{(OM)} = 2.6 \int_{\theta=0}^{\pi/6} \frac{a}{2R} \cos \theta d\theta = \frac{a}{R} \cdot \frac{3}{\pi}$$

In this case the amount of observed migration is less than in the case of the circle of radius R.

The Equilateral Triangle

Out migration from an equilateral triangle described around the same circle is

(12)

$$eP_{(OM)} = 2.3 \int_{\theta=0}^{\pi/3} \frac{a}{2\pi R} \cos \theta d\theta = \frac{a}{R} \cdot \frac{3\sqrt{3}}{2\pi}$$

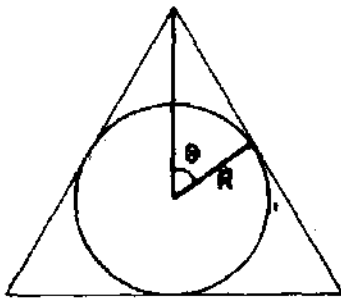


fig. 7

The Rectangle

For ease of calculation a rectangle is selected whose acute angle at the intersection of the diagonals is $\pi/3$. The quantity of out migration from this rectangle is;

(13)

$$rP_{(OM)} = 2.2 \int_{\theta=0}^{\pi/6} \frac{a}{2\pi \cdot \sqrt{3} \cdot b/2} \cos\theta d\theta + 2.2 \int_{\theta=0}^{\pi/3} \frac{a}{2\pi b/2} \cos\theta d\theta$$

$$= \frac{a}{b} \frac{8}{\pi\sqrt{3}}$$

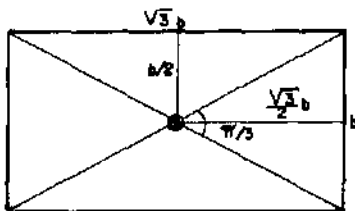


fig. 8

The volume of migration calculated above for each shape involves different areas. In the table below area is held constant for all shapes to show the influence of shape on the observed volume of migration.

Shape	Area	Perimeter Enclosing This area	Amount of Out Migration
Circle	πR^2	$2\pi R = 6.28 R$	$\frac{a}{R} = 1.000$
Hexagon	πR^2	$\frac{12 R}{\sqrt{3}} \cdot \sqrt{\frac{\pi}{2\sqrt{3}}} = 6.60 R$	$\frac{a}{R \sqrt{\frac{\pi}{2\sqrt{3}}}} \cdot \frac{3}{\pi} = 1.024$
Square	πR^2	$8R \cdot \sqrt{\frac{\pi}{4}} = 7.10 R$	$\frac{a}{R \sqrt{\frac{\pi}{4}}} \cdot \frac{2\sqrt{2}}{\pi} = 1.037$
Equilateral Triangle	πR^2	$6\sqrt{3}R \sqrt{\frac{\pi}{3\sqrt{3}}} = 8.10 R$	$\frac{a}{R \sqrt{\frac{\pi}{3\sqrt{3}}}} \cdot \frac{3\sqrt{3}}{2\pi} = 1.064$
Rectangle	πR^2	$2R(1+\sqrt{3}) \sqrt{\frac{\pi}{\sqrt{3}}} = 7.40 R$	$\frac{a}{R \sqrt{\frac{\pi}{\sqrt{3}}}} \cdot \frac{8}{\pi\sqrt{3}} = 1.115$

This table points to some interesting conclusions about the way the shape of the areal unit we use to observe migration affects our estimates.

- The circle has the lowest volume of migration. It has the shortest perimeter and thus is the most compact figure.
- In general, as the length of the perimeter increases the quantity of out migration increases.
- The figure's shape in addition to the length of its perimeter affect the amount of migration as may be seen by comparing the rectangle and the equilateral triangle. Although the equilateral triangle has a longer perimeter it produces less out migration.
- Within the limited number of regular shapes there is a difference in observed migration of 11.5%.
- All figures examined in this section are convex; one may expect to find that non-convex figures have an even greater influence on the results.

CONCLUSION

Migration estimates are affected by the way both time and areal units are aggregated. Different ways of estimating movement produce specific and predictable biases. Thus, explanatory variables selected to analyse migration estimates should not refer to more precise periods of time and age groups than the estimates themselves. Estimates spread movers and periods of possible movement over broader time spans than many researchers have assumed; this presents itself as one explanation for the rather unsatisfactory performance of migration models in explaining actual movement.

A second factor that has slowed migration models is the influence of the size and shape of different areal units on the estimates. By locating a single source of migration at the center of an areal unit and drawing only on our knowledge that the volume of migration from a point declines with distance, we examined the effect on our estimates by varying the size and shape of our units. First, we have pinpointed the areal features that make aggregating small units into larger ones a useful approach for reducing error. Second, we have found that the quantity of out migration we observe with our estimates is increased by lengthening the perimeter of the areal units we use. When handling areal units whose perimeters and shape vary, we must standardize for the effect of these differences. Within the limited number of shapes with small areal differences examined in this article, we found that areal configuration alone can affect estimates of observed migration by as much as twelve percent. If the areal units' shape were more varied and complex as in the actual administrative units used for censuses, we could expect the influence of geometry on the estimates to increase still further.

The case selected here of a single migration - producing point is highly oversimplified. For migration to occur there must be interaction with other locations in an environment composed of different opportunities structured again in space. This introduces a set of complicating factors since the migration - sending and receiving must be located within areal units in a fashion which distributes them outside the unit's geometric center. Increasingly complex features may be handled on a step-by-step basis as done in this article to produce a theoretical framework to direct the researcher in constructing a migration model and in interpreting its empirical results.

GÖÇ GÖZLEMLERİNDE ZAMANDA VE ALANDA TOPLULAŞTIRMA SORUNLARI

ÖZET

Göçe ilişkin çalışmalarımızda kullanılan veriler, yönetimsel kararlarla belirlenmiş nüfus sayımı tarihleri arasındaki zaman diliminde ve il gibi yönetimsel alan birimlerinden olan göçleri göstermektedir. Başka bir deyişle inceleme yapan kişinin amaçlarından bağımsız olarak yönetimsel amaçlara bağlı olarak zamanda ve mekanda toplulaştırılmış göç gözlemleri kullanılmaktadır. Gözlenen göç miktarı ise, zamanda ve

mekanda yapılan toplulaştırmanın düzeyine göre değişmektedir. O halde bir göç araştırmacısının kullandığı verinin özelliklerini doğru olarak değerlendirebilmesi için, göç gözlemlerinde zamanda ve mekanda toplulaştırmanın etkisini çok iyi bilmesi gerekir. Bu önemine karşın, göç gözlemlerinin toplulaştırma sorunu kuramsal olarak çok incelenmemiştir. Bu yazıda yapılacak olan söz konusu kuramsal eksikliği gidermeye dönük bir araştırmadır.

Son on yıl içinde göçe ilişkin kuramsal ve görgül çalışmalar gözden geçirildiğinde her iki tür toplulaştırmanın önemi daha iyi anlaşılmaktadır. Yüksek düzeyde toplulaştırılmış verilere dayanan hareketlilik incelemelerinde şaşılacak bir denge bulunabilir. Oysa incelemede kullanılan "gözlem birimlerinden" daha küçük birimler kullanılması halinde ise bulunan denge birdenbire yok olabilir. Eğer araştırmacı bir alansal birimi, alt birimlere bölmenin gözlemler üzerinde yapacağı etkiyi bilirse, kullandığı verilerin neteliğine bağlı olarak ulaştığı sonuçları irdeliyebilir.

Göçün zaman ve mekanda toplulaştırılması sorununu inceliyen bu yazı üç bölümden oluşmaktadır. Birinci bölümde, bir süreç olarak göç toplumsal politikalar açısından incelenirken, göçün tanım ile zamanda ve mekanda toplulaştırmanın bağıntısı kurulmaktadır. İkinci bölümde, sadece zamanda toplulaştırmanın göç gözlemlerine etkisi araştırılmaktadır. Üçüncü bölümde ise, alanda toplulaştırmanın göç gözlemlerine etkisi gösterilmektedir.

BIBLIOGRAPHY

Books

- AITCHISON, J. and BROWN, J.A. *The Lognormal Distribution*. Cambridge: Cambridge University Press, 1957.
- BURCH, T.K. "Internal Migration in Venezuela." Unpublished doctoral dissertation, Princeton University, Princeton, 1962.
- COALE, A.J. *The Growth and Structure of Human Populations: A Mathematical Investigation*. Princeton: Princeton University Press, 1972.
- COALE, A.J. and DEMENY, P. *Regional Model Life Tables and Stable Populations*. Princeton: Princeton University Press, 1966.
- ERDER L. "The Making of Industrial Bursa". Doctoral dissertation. Ann Arbor University Microfilms, (Princeton University, 1976)
- KUZNES, SIMON, ELDRIDGE and THOMAS: *Population Redistribution and Economic Growth in the United States. 1870-1950*. 3 vols. Philadelphia: The American Philosophical Society, 1964.

- LOWRY, I.S. *Migration and Metropolitan Growth: Two Analytical Models*. San Francisco: University of California, 1966.
- SHORTER, F.C. *Computational Methods for Population Projections*. New York: The Population Council, 1974.
- TEKÇE, B. *Migration to Cities in Turkey: Estimates for 1955-60 and 1960-65*. Ankara: State Institute of Statistics, 1976.
- TEKELİ, İ. "Mekan Organizasyonlarına Makro Yaklaşım ve Türkiye Üzerinde bir Deneme". Unpublished associate professorship dissertation, Ankara: Middle East Technical University, 1972.
- WILSON, A.G. *Entropy in Urban and Regional Modelling*. London: Pion Press, 1970.
- YENER, S. *1965-1970 Döneminde İller Arası Göçler ve Göç Edenlerin Nitelikleri*. Ankara: Devlet Planlama Teşkilatı, 1977.

Articles

- CAVALLI-SFORZA. The Distribution of Migration Distances: Models and Applications to Genetics *Entretiens de Monaco*, May, 1962.
- GEDİK, A. "A Causal Analysis of Internal Migration in Turkey 1965-1970." University of Washington, doctoral dissertation in progress.
- EVERT, S.L. A Theory of Migration, *Demography*, v.3, 1966, pp.411-511.
- KULLDORF, G. Migration Probabilities, *Lund Studies in Human Geography*, No.14, Lund University, 1955.
- MILLER, E. A Note on the Role of Distance in Migration Costs of Mobility Versus Intervening Opportunities. *Journal of Regional Science*, v.12. n.3, 1972, pp.475-479.
- MORRILL, R.L. The Distribution of Migration Distances, *Papers of the Regional Science Association*, v.11, 1963, pp.75-84.
- ROGERS, A. and F. WILLEKENS Spatial Population Dynamics, *Papers of the Regional Science Association*, v.36, pp.3-34.
- WOLPERT, J. Behavioral Aspects of the Decision to Migrate, *Regional Science Association Papers*, v.15, 1965, pp.159-169.