

# Interprovincial Migration Flows and the Exponential Interaction Model

By LAURI HAUTAMÄKI

Institute of Geography at the University of Helsinki

## Problem of study, and study material

Efforts have been made to explain the migration movement with models of many types, but no perfect model has yet been evolved, and it is evident that none will be. There has been little methodical development of this kind in Finland.

As a rule the migration movement has been studied unilaterally: movement to one region from other regions has been examined. The reason for this has evidently been a lack of necessary data: it was comparatively easy to list the migration movement of a given locality with sufficient accuracy, but not the movements between all localities.

When migration is studied multilaterally, from every region to every other, the great problem is to determine the direction of flows of migration: there is more movement from Finnish Lapland to Uusimaa, for instance, than from Uusimaa to Lapland.

The purpose of this article is to evolve a model to explain with maximum efficiency the flows of migration between administrative provinces in Finland in 1966. The objective is a model of adequate simplicity which is still capable of explaining the bulk of the migration movement. This article deals with the same problem, which was examined in my previous study (Hautamäki 1971). It was possible to continue the examination of this problem, because new material was at hand.

Administrative provinces were chosen as objects of study because appropriate material is available, and they are sufficiently large for chance not to be an important factor. Flows of migration between them were listed for 1966. Distances were calculated along main roads between main centres. An exception was Ahvenanmaa, (Åland), to which no main road leads. The sea journey to Turku was converted to a main road journey by multiplying the time in hours spent at sea by 60, which is the average speed of a car in km per hour. The average income per person was available in ready form. Populations and flows of migration between administrative provinces were obtained from official statistics. The proportion of Swedish-speaking population and the level of unemployment were listed. The latter was ascertained at the Labour Ministry.

### Geographical migration theories

Geographers have paid attention to the interaction theory, which can be applied to migration, too. There are two kinds of interaction models. One is based heavily on physical analogies and the other is empirical, aiming to generalize the detailed findings in a mathematical form (Haggett 1965). The results of Ravenstein (1885, 1889) have been the starting point for later migration studies, which have been made in abundance (Isard et al. 1960). In recent times such studies are few, perhaps, because it is difficult to find anything theoretically new.

The gravity model is one of the best contributions of physical science to the geographers. Although this model may have been implied in the writings of Carey (1858—1859), Ravenstein (1889), Young (1924) and Reilly (1929), the largest impetus in its development stems from the works of Stewart and Zipf, who simultaneously worked on the model from independent angles (Isard et al. 1960).

Dodd (1950) has defined the elementary interaction model as follows:

$$(1) \quad I = \frac{K T P_a P_b A_a A_b}{D}$$

*I* stands for interaction. *K* is constant, *T* is a time element, *P<sub>a</sub>* and *P<sub>b</sub>* are the populations of groups *a* and *b*, and *A<sub>a</sub>* and *A<sub>b</sub>* are the specific indices of per capita activity, an example can be average income per person. *D* is a space dimension. The formula can be applied to many different cases, for example various interaction forms.

Absorption theory is another analogy, which has been used in migration study. Johnsson (1952) suggested that migration from the centre might be likened to the emission of a ray of light. Light and, in the same way, migrating population are gradually absorbed by the medium in proportion to flow per unit of distance. However, there are many difficulties in applying this idea.

Elementary interaction models are simple and effective guides in describing movement, but in practice they meet many difficulties (Haggett 1965). There are many problems in defining each component: mass and distance, and the relations between mass and distance.

Mass has conventionally been equated with population size. The use of population has many difficulties (Isard et al. 1960). The population must be weighted in some way, for example by its mean per capita income. Other indices are commodity output (Warntz 1959) and retail sales (Dunn 1956). Distance can be measured also in many ways. The simplest way is that of the straight-line distance between the two points. However, distance is a more complex function and straight-line distance is very seldom

useful. In many studies, time or transportation costs are a more appropriate measure (Harris 1954). The migrants' view of distance is not a simple geographical one (Huff 1960). Therefore Hägerstrand (1957) has used the logarithmic transformation of distance. Olsson (1965) has examined the idea of distance very effectively.

The relations between mass and distance are also the difficult problem. Isard et al (1960) has shown that it is possible to fit a quadratic rather than a linear function to a set of interaction data.

Dodd (1950) and Carrothers (1956) suggest a refinement of the gravity model which uses different exponents for mass and distance. Some authors have made attempts to solve this. Many authors have suggested that the gravity model must be modified by using different exponents for mass and distance (Stewart 1942, 1948, Dodd 1950, Ikle 1954, Andersson 1955, Carrothers 1956). Exponential models can be solved by taking logarithms of both sides of the formula and then using linear regression analysis. The unknowns may then be determined. Until further systematic research is pursued on the problem of exponents it will remain unresolved (Isard et al. 1960).

Isard et al (1960) and Kariel (1963) has emphasized regional complementarity in modifying the gravity theory. Isard has attempted to build a simple model

$$(2) \quad M_{ij} = (P_i/d_{ij}) f(Z_j)$$

where  $M_{ij}$  is migrants from area  $i$  to area  $j$ ,  $P_i$  is the population of area  $i$ ,  $d_{ij}$  is the distance between regions  $i$  and  $j$ .  $f(Z_j)$  is a function of  $Z_j$ ,  $Z_j$  measures the attractive force of destination  $j$ . Stouffer (1940) has developed the intervening opportunity model:

$$(3) \quad M_{ij} = (N_j/N_{ij}) \cdot k.$$

$N_j$  is the number of opportunities at location  $j$ ,  $N_{ij}$  is the number of opportunities in the intervening area between  $i$  and  $j$ ,  $k$  is constant. This model has been verified by empirical studies (Isbel 1944 and Folger 1953), and it gave a better prediction than the simple gravity model.

Hägerstrand (1957) has contributed the migration theory with the following model

$$(4) \quad M_{ij} = (V_j I_j)/P_j \cdot k$$

$V_j$  are the vacancies at  $j$ ,  $I_j$  is the information level about these vacancies existing at the source  $i$ ,  $P_j$  is the population of destination  $j$ . The volume of migration is related to the vacancy density of the destination and the level of information about it. Distance is compensated by the level of information.

There are many possibilities of describing migration in a mathematical form. However, they do not exactly fit the problem discussed earlier. The exponential applications of the gravity model may be the most profitable, because the material for this kind of model can be collected and the theoretical background is adequate.

## Results of previous study

### *Absolute migration flows*

In the early phase of study (Hautamäki 1971) various models were tested which are not given here for reasons of space. The following proved best:

$$(5) \quad M_{ij} = \frac{P_i^x P_j^y A_i^z A_j^q}{D_{ij}^w}$$

$M_{ij}$	= number of migrants from region i to region j
x, y, z, q and w	= unknown exponents
$P_i$ and $P_j$	= populations of regions i and j
$A_i$ and $A_j$	= average income per person in regions i and j
$D_{ij}$	= distance (km) between regions i and j

The above model is useful because each factor has a different exponent. For the solution of unknown exponent regression analysis was used, and the formula was converted logarithmically so that a linear solution could be used. Correlations between variables (Pearson's r) were calculated. Collinearity did not noticeably occur. The greatest correlation between explanatory variables was considerably below 0.40. Collinearity is not a serious drawback, because the main objective was a model which would, as a whole, explain as much as possible. Dependencies were not tested for linearity, but at least according to Rikkinen (1970) the dependence between distance and standardized flows of migration appears to be almost linear.

A solution was reached by so-called stepwise regression analysis, in which the variable with greatest explanatory power is first chosen as explainer, then the other variable with maximum explanatory power explaining the residual is chosen, and so on. The results of regression analysis are here shown in Table 1.

Migration between administrative provinces can thus be explained by the following formula very well.

$$(6) \quad M_{ij} = e^{-18} \cdot \frac{P_i^{1.3} \cdot P_j^{1.2}}{D_{ij}^{1.0}} \cdot \frac{A_j^{0.8}}{A_i^{1.2}} \quad (\text{For explanations see formula 5})$$

Table 1. Results of stepwise regression analysis (5 steps).

	R <sup>2</sup>		b	$\beta$	t	expl. loss
1. ....	50	Dij	-2.3	-0.71	-11.4	50
2. ....	67	Dij	-1.8	-0.55	10.0	26
		Pi	1.0	0.43	8.0	16
3. ....	90	Dij	-1.0	-0.30	-9.3	6
		Pi	1.3	0.57	18.8	27
		Pj	1.2	0.54	17.7	24
4. ....	91	Dij	-1.0	-0.29	-9.2	6
		Pi	1.3	0.59	19.9	28
		Pj	1.2	0.55	18.6	24
		Ai	-1.3	-0.10	-3.6	1
5. ....	92	Dij	-1.0	-0.30	-9.6	6
		Pi	1.3	0.59	20.0	27
		Pj	1.2	0.53	18.1	22
		Ai	-1.2	-0.09	-3.4	1
		Aj	0.8	0.06	2.2	0

R<sup>2</sup> = explanation per cent

b = regression coefficient (exponent value)

$\beta$  = standardized regression coefficient

t = t-test value (d.f. 131)

expl. loss = percentage by which explanatory power of model is reduced if variable concerned is omitted.

Other explanations see formula 5.

Although the migration is affected at this level by many other factors of which the above formula takes no account, their proportion is surprisingly small, some 8 %. This also includes the element of chance. Language difference, emigration, population growth differences and contacts between regions naturally affect flows of migration in various ways.

In general it seems that at this level the migration movement is comparatively easy to forecast. Olsson (1965) concluded that the element of chance in the movement is very small.

#### Standardized migration flows

Flows of migration were also examined in standardized form, i.e. by dividing the number of migrators from one region to another by the population of the region of departure. In this way the divergence arising from the difference in the size of population is eliminated. The same stepwise regression analysis was used as in the examination of absolute flows of migration.

Missing from Table 2 are regression coefficients (b), because flows of migration are related to the population of the region of departure and are not comparable with absolute flows; standardized coefficients ( $\beta$ ) are

Table 2. Dependence of standardized migration flows on different factors.

	R <sup>2</sup>		$\beta$	t	%-loss
1. ....	36	Aj	0.60	8.5	36
2. ....	60	Aj	0.62	11.1	38
		D	-0.49	-8.9	24
3. ....	67	Aj	0.56	10.9	31
		D	-0.38	-7.1	13
		Pj	0.29	5.3	7
4. ....	69	Aj	0.55	10.9	29
		D	-0.38	-7.1	13
		Pj	0.30	5.4	7
		Ai	-0.11	-2.2	1

R<sup>2</sup> = explanation per cent

b = regression coefficient (exponent value)

$\beta$  = standardized regression coefficient

t = t-test value (d.f. 131)

expl. loss = percentage by which explanatory power of model is reduced if variable concerned is omitted.

Other explanations see formula 5.

comparable. However, standardized regression coefficients of standardized migration flows differ from those of absolute flows (Table 1).

Standardized migration flows can be less adequately explained than absolute flows. In standardized flows the population of the departure region is not taken as an explanatory variable, because the migrators were divided by it. The unknown exponents obtained by regression analysis are not examined separately.

It was noted that the percentage of Swedish speakers and the proportion of unemployment may affect the migration flows. The material concerning these variables was collected and regression analysis was able to continue.

### Results of further study

The correlation matrix between all explanatory variables included, and between both variables to be explained, is shown in Table 3. The correlation between the new variables adopted (proportion of Swedish-speaking population and unemployment percentage), and the correlation with former explanatory variables is very large, in many cases over 0.500. This hinders examination of the significance of individual variables. Correlations between explanatory variables in Tables 1 and 2 remain below 0.400.

The correlation between absolute and standardized migration flows is extremely large but still less than one: it is 0.88. The difference arises

Table 3. Correlation matrix of all variables.

	1	2	3	4	5	6	7	8	9	10	11
$d_{ij}$	1.000										
$P_i$	-0.365	1.000									
$S_i$	0.366	-0.722	1.000								
$A_i$	0.029	0.165	0.397	1.000							
$P_j$	-0.361	-0.091	0.066	-0.015	1.000						
$A_j$	0.036	-0.015	-0.036	-0.091	0.165	1.000					
$S_j$	0.246	0.025	-0.077	-0.055	-0.280	0.610	1.000				
$U_j$	0.054	-0.002	0.053	0.049	0.019	-0.537	-0.800	1.000			
$U_i$	0.058	0.019	-0.586	-0.537	-0.002	0.049	0.073	-0.091	1.000		
$M_{ij}$	-0.708	0.636	-0.570	-0.014	0.598	0.136	-0.202	0.009	0.121	1.000	
${}_sM_{ij}$	-0.698	0.256	-0.387	-0.176	0.732	0.317	-0.124	-0.119	0.216	0.883	1.000

$d_{ij}$  = Distance between regions  $i$  and  $j$

$P_i$  and  $P_j$  = Population of departure region  $i$  and arrival region  $j$

$S_i$  and  $S_j$  = Proportion of Swedish-speaking population of total population in regions  $i$  and  $j$

$A_i$  and  $A_j$  = Average income level in regions  $i$  and  $j$

$U_i$  and  $U_j$  = Unemployment (%) in regions  $i$  and  $j$

$M_{ij}$  = Migrants from region  $i$  to region  $j$

${}_sM_{ij}$  = Standardized migration flows from region  $i$  to region  $j$

because in standardized migration flows the various population figures of the departure region are eliminated. For this reason the population of the departure region does not correlate with the liveliness of the migration movement as strongly as in absolute migration flows (Table 3). The Swedish-speaking percentage also correlates less with standardized than with absolute migration flows. On the other hand, the population and income level of the arrival region, and unemployment in the arrival and departure regions correlate considerably more with standardized than with absolute migration flows. The distance between regions ( $d_{ij}$ ) correlates about equally with standardized and absolute migration flows.

Distance between regions, population of arrival and departure county and Swedish-speaking percentage of the departure region correlate most with absolute migration flows. Populations correlate positively, which was to be expected from previous studies. On the other hand, distance and Swedish-speaking percentage correlate negatively. The negative correlation of the latter may be explained on the one hand by the small population of Swedish-speaking regions (Ahvenanmaa), and on the other by a less vigorous migration movement and a greater willingness to emigrate among Swedish-speaking people.

Population of the arrival region correlates most with standardized migration flows; it is followed by distance. Other correlations are distinctly smaller, below 0.400.

Table 4 gives the results of regression analysis for both standardized and absolute migration flows. For the former the explanation percentage was 89, and for the latter 93. Especially notable was the greater explanation percentage of standardized migration flows compared with Table 2. This is probably due mainly to inclusion of the part played by unemployment and Swedish-speaking people, which may include the effect on the population in the departure region.

If the regression coefficients for Table 1 are compared with Table 4, it will be noted that they are approximately the same for the best explanatory variables ( $D_{ij}$ ,  $P_i$  and  $P_j$ ), but quite different for the weaker explanatory variables. An important reason for this difference is the collinearity of the explanatory variables, which has been mentioned earlier. There are thus no ground for drawing more far-reaching conclusions from individual regression coefficients in this connection. It can only be noted that the regression coefficients of the most significant variables are probably reliable in the regression analyses shown in Table 1, and that they have been used in the construction of a model for the migration movement (Formula 6).

Figures 1 and 2 enable us to examine how far the regression models of Table 4 tally with actual migration flows. A natural logarithm has been taken of the figures depicting migration flows.



Table 4. Dependence of migration flows on all the measured independent variables.

Source of variation	Sum of squares		Degree of freedom		Mean square	
	A	B	A	B	A	B
regression	502.96	161.22	9	9	55.88	17.91
residual	38.01	20.60	122	122	0.31	0.17
total	540.98	181.81	131	131		

F-value A = 179.34, B = 106.11  
 Multiple correlation squared A = 0.93, B = 0.89

Independent variable	Regression coefficient		Its standard deviation		Standardized coefficient		Student t		Explanation	
	A	B	A	B	A	B	A	B	A	B
d <sub>ij</sub>	-1.14	-	0.12	0.08	-0.34	-0.48	-9.29	-10.48	0.049	0.102
P <sub>i</sub>	1.33	-	0.15	-	0.60	-	8.71	-	0.043	-
S <sub>i</sub>	0.00	-	0.00	0.00	0.06	-0.01	0.68	-0.21	0.000	0.000
A <sub>i</sub>	-0.45	-	0.52	0.47	-0.03	-0.07	-0.86	-1.28	0.000	0.001
P <sub>j</sub>	1.11	-	0.06	0.04	0.50	0.53	16.55	13.63	0.157	0.172
A <sub>j</sub>	1.50	-	0.46	0.34	0.11	0.26	3.23	6.04	0.006	0.033
S <sub>j</sub>	0.02	-	0.07	0.05	0.02	0.03	0.34	0.41	0.000	0.000
U <sub>j</sub>	0.73	-	0.34	0.24	0.10	0.08	2.16	1.34	0.002	0.001
U <sub>i</sub>	1.05	-	0.35	0.23	0.15	0.21	2.99	3.62	0.005	0.012

Constant term A = -28.27, B = -11.14

Other explanations see table 3.

Figure 1.  
 Theoretical and empirical migration flows between administrative provinces in Finland in 1966. Theoretical flows are calculated according to the formula shown in Table 4 A. Columns = number of in-migrants (natural logarithms) to each region. Figures 1—12 mean administrative provinces. In-migrants from each region are presented in the same order as clock hours 1—12, white column = theoretical migration flows, ruled column = empirical migration flows.

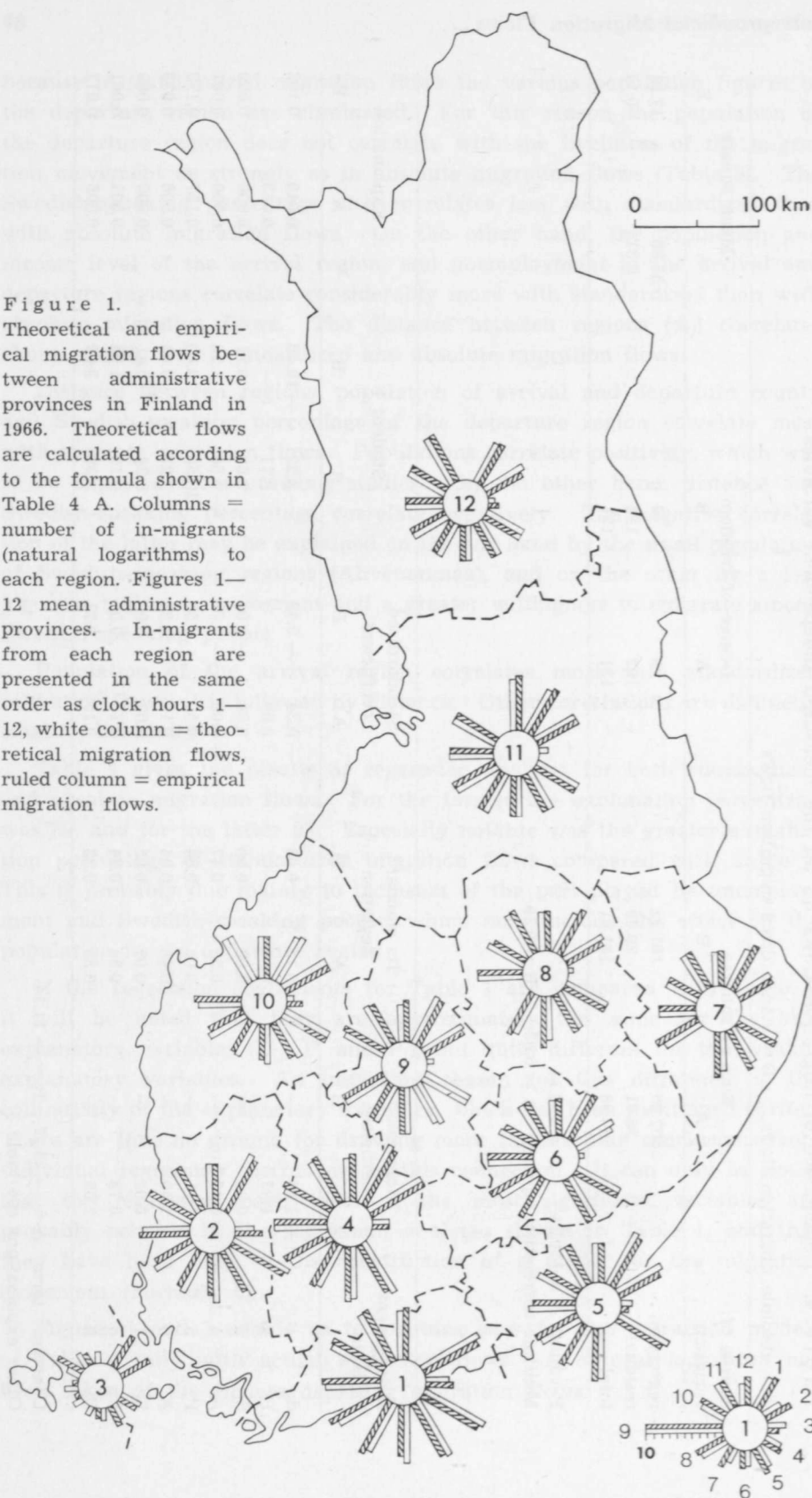
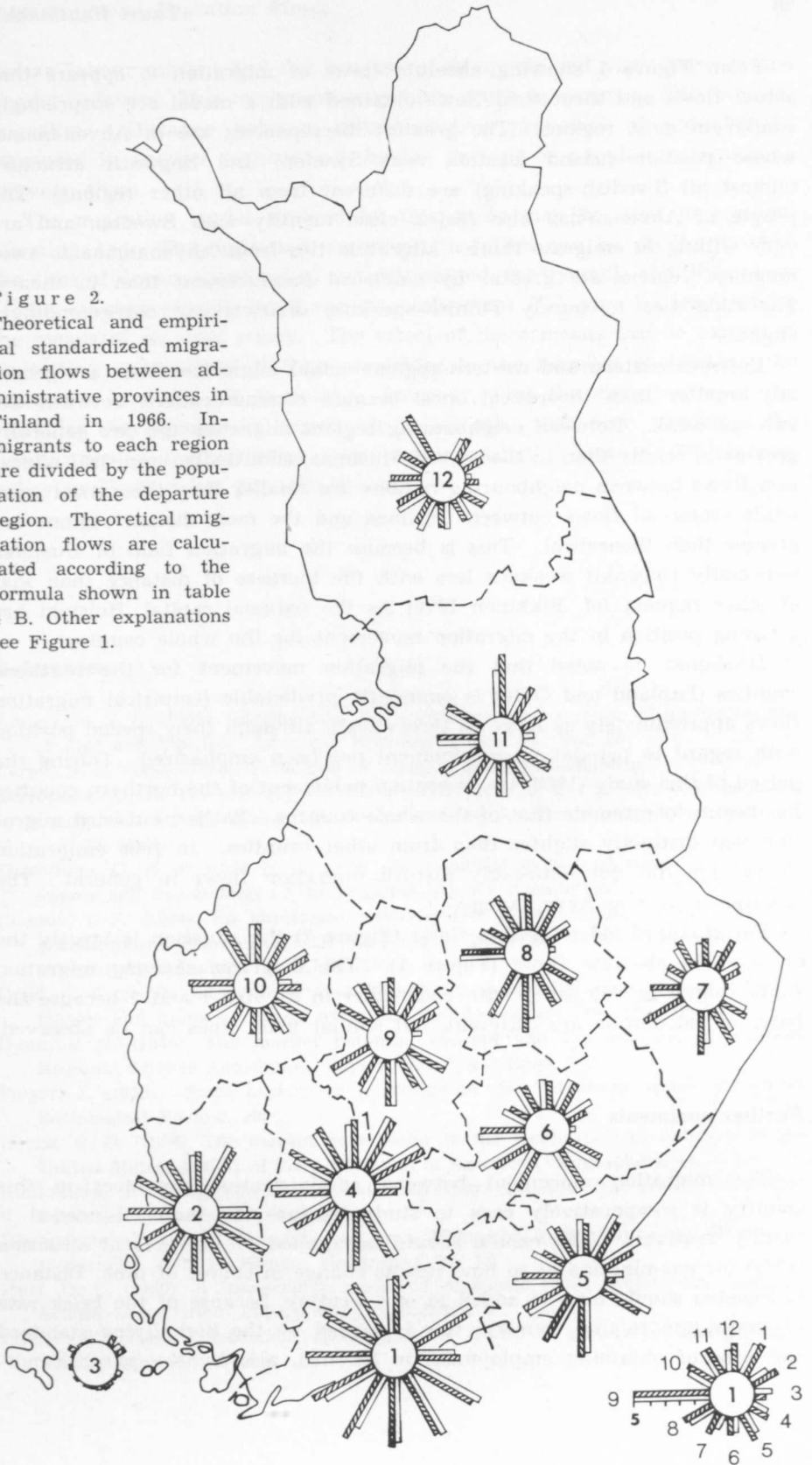


Figure 2.  
 Theoretical and empirical standardized migration flows between administrative provinces in Finland in 1966. In-migrants to each region are divided by the population of the departure region. Theoretical migration flows are calculated according to the formula shown in table 4 B. Other explanations see Figure 1.



From Figure 1 showing absolute flows of migration it appears that actual flows and theoretical flows obtained with a model are surprisingly similar in most regions. The greatest discrepancies are in Ahvenanmaa, whose position (island location near Sweden) and linguistic structure (almost all Swedish-speaking) are different from all other regions. The people of Ahvenanmaa also feel a close identity with Sweden, and are very willing to emigrate there. Migration ties from Ahvenanmaa to two-language regions are greater by empirical measurement than in theory. Migration ties to purely Finnish-speaking districts are correspondingly slight.

Between eastern and western regions actual migration flows are generally smaller than theoretical ones, because communications between the two are weak. Between neighbouring regions migration ties are generally greater in reality than in theory. In Uusimaa, admittedly, empirical migration flows between neighbouring regions are smaller than theoretical ones, while empirical flows between Uusimaa and the most distant regions are greater than theoretical. This is because the migration field of Uusimaa (especially Helsinki) weakens less with the increase of distance than that of other regions. (cf. Rikkinen 1970) As the national capital, Helsinki has a strong position in the migration movement for the whole country.

It should be noted that the migration movement for the northern counties (Lapland and Oulu) is eminently predictable (empirical migration flows approximately as large as theoretical), although their special position with regard to population development has been emphasized. During the period of this study (1966) the migration movement of the northern counties has begun to resemble that of the whole country. Earlier, outward migration was distinctly slighter than from other counties. In 1966 emigration to Sweden did not noticeably disturb migration flows in general. The situation may now have changed.

For standardized migration flows (Figure 2) the situation is largely the same as in absolute flows (Figure 1). The diagrams showing migration flows cannot be compared with each other in Figures 1 and 2 because the bases of calculation are different, but similar main lines can be observed.

### Further comments

The migration movement between administrative provinces in this country is comparatively easy to study because the material needed is readily available. The results should be applied to the present situation (1971) for examination as to how results change in course of time. Distance to Sweden should also be added to our scrutiny, because of the brisk rate of emigration to that country: this is caused by the high living standard and ease of obtaining employment in Sweden, also by the joint Scandi-

navian labour market and the non-use of passports and visas between the Nordic countries. This movement may have much in common with internal migration. Distance between administrative provinces should also be measured in another manner, e.g. by the amount of population remaining between them.

The migration model evolved by this study may be used in regional planning, too. The migration movement in future can be predicted when the aims of development policy are determined, for example the differences of the average income level and unemployment (%) may be lessened by the means of regional policy. The effect of these means can be examined by setting new values of these variables to the formula evolved by the regression analysis. It is notable, however, that the significance of the variables may change in future. It is not safe, that even nowadays the significance of the variables examined is realistic, because the regional policy in Finland has been ineffective.

## References

- Andersson, T. R. (1955). Intermetropolitan migration: a comparison of the hypotheses of Zipf and Stouffer. *American Sociological Review*, Vol. 20.
- Carey, H. C. (1958—1859). *Principles of social science*. Philadelphia.
- Carrothers, G. A. P. (1956 A). An historical review of the gravity and potential concepts of human interaction, *Journal of the American Institute of Planners*, Vol. 22, No 2.
- Carrothers, G. A. P. (1956 B). Gravity and potential models of spatial interaction, *Papers and Proceedings of the Regional Science Association*, Vol. 2.
- Claeson, C-F. (1964). En korologisk publikanalyt: Framställning av demografiska gravitationsmodeller med tillämpning vid omlandsbestämning på koordinatkarta, *Geografiska Annaler*, Vol. XLVI, No 4.
- Dodd, S. C. (1950). The interactance hypothesis. A gravity model fitting physical masses and human groups., *American Sociological Review*, Vol. 15.
- Dunn, E. S. (1956). The market potential concept and the analysis of location. *Regional Science Association. Papers and Proceedings*, 2.
- Folger, J. (1953). Some aspects of migration in the Tennessee valley. *American Sociological Review*, 18.
- Harris, C. D. (1954). The market as a factor in the localization of industry in the United States. *Annals of the Association of American Geographers*, 44.
- Hautamäki, L. (1971). Interprovincial migration and the non-linear interaction hypothesis, *Fennia* 102, Helsinki.
- Hägerstrand, T. (1957). Migration and area. Migration in Sweden, *Lund Studies in Geography. Ser. B. Human Geography* No 13.
- Huff, D. L. (1960). A topographic model of consumer space preferences. *Regional Science Association, Papers and Proceedings*, 6.
- Ikle, F. C. (1954). Sociological relationship of traffic to population and distance, *Traffic quarterly*, Vol. 8.

- Isard, W., D. F. Bramhall, G. A. P. Carrothers, J. H. Cumberland, L. N. Moses, D. O. Price and E. W. Schooler (1960). *Methods of regional analysis: an introduction to regional science*. New York.
- Isbell, E. C. (1944). Internal migration in Sweden and intervening opportunities. *American Sociological Review*, 9.
- Johnson, O. H. (1952). En stads flyttnings- och födelseortsfält. *Svensk Geografiska Årsbok*, 28. 115—122.
- Kariel, H. G. (1963). Selected factors areally associated with population growth due to net migration. *Annals of the Association of the American Geographers*, 53.
- Mackey, J. R. (1968). The interactance hypothesis and boundaries in Canada: A preliminary study, *Spatial Analysis*. Ed. by Berry, B. J. and Marble, D. F., 122—129.
- Olsson, G. (1965). Distance and human interaction. A migration study. *Geografiska Annaler*, 47 B.
- Ravenstein, E. G. (1885), (1889). The laws of migration. *Journal of the Royal Statistical Society*, 48, 52.
- Reilly, W. J. (1929). Methods for the study of retail relationships. *University of Texas, Bulletin*, 2944.
- Rikkinen, K. (1970). A circulation model of interprovincial migration in Finland. *Fennia* 99: 8.
- Stewart, J. Q. (1942). A measure of the influence of a population at a distance, *Sociometry*, Vol. 5.
- Stewart, J. Q. (1948). Demographic gravitation: Evidence and applications, *Sociometry*, Vol. 11.
- Stouffer, S. A. (1940). Intervening opportunities: A theory relating mobility and distance, *American Sociological Review*, Vol. 5.
- Warntz, W. (1959). *Toward a geography of price*. Philadelphia.
- Zipf, G. (1946). The  $P_1P_2/D$  hypothesis on the intercity movement of persons, *American Sociological Review*, Vol. 11.