MIGRATION MODELS AND THEIR SIGNIFICANCE FOR POPULATION FORECASTS

H. ter Heide[1]

INTRODUCTION: THE P_1P_2/D HYPOTHESIS

IN recent years, demographers in several countries have developed mathematical models describing, and to some extent explaining, patterns of internal migration. It may be considered worthwhile to review a number of these models, and to evaluate, by reciprocal comparison, their respective merits and shortcomings. This may be the more interesting to American readers as some of the European studies on the subject are probably unknown in the United States.

Among the composers of migration models have been several who have claimed that their models could, in principle at least, be used in regional population projections. However, to the best of the present author's knowledge, no such practical use of any of the models has yet been made. The proposed review of the various models will make it possible to evaluate the significance of migration models for population projections and forecasts in a general way. Thus, the two objects of this article have been stated.

The most promising point of departure for a treatise on migration models is Zipf's well-known P_1P_2/D hypothesis. It is true that not all models are based on Zipf's formula. This formula does, however, provide a fixed point from which even those models which are not derived from it may be seen in their proper perspective.

It will be remembered that the P_1P_2/D hypothesis states that migration—as well as other types of interregional exchange of persons, goods, or information—is directly proportional to the product of the populations of the two regions involved, and

[1] Netherlands Government Physical Planning Service, The Hague, The Netherlands.

inversely proportional to the distance between the regions. In formula, this reads

$$M = \frac{P_1 P_2}{D} \qquad [1]$$

where M indicates gross migration between regions 1 and 2.[2] As stated, this is a logical rather than a mathematical formula; we will give it the following more general form:[3]

$$M_{gh} = kP_g P_h D_{gh}^{-\alpha} \qquad [2]$$

where k and α are positive variables, the values of which are dependent on situational factors.

Scandinavian authors are inclined to use a slightly different formula, i.e.:

$$m = kD^{-\alpha} \qquad [3]$$

where m indicates the relative volume of migration between two regions or the gross migration rate, which is defined

$$m_{gh} = \frac{M_{gh}}{P_g P_h} \qquad [4]$$

The Scandinavian formula has been introduced by Kant.[4] It is clear that formulas (1) and (2) are essentially alike, provided that it is not considered necessary to raise the population figures in Zipf's formula to any power other than unity. Anderson has tentatively suggested that the population figures might be given an exponent other than unity.[5] His argument notwithstanding, we shall maintain in the next section that no exponent other than 1 is allowed.

[2] Zipf, George K.: HUMAN BEHAVIOR AND THE PRINCIPLE OF THE LEAST EFFORT. Cambridge, Mass., Addison-Wesley Press, 1949, pp. 386–409. Cf. also: Zipf, George K.: The $P_1 P_2/D$ Hypothesis: on the Intercity Movement of Persons. *American Sociological Review,* 1946, 11: 677–686.

[3] With one exception, we shall use our own notation throughout this article, without indicating where it differs from the various notations used by the authors cited.

[4] Kant, Edgar: Den inre omflyttningen i Estland i samband med de estniska städernas omland. *Svensk Geografisk Årsbok.* 1946, 22: 83–124 (with English summary).

[5] Anderson, Theodore R.: Intermetropolitan Migration: A Comparison of the Hypotheses of Zipf and Stouffer. *American Sociological Review,* 1955, 20: 287–292,

Both Zipf's and Kant's formulas have been tested a number of times: the former by Folger, Anderson, and Lövgren,[6] the latter, apart from by Kant himself, by Agersnap, Lövgren, and Hägerstrand.[7] The results of the tests have been satisfactory. The main weakness of the models is a tendency to overestimate short-distance migration. This is a mathematical flaw which is common to Pareto-type formulas; for when x (in this case, distance) approaches 0, y (migration) approaches infinity. In the realm of traffic-analysis, where the same difficulty was encountered, the use of "normal" formulas has been suggested. For instance:[8]

$$m = ke^{-D^2} \qquad\qquad [5]$$

To the author's knowledge, migration has never been studied with the aid of a simple "normal" formula, though Kulldorf[9] has used a logarithmico-normal formula in migration analysis. The practical value of this refinement is limited, since migration studies are generally concerned with regions of a sufficient size for the problem to remain of minor importance.

THEORETICAL CONSIDERATIONS

It will be found profitable to base the further development of our argument on a consideration of the meaning of the P_1P_2/D relationship.

Since Ravenstein wrote his famous papers,[10] the inverse relationship between the volume of migration and distance has been a matter of common knowledge. Empirical confirma-

[6] Folger, John: Some Aspects of Migration in the Tennessee Valley. *American Sociological Review*, 1953, 18: 253–260; Anderson, *op. cit.*; Lövgren. Esse: The Geographical Mobility of Labour. *Geografiska Annaler*, 1956, 38: 344–394.

[7] Agersnap, Torben: Studier over indre Vandringer i Danmark. *Acta Jutlandica*, 24: 1952, suppl. B, København, Ejnar Munksgaard, 1952; Lövgren, *op. cit.*; Hägerstrand, Torsten: Migration and Area, *in* MIGRATION IN SWEDEN: A SYMPOSIUM, Lund, Gleerup, 1957, pp. 113 ff.

[8] This formula has been suggested by the Dutch townplanner L. H. J. Angenot.

[9] Kulldorf, G.: MIGRATION PROBABILITIES, Lund, Gleerup, 1955.

[10] Ravenstein, E. G.: The Laws of Migration. *Journal of the Statistical Society*, 1885, 48: 167–219; *also* The Laws of Migration: Second Paper. *Journal of the Royal Statistical Society*, 1889, 52: 241–289.

tion of this relationship is abundant, in North American and European as well as in some Asian countries. Three circumstances have been mentioned by diverse authors to explain this inverse relationship:

1. The expense and difficulty of traveling over long distances.[11]

2. The wish to maintain contacts, either of a personal or a business nature, with the region one leaves behind.[12]

3. The fact that information concerning opportunities is easier to be had for regions at shorter distances.[13]

Considering the evidence from diverse studies which we will not mention now, it is indeed likely that all three of these factors operate to varying extents; and some of them are specifically required to explain the different degrees of susceptibility to distance of various categories of migrants. However, this need not detain us here. For a general analysis of the inverse migration-distance relation it is sufficient to work with the third, the information factor, because of the logical interrelationship between this factor and the first two: if travel between two regions is cheap and easy, and contact between them is easily maintained, this in fact means that information passes freely between them.

The recognition of the significance of the information factor in explaining the inverse migration-distance relation throws light on the role of the population variables in the Zipf formula as well. If one tried to explain the position of these variables without recourse to the information factor, the line of reasoning would have to be as follows: Assuming that migrants are selected at random, the probability that in a country with

[11] Sorokin, P. and Zimmerman, C. C.: PRINCIPLES OF RURAL-URBAN SOCIOLOGY, New York, 1929, pp. 593–594; Westefeld, Albert: The Distance Factor in Migration. *Social Forces*, 1940, 19: 213–218; Kant, *op. cit.*, p. 107 ff, p. 122; Nelson, Phillip: Migration, Real Income and Information. *Journal of Regional Science*, 1959, 1: 43–74.

[12] Westefeld, *op. cit.;* Nelson, *op. cit.*

[13] Sorokin and Zimmerman, *loc. cit.;* Westefeld, *op. cit.;* Lövgren, *op. cit.*, p. 356; Nelson, *op. cit.*

population P_t an internal migrant comes from region g with populaton P_g is P_g/P_t. This takes care of the region of out-migration; but the probability of his going to region h with population P_h is not necessarily P_h/P_t. With a random distribution of migrants, in-migration might be expected to be proportionate to the area of a region rather than to its population size. However, the experiment of substituting area for population in the model was recently tried in the Netherlands by Somermeijer, and the results were not at all favorable.[14]

An entirely different picture results when the P_1P_2/D hypothesis is considered to be descriptive of the circulation of information, and only indirectly of the pattern of migration. As has been lucidly shown by Miller,[15] on the assumption that the chances of a given piece of information originating or terminating with any person are distributed at random, the probability of a person in g with population P_g receiving a message from a person in h with population P_h is, by the rule of multiplication of probabilities, P_gP_h/P_t^2. So is, consequently, the probability of a person migrating from g to h on account of this message. The factor $1/P_t^2$ may be considered as a constant for any country within a limited time-span.

Of course, in reality chances of sending or receiving messages are unequal. For this reason one may consider attaching weights to the population figures in the formula. For instance, if the average person in g is considered twice as likely to send or receive information as the average person in the country as a whole, and the average person in h half as likely, the generalized Zipf formula should read:

$$M_{gh} = k \, 2P_g \, \tfrac{1}{2}P_h \, D_{gh}^{-a} \qquad [6]$$

It should be noted that this consideration can never occasion

---

[14] Personal communication from Mr. W. H. Somermeijer. The experiment under consideration has not been mentioned in his article to which we shall refer later.

[15] Miller, George A.: Population, Distance and the Circulation of Information. *American Journal of Psychology*, 1947, 60: 276–284. Cf. Zipf, George K.: Some Determinants of the Circulation of Information. *American Journal of Psychology*, 1946, 59: 400–421; and Zipf's comments on Miller's remarks.

exponents to be attached to the population figures, but only weights. Hence our contention that the population figures should not be raised to any power other than 1.

The suggestion to weight the population variables may be found in studies by Dodd, Stewart, Isard and Somermeijer.[16] Insofar as the population of the area of out-migration is concerned, however, the weights represent more population characteristics than just susceptiblity to the reception of information. Since the formula is symmetrical and describes gross migration, this holds in fact for both weights. The discussion of the question of weighted population figures will therefore be postponed till later in this article.

FURTHER CONSIDERATION OF THE DISTANCE FACTOR

The susceptibility of communication to sheer distance is an established fact.[17] This explains why the volume of migration falls off with increasing distance: the possibilities for getting information concerning opportunities which warrant migration decrease as distance increases.

However, it stands to reason that the passing of information is not only influenced by sheer geographical distance, but also by what might be termed technical distance: the availability of technical means of communication and transport. Besides,

---

[16] Dodd, Stuart C.: The Interactance Hythothesis. *American Sociological Review,* 1950, 15: 245–256; Stewart, John Q.: Potential of Population and Its Relationship to Marketing, *in* THEORY OF MARKETING (R. Cox and W. Alderson, eds.), Homewood, Illinois, Richard D. Irwin, Inc., 1950; and other papers by Stewart and William Warntz; Isard, Walter, *et al.:* METHODS OF REGIONAL ANALYSIS: AN INTRODUCTION TO REGIONAL SCIENCE, New York and London, Wiley & Sons, 1960, p. 543; Somermeijer, W. H.: Een analyse van de binnenlandse migratie in Nederland tot 1947 en van 1948–1957. *Statistische en Econometrische Onderzoekingen,* 1961: pp. 115–174 (with English summary), p. 144.

[17] The inverse relationship between distance and intensity of communication has been proved by social psychologists and sociologists both in laboratory and field experiments and in social reality. To mention only a few studies: Bavelas, A.: Communication Patterns in Task-oriented Groups. *Journal of the Acoustical Society of America,* 1950: 725–730; Dodd, Stuart C.: Testing Message Diffusion from Person to Person, *Public Opinion Quarterly,* 1952, 16: 247–262; Dodd, Stuart C.: Testing Message Diffusion in Controlled Experiments. *American Sociological Review,* 1953, 18: 410–416; Festinger, Leon; Schachter, Stanley and Back, Kurt: SOCIAL PRESSURES IN INFORMAL GROUPS, New York, Harper & Bros., 1950; Kuper, Leo: Blueprint for Living Together, *in* LIVING IN TOWNS, London, Cresset Press, 1953.

there is the sociological category of social distance. This implies that people are less likely to communicate the more unequal they are as regards cultural and social status traits.

The significance of one aspect of technical distance has been studied in Denmark by Agersnap. Using Kant's formula to analyze migration into the city of Århus, Agersnap found different constants and exponents when he differentiated between out-migration communities with or without railway stations. For communities with stations, the function reads

$$m = 4775D^{-1.4} \qquad [7]$$

and for communities without stations

$$m = 13050D^{-1.8} \qquad [8]$$

That is to say, for those communities not having railway stations, out-migration was more susceptible to the influence of geographic distance.[18]

Though the significance of the technical distance factor has thus been shown, it remains doubtful whether it should be considered worthwhile to introduce this factor in migration models. The aspects of technical distance are as manifold as are the means of communication or transport that have been invented, and it would be very difficult indeed to aggregate all of these aspects into one variable which could be inserted in the formula. Normally, measuring the distance along roads, railways or waterways—whatever are the main means of transportation—seems to be sufficient. Special attention need only be paid to the technical distance factor when trying to adapt a migration model for use in countries which show wide variations in development; e.g., some of the West African countries, where the coastal areas are usually more highly developed than the interior.

The concept of social distance accounts for the fact that, generally, the volume of internal migration greatly exceeds that of international migration, even if geographical distance is kept constant and if there are no legal barriers to international mi-

---

[18] Agersnap, *op. cit.*, pp. 47–48.

gration. The cultural and linguistic differences which usually exist between different countries are sufficient explanation for this phenomenon. For an example of the operation of the social distance factor on the pattern of internal migration, a case where diverse languages were spoken within a country may be mentioned. Making use of data compiled by Thirring, Kant has shown that in Hungary, before World War I, migration to Budapest from the Székely—Udvarhely region in Transsylvania was much heavier than from the much nearer Kolozsvár (Cluj)—Temesvár areas.[19] The Székely region (which is now, with the rest of Transsylvania, part of Romania) houses a Hungarian-speaking population, as opposed to the Romanians in western-Transsylvania.

The one known migration model which includes a social distance factor is Somermeijer's. He took advantage of the fact that in the Netherlands one of the most important social characteristics, and one about which census data are available, is religion. By a simple mathematical procedure, Somermeijer calculated a measure R_{gh} for the difference in religious composition of the population between any two regions g and h, and inserted this measure in the Zipfian formula as follows:

$$M_{gh} = kP_gP_hD_{gh}^{-a}(1 + \beta R_{gh})^{-1} \qquad [9]$$

As R (unlike D) may be = 0, a Pareto distribution was not feasible for this variable, and a hyperbolical function was chosen instead. In actual fact, β was found to be relatively small (which cannot, however, be taken as definitely disproving the influence of religion on the pattern of migration).[20]

Further experiments with social distance factors in migration models would be welcome, though it will not always be easy to find a suitable indicator. There are signs, however, that in the United States, too, something of the nature of social dis-

[19] Kant, Edgar: Umland Studies and Sector Analysis, *in* STUDIES IN RURAL-URBAN INTERACTION, Lund, Carl Bloms, 1951. This is the English summary of a study published previously in Swedish: Kant, Edgar: Omlandsforskning och sektoranalys, *in* TÄTORTER OCH OMLAND (G. Enequist, ed.), Uppsala, Lundequist, 1951.

[20] Somermeijer, *op. cit.*

tance is operative in shaping the pattern of internal migration. Otherwise it would be hard to explain why several authors report less migration across state boundaries than simple (geographical) distance models would lead one to expect.[21] Hypothetically, it might be argued that the social distance which, in consequence of cultural differences, exists between the North and the South of the Union, must be of some measurable significance for the internal migration pattern, too.

OPPORTUNITIES

People, it may be imagined, migrate to a certain place because they expect to find opportunities there: opportunities *to get a job, a new house, education in specific schools, etc.* The formulas we have mentioned up till now treat migration as if these opportunities were distributed randomly over the country. Of course, in reality this is not the case. Therefore, particularly to describe and explain net migration, it is necessary to extend the models so as to include factors indicative of the distribution of opportunities.

In the case of several of the models which have been developed to include such factors, Zipf's P_1P_2/D hypothesis is bypassed. Stouffer, in particular, emphatically denied any direct relevance of the distance factor. His "intervening opportunities hypothesis" proposes that the number of persons going a given distance is directly proportional to the number of opportunities at that distance and inversely proportional to the number of intervening opportunities.[22] Later, Stouffer took the logical step of also recognizing intervening "competing migrants": ". . . . everything else being equal, the attractiveness of city Y for migrants from city X will depend, at least to some extent, on how many potential migrants are closer to Y

[21] Anderson, *op. cit.*, p. 290; Hillery, George A. and Brown, James S.: Some Conclusions on Migratory Streams from a Study of the Southern Appalachians. [Paper delivered at the 1961 meeting of the Population Association of America.] See *Population Index*, 1961, 27: 210–211.

[22] Stouffer, Samuel A.: Intervening Opportunities: A Theory Relating Mobility and Distance. *American Sociological Review*, 1940, 5: 845–867.

than are the potential migrants in X."[23] A few years previously, Porter had already introduced a model, mathematically rather complicated, developed according to the same principles which underlie Stouffer's second model.[24]

The intervening opportunities hypothesis has been tested with, as Isbell put it, "encouraging" results by Stouffer himself, Bright and Thomas, Isbell, Strodtbeck, Folger, and Hägerstrand.[25] Diverse operational definitions have been used, but most authors have accepted Bright and Thomas' working definition of opportunities as the actual number of persons living in the region concerned but born elsewhere. The main difficulty with the intervening opportunities hypothesis is the appearance of what Bright and Thomas have called "non-competitive opportunities." That is to say, different types of opportunities— jobs, houses, schools, etc.—are relevant for different migrants, and they cannot all be represented by one indicator. Besides, leaving out the population and distance variables as such is going to the other extreme. It is true that intervening opportunities do indicate distance in a sense, but neither the communication aspect of the distance factor nor the economic and social costs aspect is fully represented in the intervening opportunities concept. We do not feel, therefore, that Stouffer's model brings us much further along.

The same may be said of the additive models proposed by Price and by Tarver.[26] Price introduced his model to the 1959 World Population Conference in the following words: "In order

[23] Stouffer, Samuel A.: Intervening Opportunities and Competing Migrants. *Journal of Regional Science,* 1960, 2: 1–26, p. 7.

[24] Porter, R.: Approach to Migration through its Mechanism. *Geografiska Annaler,* 1956, 38: 315–343.

[25] Bright, Margaret L. and Thomas, Dorothy Swaine: Interstate Migration and Intervening Opportunities. *American Sociological Review,* 1941, 6: 773–783; Isbell, Eleanor C.: Internal Migration in Sweden and Intervening Opportunities. *American Sociological Review,* 1944, 9: 627–639; Strodtbeck, Fred L.: Equal Opportunity Intervals. *American Sociological Review,* 1949, 14: 490–497; Folger, *op. cit.;* Hägerstrand, *op. cit.,* pp. 120 ff.

[26] Price, Daniel O.: A MATHEMATICAL MODEL OF MIGRATION SUITABLE FOR SIMULATION ON AN ELECTRONIC COMPUTER, (*Proceedings of the International Population Conference,* Wien, 1959), Wien, International Union for the Scientific Study of Population, 1959; Tarver, James D.: Predicting Migration. *Social Forces,* 1961, 39: 207–213.

to have something specific to deal with, let us consider a simple case where we have five individuals with four characteristics each and three areas, each area having three characteristics. The individuals will be symbolized as capital I_i, with $_i$ varying from 1 to 5. The four characteristics of these individuals will be A_i, B_i, C_i, and D_i. The areas will be symbolized as A_j, with j varying from 1 to 3. The characteristics of these areas will be X_j, Y_j, and Z_j. . . . Let us symbolize the probability of individual I (I_i) moving from area J (A_j) to area K (A_k), as: P $[I_iA_jA_k]$. In terms of the symbols given above we can then write:

$$P[I_iA_jA_k] = f\ (A_i,\ B_i,\ C_i,\ D_i,\ X_j,\ Y_j,\ Z_j,\ X_k,\ Y_k,\ Z_k)"$$

Further development of this formula gives:

$$P(I_iA_jA_k)= f_1(A_i)+ f_2(B_i)+ f_3(C_i)+ f_4(D_i)+\ f_5(X_j)+ f_6(Y_j)$$
$$+ f_7(Z_j)+ f_8(X_k)+ f_9(Y_k)+ f_{10}(Z_k)+ f_{11}(X_j - X_k)+ f_{12}(Y_j - Y_k)+$$
$$f_{13}(Z_j - Z_k).$$

(It should be noted that in these two formulas we have used Price's notation rather than our own. Price's P, for instance, which denotes probability, should not be confused with our population variable.)

Price proposes to include in this model such variables as color, sex, age, employment status, marital status, etc. Work on the model has been suspended awaiting the availability of 1960 census data.

As compared to the P_1P_2/D hypothesis, this model is, in a sense, a step backward. The "opportunities" factors are introduced only at the expense of losing the communication factor. The inverse relation between volume of migration and distance, since Ravenstein one of the best-established migration phenomena, does not appear in the model. Also there are reasons to consider multiplicative formulas mathematically superior to additive models.

A model which does take advantage of the Zipf hypothesis is the one recently developed in the Netherlands by Somermeijer.[27] We have already seen that he introduced a social distance factor by inserting a variable measuring religious differentiation. From the resulting formula, which we reproduced as nr (9) above, Somermeijer derives a model describing net migration as follows:

It is assumed that formula (9) also describes migration from g to h (or from h to g), provided factors indicating the relative attractiveness of g and h respectively are introduced. The difference between these "attractiveness factors" decides the amount and the direction of migration. That is to say: let $F_{m,g}$ indicate the value of an attractiveness factor m in g, and $F_{m,h}$ that of the corresponding factor in h, then if

$$F_{m,h} - F_{m,g} > 0$$

migration will take place from g to h—as far as this factor is concerned, anyway—; and vice versa. Of course there are a number of attractiveness factors, each supposedly influencing a different class of migrants, with the result that migration in both directions takes place. To an even larger extent, the fact that the attractiveness factors may have different subjective values for different individuals must be considered to be responsible for the appearance of migratory streams in opposite directions. For that reason, Somermeijer introduces the attractiveness factors in his formulas as $\overline{F}_{m,h}$ and $\overline{F}_{m,g}$, indicating the average values of $F_{m,\,h_{i_h}}$ and $M_{m,\,g_{i_g}}$ for all individuals i_h in h and i_g in g resp. In practice of course the objective values of the factors are used.[28]

---

[27] An important source of inspiration for Somermeijer which should be mentioned here is the prewar British study by Makower *et al.* (Makower, H., Marschak, J. and Robinson, H. W.): Studies in Mobility of Labour: a Tentative Statistical Measure, OXFORD ECONOMIC PAPERS, 1938, 1: 83–123; Studies in Mobility of Labour: Analysis for Great Britain, Part I, *ibid.*, 1939, 2: 70–97; *ibid.*, Part II, 1940, 5: 39–62.

[28] A word concerning the social distance factor is in order here. Theoretically, this factor has an "attractiveness" aspect as well as the communication aspect, which we have hitherto considered. For example, religious differences between two regions

The factors should, obviously, be combined, after having been weighted.[29] In addition, some migration takes place as a direct result of the differences in the (subjective) values of the attractiveness factors for different individuals (even apart from accidental factors). To provide for these inter-individual differences and accidental factors the constant C_{0m} is introduced. Ideally, the C_{0m}'s for opposite migration streams cancel each other out. We now have

$$M_{g \to h} = \tfrac{1}{2} P_g P_h D_{gh}^{-\alpha} (1 + \beta R_{gh})^{-1} \underset{m}{\sum} \{ C_{0m} + C_{lm} (\overline{F}_{m,h} - \overline{F}_{m,g}) \}$$
[10]

and, of course

$$M_{h \to g} = \tfrac{1}{2} P_g P_h D_{gh}^{-\alpha} (1 + \beta R_{gh})^{-1} \underset{m}{\sum} \{ C_{0m} - C_{lm} (\overline{F}_{m,h} - \overline{F}_{m,g}) \}$$
[11]

These two formulas describe migration streams in opposite directions. By definition, adding the formulas should give us gross migration. The result of the addition reads:

$$M_{gh} = P_g P_h D_{gh}^{-\alpha} (1 + \beta R_{gh})^{-1} \underset{m}{\sum} C_{0m}$$
[12]

This is the same as formula (9), except that the constant k has now been replaced by $\underset{m}{\sum} C_{0m}$. To calculate the balance of migration in h as far as migration between g and h is concerned one naturally substracts formula (11) from formula (10):

$$M_{h(g)} = P_g P_h D_{gh}^{-\alpha} (1 + \beta R_{gh})^{-1} \underset{m}{\sum} C_{lm} (\overline{F}_{m,h} - \overline{F}_{m,g})$$
[13]

Somermeijer tested his model with good results on migration

may lessen the attractiveness of one for migrants from the other. It might be argued that, therefore, the social distance factor might be accorded similar treatment in the model as the F-factors. In fact, however, this is not possible, because while migration is, in principle, positively correlated with the differences between $\overline{F}_{m,h}$ and $\overline{F}_{m,g}$, it is inversely related to the differences in religious composition of the two populations, even when the "attractiveness" aspects of these differences are considered. Hence the negative power to which the social distance factor is raised.

[29] The desire to introduce several attractiveness factors separately into the formula —so as to measure their influence separately—precludes the possibility of treating them like the "specific indices of level" in Dodd's interactance model as simple multipliers of the population variables. Dodd, Stuart C.: The Interactance Hypothesis. *American Sociological Review*, 1950, 15: 245–256.

between Dutch provinces, using attractiveness factors such as per capita income, per cent unemployed, degree of urbanization, recreational resources and quality of dwellings. Fitting the constants as well as possible by iteration, correlation co-efficients for the effects of the attractiveness factors on net migration in the neighborhood of .9 were obtained.[30]

MIGRATION DIFFERENTIALS

It is a matter of common knowledge that different categories of people show different migration rates. The age differential, for example, has been definitely established: migration generally reaches a peak in the young adult ages (20–29 age bracket). Differentiation according to sex, marital status, family composition, occupation, etc., has been less well established. Dorothy Thomas' 1938 RESEARCH MEMORANDUM is still the most comprehensive source of information on differential migration data.[31]

Differential migration is an integral part of Price's model; as a matter of fact, he expressly recognizes "characteristics of individuals" in his formula. However, in Somermeijer's model as well provision for the inclusion of differential migration factors may be made. Somermeijer himself suggests weighting the population variables in case the populations in g and h differ in composition as regards categories with different migration tendencies.[32]

As far as differential migration is concerned the main difficulty is to get sufficient information to calculate the differentials. Usually, differential migration figures according to age,

[30] This treatise deliberately omits to discuss two other recent models, viz. Thomlinson's and the one submitted to the 1961 World Population Conference by Muhsam. It is felt that both these models are theoretically not sufficiently evolved to be of much significance for our discussion. Thomlinson's model, especially, explains very little, as it does not specify which opportunities are represented in the rates which are calculated, and besides it begs the question insofar as certain hypotheses have to be accepted without proof as premises from which the model is developed. Thomlinson, Ralph: A Model for Migration Analysis. *Journal of the American Statistical Association*, 1961, 56: 675–686.

[31] Thomas, Dorothy Swaine: RESEARCH MEMORANDUM ON MIGRATION DIFFERENTIALS, New York, Social Science Research Council, 1938.

[32] Somermeijer, *op. cit.*, p. 144.

marital status, etc. are available for a few years only, if at all. Price, in his 1959 paper, mentions 1947–1948 data for the United States, and comments that for some other differentials he will have to await the 1960 Census.[33] Even in the Netherlands with their continuous population register, data on migration by age are available for 1950–1951 only (they will become available for 1960), while any statistics that exist on migration by occupation are grossly unreliable. (Sex and marital status differentials, on the other hand, are continuously available.)

The limited availability of differential migration data has serious consequences, especially where migration models are required for population projection purposes, because there is no reason to think that, for instance, the age-structure of the migrant group which has been established in any one year will forever remain the same. As a matter of fact, it may be considered quite likely that the changing age-structure of the population at large will influence age-differentials in migration.

REGIONAL POPULATION PROJECTIONS

The question whether the reviewed migration models have any significance for regional population projections now arises. Price as well as Somermeijer are inclined to suggest that it may be possible to make use of migration models when trying to develop population projection models.[34] To evaluate this "claim" it is necessary first to review existing methods of regional population forecasting.

It is clear that the use of migration models will come into consideration only when distribution methods for regional population forecasts are dealt with, i.e. methods by which the total population of a country according to a national forecast is distributed over the regions. The simpler migration formulas, such as Zipf's hypothesis and even the intervening opportunities model, may be used, and in fact have been used, to

---

[33] Price, *op. cit.,* p. 667.
[34] Price, *op. cit.,* p. 672; Somermeijer, *op. cit.,* p. 121.

describe migration to and from a certain town or region. Since these models are incomplete as far as migration factors are concerned, nobody would suggest that they might be used for population projections, and in fact they are unable to predict net migration. The more advanced models, such as Price's and Somermeijer's, are distribution models themselves: they describe patterns of migration between all of the regions—states, provinces—into which a country may be divided, rather than migration between one of these regions and the others.

Our discussion of distribution methods for regional population forecasts will deal mainly with the methods developed over the past thirty years in the Netherlands. In this country, the impetus for developing distribution methods issued from the discovery that local and regional authorities, when preparing forecasts for their particular territories, tended (and tend) to overestimate their share in the national population, which, if all regions had prepared forecasts, would result in the sum of the combined regional forecasts exceeding the national forecast. To check this tendency, some authors tried to find a relation between population development in the region under study and in the country as a whole, which involved putting the regional forecast into the framework of a national forecast.[35] As early as 1935, Van Lohuizen and Delfgaauw suggested taking the national forecast as a starting point, and developing models by which the total population according to the national forecast would be distributed over the regions.[36] To implement this suggestion the Standing Committee of the Government Physical Planning Service appointed a Committee for Regional Population Forecasts, which started its work just

[35] Angenot, L. H. J.: DE TOEKOMSTIGE LOOP DER BEVOLKING IN NEDERLAND EN IN HET HAVENGEBIED VAN ROTTERDAM, Rotterdam. Vereeniging Nieuw Rotterdam, 1934. An American example of the same type of projection may be found in the report, FUTURE DEVELOPMENT OF THE SAN FRANCISCO BAY AREA, 1960–2020, prepared and published for U.S. Army Engineer District, San Francisco Corps of Engineers, by Office of Area Development, Business and Defense Services Administration, U.S. Department of Commerce, Washington, D.C., U.S. Govt. Printing Office, 1959.

[36] van Lohuizen, Th. K. en Delfgaauw, G. Th. J.: De toekomstige verdeeling van bevolking en van woningvoorraad over het land, *Tijdschrift voor Volkshuisvesting en Stedebouw,* 1935, 16: 203–209, 226–230.

after World War II. This committee has since issued two reports.[37]

A number of distribution formulas have been devised both by the committee and by others. To provide an example, one of the most simple formulas may be mentioned here. This one simply postulates that population growth will be divided over the country in the same way as in the past, that is to say in the period chosen as a base. Symbolizing the national population as P and the population of any region i as p_i, and adding subscripts 1, 2 and 3 to indicate, respectively the beginning of the base period, the end of the base period, and the projection date, we get

$$p_{i3} = p_{i2} + (p_{i2} - p_{i1}) \frac{P_3 - P_2}{P_2 - P_1} \qquad [14]$$

It is possible to distinguish between demographic and economic distribution formulas. The latter result in projections of the economically active rather than the total population, though the total population can of course be derived from the employment projection. The one above is a typical demographic formula, as was Van Lohuizen and Delfgaauw's original 1935 model. Among the economic formulas may be numbered the input-output models developed in the United States by Leontieff and Isard.[38] The Committee for Regional Population Forecasts worked with an economic model closely connected with the basic-nonbasic ratio.[39] Steigenga discussed this type of economic model in several papers.[40] While both the Com-

[37] Commissie voor Regionale Bevolkingsprognoses: EERSTE RAPPORT, 's-Graven-hage, Staatsdrukkerij, 1952; Commissie voor Regionale Bevolkingsprognoses: TWEEDE RAPPORT, 's-Gravenhage, Staatsdrukkerij, 1959. Both reports include English summaries. The committee's present program, from which we quote an example (formula 14), has for a large part been carried out by J. Godefroy. Chairman of the committee is L. H. J. Angenot, and this writer serves as secretary to the committee.

[38] Leontieff, Wassily W.: THE STRUCTURE OF AMERICAN ECONOMY, 1919–1939, New York, Oxford Univ. Press, 1951; Leontieff, Wassily W.: *et al.*, STUDIES IN THE STRUCTURE OF AMERICAN ECONOMY, New York: Oxford Univ. Press, 1953; Isard, *op. cit.*, esp. Chpt. 8.

[39] Commissie voor Regionale Bevolkingsprognoses: TWEEDE RAPPORT, Deelrapport E.

[40] Steigenga, Willem: Het vraagstuk der regionale bevolkingsprognose, *Tijdschrift*

mittee and Steigenga took advantage of the differentiation in basic and nonbasic groups of industries available in Dutch statistics,[41] other methods of assessing the basic-nonbasic ratio, which can also be used in other countries and have in fact been used in the United States, have been developed and applied to (albeit non-distributive) local projections by Klaassen c.s. and others.[42]

THE SIGNIFICANCE OF MIGRATION MODELS FOR POPULATION PROJECTIONS

At first glance, it might be thought that migration models could be used to advantage in the sphere of demographic projection formulas. In combination with regional natural increase projections, to be derived from the national projection by some kind of ratio method or brought in line with the national projection by iteration, migration projections based on application of migration models might yield acceptable forecasts. However, in general the only reason to restrict oneself to demographic projections is the lack of data concerning economic aspects, and since it is difficult to imagine a set of attractiveness factors in a migration model which would not include economic factors, this same lack of data makes the application of migration models virtually impossible.

As regards economic projections, the original idea was that,

voor Economische en Sociale Geografie, 1954, 45: 80–88 (with English summary); Steigenga, Willem: INDUSTRIALIZATION-EMIGRATION: THE CONSEQUENCES OF THE DEMOGRAPHIC DEVELOPMENT IN THE NETHERLANDS, The Hague, Mart. Nijhoff, 1955, pp. 26 ff.

[41] Thanks to: Stuwende en verzorgende bedrijven (Rapport Commissie van Lohuizen), *Tijdschrift voor Economische en Sociale Geografie*, 1952, 43: 80–87.

[42] Klaasen, L. H., Van Dongen Torman, D. H. en Koyck, L. M.: HOOFDLIJNEN 1970, Leiden, Stenfert Kroese, 1949, pp. 69 ff.; Schoonhoven, J.: Een methode tot daarvan voor de bevolkingsprognose in kleinere gebieden. *Tijdschrift voor Economische en Sociale Geografie*, 1953, 44: 157–159; Alexanderson, Gunnar: THE INDUSTRIAL STRUCTURE OF AMERICAN CITIES, Stockholm, Almqvist & Wiksell, 1956, p. 14 ff.; FUTURE DEVELOPMENT OF THE SAN FRANCISCO BAY AREA (cf. note 32), p. 34 ff. The methods developed in these four studies are based on the same principle, but vary slightly as regards technical details.

in connection with true economic models, migration projections are not only superfluous but positively confusing. Steigenga, for instance, states in so many words that the migration forecast is an outcome, not one of the data of an economic projection.[43] The same is implied in the models developed by Isard.[44] In non-distributive projections, the idea that the economic rather than the demographic development should be the subject of the projection is quite firmly established.

However, Steigenga's proposition implies the premise that migration is regulated by economic factors only. In the not too distant past, this may have been a safe premise. At present, however, such phenomena as, to mention only one example, the increasing migration into Florida and other "good climate" areas in the United States cast doubt as to whether economic factors are still of exclusive significance. In many countries, it is now impossible to calculate regional population projections without taking into account the considerable increase in commuting, sometimes over long distances.

These considerations indicate the significance which migration models might now have for regional population projections; but they also indicate what is to be required of the models if they are to be of any practical use. In the first place, the attractiveness factors should faithfully portray the real factors, economic as well as other, that regulate internal migration. This requirement is not met, in the author's opinion, by applying an extensive set of variables which are derived from automatic correlation procedures without any attempt at theoretically explaining the correlations. This is the method followed by Tarver.[45] More faith is to be put in Somermeijer's method of deriving variables from, at least, *ad hoc* theoretical considerations.[46] It is clear, however, that much will still have to be done in the way of developing theoretically sound propositions as regards the factors regulating internal migration. Studies

---

[43] Steigenga, *Het vraagstuk der regionale bevolkingsprognose*, p. 82.
[44] Isard, *op. cit.*, p. 582 ff.
[45] Tarver, *op. cit.*
[46] Somermeijer, *op. cit.*, p. 123 ff.

like the one recently published by Spiegelglas contribute to this aim.[47]

Besides, it should be reasonably probable that the influence of certain attractiveness factors will remain the same over a period of time, if migration models are required for projection purposes. This will be the less probable in the case of variables which, themselves, develop according to non-linear functions. In the Netherlands, for instance, the population occupied in agriculture has declined considerably during the past decade. Obviously, this decline will not continue in the same way for more than a comparatively short period: a limit will soon be reached. Therefore, this variable cannot be used in projection formulas of any description, even though, it might be a satisfactory attractiveness factor for describing past migration.

A last consideration is that attractiveness factors should be chosen in such a way as to make the calculation of a projection of these factors themselves possible. For, to apply migration models the independent variables, that is to say the attractiveness factors, will have to be known. Ideally the projections of the attractiveness factors should be independent of the population projection. For instance, a projection of Somermeijer's main attractiveness factor, degree of urbanization, would have to be based on the population projection itself. Therefore attempts to apply this factor in a migration projection model would entail complicated iteration processes.

SUMMARY

Our review of diverse migration models has led to the conclusion that, up till now, the most satisfactory model is the one developed by Somermeijer. In this model, the theoretical and mathematical advantages of a multiplicative formula based on

[47] Spiegelglas, Stephen: The Role of Industrial Development as a Factor Influencing Migration to and from Wisconsin Counties, 1940–1950. *Journal of Farm Economics*, 1961, 43: 128–137.

While the present article awaited publication a study containing a nationwide analysis of internal migration in the United States on lines not incomparable to Spiegelglas' method was published: BLANCO, C.: THE DETERMINANTS OF REGIONAL FACTOR MOBILITY, Rotterdam, Nederlandsche Economische Hoogeschool, 1962.

Zipf's P_1P_2/D hypothesis are maintained, while an efficient method has been developed to insert attractiveness factors into the formula.

Recent experience with distribution methods for regional population forecasts has indicated that, in principle at least, migration models might be of some use in calculating regional population projections. They would have to be used in conjunction with other methods. However, to make the use of migration models possible, the attractiveness factors should (a) be theoretically relevant, (b) show a connection with migration which may be considered relatively permanent under different circumstances, and (c) be such that they, themselves, can be made the subject of an independent projection.

Our general conclusion will have to be that migration models show some promise of eventually being applicable in connection with regional population forecasts, but that much further study will be necessary before this promise will come true.