Pertanika 2(2), 95–100 (1979)

An Econometric Analysis of Interstate Migration in Peninsular Malaysia

AHMAD MAHDZAN AYOB

Faculty of Resource Economics and Agribusiness, Universiti Pertanian Malaysia, Serdang, Selangor.

Key-words: econometric model; inter-state migration; regression; Peninsular Malaysia.

RINGKASAN

Suatu model ekonometrik telah digunakan untuk menerangkan migrasi antara negeri di Semenanjung Malaysia dengan menggunakan data Banci Penduduk 1957. Penentu-penentu migrasi antara negeri ialah jumlah pekerja di bidang industri kedua dan ketiga pada destinasi, bilangan penduduk pada tempat asal, jarak jauh antara destinasi dengan tempat asal dan satu dummy variable *yang menunjukkan samada negeri-negeri berkenaan terletak sebelah menyebelah dan dilalui oleh jalan kereta atau keretapi.*

SUMMARY

An econometric model is used to explain inter-state migration in Peninsular Malaysia using the 1957 population census data. The determinants of interstate migration are the size of employment in the secondary and tertiary industries at the destination, the size of the population at the origin, distance between the two points, and a dummy variable indicating whether states share a common boundary and traversed by a road or railway.

INTRODUCTION

Economists generally agree that inter-regional migration is a useful mechanism by which inter-regional differences in factor endowments are removed for the benefit of all regions. Inter-regional migration is a substitute for the flow of goods which may not occur because of certain artificial restrictions in the economy. The skills of workers may therefore be transferred between regions thus enabling certain regions to produce goods which, hitherto, were not available via the process of normal trade.

Concern has, however, been expressed by regional political leaders of the dangers of uncontrolled out-migration from certain rural areas. Such an exodus into the urban areas is seen as a threat to the rural agricultural economy which is facing a labour shortage.

The objective of this paper is primarily to identify the determinants of inter-regional migration in Peninsular Malaysia. Most people agree from intuition that the rural-urban flow of population results from a push and a pull factor. The differentials in the quality of life between the two sectors provide the push and pull at the same time. But this is an over-simplication of reality, and some in-depth study is required to explain the phenomenon. Variables that are likely to influence migration must be determined, quantified and analyzed through rigorous statistical techniques.

LITERATURE REVIEW

"The theory and empirical analysis of inter-regional migration is a field which economists and social scientists have devoted lavish attention, though, unfortunately, without producing clear-cut unambiguous results." (Richardson, 1973, *p.* 88).

Richardson (1969, p. 295) has summarised neoclassical migration theory which predicts that in a situation characterised by inter-regional differentials in real wages, labour will migrate from the low-wage to the high-wage regions until real wages are equalised. This conclusion rests upon several critical assumptions: a comparative static framework; homogeneous labour; constant

Key to author's name: Ayob, A. M.

returns to scale; zero migration costs, perfectly competitive labour market; workers' move in response to wage differentials alone. Because of these unrealistic assumptions, later researchers made various modifications to the neoclassical model.

Raimon (1962) showed that inter-state migration was strongly correlated with *per capita* income differentials and with differences in job vacancies.

Blanco (1963) stressed the role of unemployment differentials and showed that 85 percent of the variation in the rate of migration between U.S.A. states, 1950 - 57, could be explained by two factors - namely, changes in the level of unemployment and changes in the number of military personnel in each state.

Lowry (1966) modified the Blanco model and then combined it with a relative wage and gravity-type formulation:

$$M_{ij} = F\left(\frac{U_i}{U_j}, \frac{W_j}{W_i}, \frac{L_iL_j}{D_{ij}}\right) \quad (1)$$

where M_{ij} is migration from region i to region j, U is the unemployment expressed as a percentage of the civilian labour force; L is the number of persons in non-agricultural labour force, W is the manufacturing wage, D is the air-line distance.

The model is based on very reasonable assumptions. For example, it is assumed that people are motivated to migrate from low-wage to high-wage and from high unemployment to low unemployment areas. Equilibrium is reached when $W_i = W_j$ and $U_i = U_j$. The model explained about 68 percent of the observed migration flows between 90 Standard Metropolitan Statistical Areas (SMSAs) for the 1955 - 60 period.

Zipf (1946) formulated a crude gravity model to analyze inter-urban migration. The model is often referred to as the P_iP_j/D hypothesis where P_i is the population at point of origin and P_j the population at the destination and D is the distance between the two points.

Stouffer (1960) introduced the concept of intervening opportunities in an attempt to explain inter-regional migration.

Yap (1977) made an excellent review of the internal migration literature pertaining to developing countries. The survey focused on "information relevant to policy concerns about the possibility and desirability of regulating, or accidentally affecting, movement to the cities." Yap discussed two kinds of empirical work, namely studies on the determinants of migration, and the "empirical material on the migrant experience in urban locations." The majority of the studies reviewed revealed that "while migration is controllable within limits, conditions in cities exert a strong pull and that, in spite of large flows, migrants are being assimilated into the urban economy and are contributing to urban output growth."

A Previous Related Study in Malaysia

Soon (1975) analyzed trends and patterns of internal migration in Peninsular Malaysia. From the various tabular analyses he found that rural-urban migration as a whole was insignificant. The pattern of migration as a whole which emerged was one of urban to urban or rural to rural. Rural development was cited as a causal factor in retaining people in the rural sector.

THE MODEL

The model used in the present paper to analyse population movement between regions in Malaysia is an adoption of and a marriage between Zipf's model (1946) and Stouffer's (1960). Zipf hypothesized that the total flow of population between two points i and j with respective population P_i and P_j and separated by the shortest transportation distance D_{ij} will be proportional to the quantity P_iP_j/D_{ij}:

$$M_{ij} = C\frac{P_iP_j}{D_{ij}} \quad (2)$$

where M_{ij} is the total flow of population from point i to point j and C is a constant of proportionality.

Stouffer on the other hand argued that there was no necessary relationship between mobility and distance but suggested that the availability of opportunities at any given distance is the critical factor.

For our purpose, we hypothesize that distance between original point i and destination j is important in determining the volume of population flow between i and j. At the same time we also agree with Stouffer who maintains that the availability of opportunities at any given distance is the "critical factor" in determining the size of the flow.

Given the soundness of both these authors, it is hypothesized that the size of population migrating from point i to point j, M_{ij}, is dependent on the size of population resident at i, P_i, the "size" of employment opportunities, E_j, available at j and the distance D_{ij}, between the two points. Mathematically this may be expressed as follows:

$$M_{ij} = f(P_i, E_j, D_{ij}) \qquad (3)$$

The theoretical restrictions imposed upon the model are as follows:-

$$\frac{\partial M_{ij}}{\partial P_i} > 0 \qquad (4)$$

$$\frac{\partial M_{ij}}{\partial E_j} > 0 \qquad (5)$$

$$\frac{\partial M_{ij}}{\partial D_{ij}} < 0 \qquad (6)$$

Whether these restrictions are satisfied remains to be determined by the data. Inequality (4) indicates, *ceteris paribus,* that the bigger the population at the source point, the bigger is the number migrating. The variable P_i therefore is a measure of the population pressure felt at i. Inequality (5) indicates, *ceteris paribus,* that the bigger is the employment opportunity at j, the larger is the flow towards j. Inequality (6) indicates that the migration flow is inversely related to the distance between the two points.

The model embodied in equation (3) is merely a general economic model which does not lend itself to statistical estimation. In order to statistically establish the causal relationship between the dependent variable and the explanatory variables in equation (3) several mathematical functional forms are hypothesized. This procedure is legitimate because theory seldom specifies the functional forms of a model and therefore allows the researcher some flexibility in the choice of the functional form. In the absence of a priori knowledge of the behaviour pattern of the dependent variable the following alternative forms are chosen:

$$\text{(I)} \quad M_{ij} = \beta_0 + \beta_1 P_i + \beta_2 E_j + \beta_3 D_{ij} + \epsilon_{ij}$$

$$\text{(II)} \quad M_{ij} = \beta_0 + \beta_1 P_i + \beta_2 E_j + \beta_3 D_{ij} + \beta_4 D_{ij}^2 + \mu_{ij}$$

$$\text{(III)} \quad M_{ij} = \beta_0 + \beta_1 P_i + \beta_2 E_j + \beta_3 D_{ij} + \beta_4 D_{ij}^2 + \beta_5 (DSQP) + \beta_6 (DE_{ij}) + \beta_7 (PE_{ij}) + \omega_{ij}$$

$$\text{(IV)} \quad \ln M_{ij} = \beta_0 + \beta_1 \ln P_i + \beta_2 \ln E_j + \beta_3 \ln D_{ij} + \beta_4 Y_{ij} + \xi_{ij}$$

where the newly appearing variables are defined as follows:

$DSQP = (D_{ij}^2) \times P_i$)
$DE_{ij} = D_{ij} \times E_j$) Model III
$PE_{ij} = P_i \times E_j$)

Y_{ij} = dummy variable which equals 1 if the two points (regions) are adjacent *and* connected by road/railway, 0 otherwise. (Model IV).

$\beta_0, \beta_1, \ldots, \beta_7$ are parameters to be estimated.

ϵ_{ij}, μ_{ij}, ω_{ij} and ξ_{ij} are random disturbance terms assumed to be normally distributed with zero means and constant variances.

SOURCE OF DATA AND MEASUREMENT OF VARIABLES

Sources of Data – The major source of data for this study is the 1957 Population Census of the Federation of Malaya published by the Department of Statistics. A minor source is the mileage chart of Shell Oil Company of Malaysia.

The use of 1957 Population Census data will no doubt invite criticism as these data are now considered out-dated. Unfortunately the latest census conducted in 1970 in Peninsular Malaysia did not gather statistics comparable with the 1957 Population Census, the former being less adequate for the purpose of this paper. For instance, there was no information whatsoever on migration flows between the states, which the present study is concerned with. It must be reiterated that the major purpose of the paper is to test some well-known hypotheses regarding migration flows using Malaysian data. While the use of 1957 data may be objected on the grounds of "obsolescence", its recourse may still throw some light on the aggregative behaviour of Malaysians at a certain point in time. Since the basic needs of the Malaysian family – employment and material comfort – have not changed

TABLE 1

Regression results of population movement models (Malaysia 1957)

Independent Variables	Regression Coefficients for Model			
	I	II	III	IV†
Intercept	4,654* (2.774)‡	10,737* (5.330)	8,242* (3.383)	−4.096* (−3.719)
Employment$_j$	.043* (4.870)	.045* (5.540)	.050* (2.520)	.851* (13.733)
Population$_i$	6.207* (4.173)	6.580* (4.830)	2.930†† (1.160)	1.099* (13.859)
Distance$_{ij}$	−27.807* (6.991)	−92.650* (6.44)	−69.960* (−5.270)	−.909* (−7.569)
DSQ*POP	–	–	−.000042** (−2.338)	–
DIST*EMPLOY	–	–	−.000230* (4.01)	–
POP*EMPLOY	–	–	.000085* (3.91)	–
DUMMY *(Y)*	–	–	–	1.063* (5.884)
(DISTANCE)^2	–	.124* (4.660)	.137* (5.580)	–
R-squared	.49	.57	.70	.876
F-statistic	33.35*	35.34*	34.17*	186.219*

† Model 4 was estimated in logs, except the dummy variable which takes 0,1 values.
* Significant at 0.01 level.
** Significant at 0.02 level.
†† Significant at 0.25 level only.
‡ Figures in parentheses are Student's t-statistic used to test the null hypothesis that the regression coefficient is not different from zero.

drastically since 1957, the use of "historical" data need not be as objectionable as it seems.

Measurement of Variables – The 1957 Population Census provides data on population movement among all states in Peninsular Malaysia. On the day of the Census, all people in each state were required to state their state of birth. Thus M_{ij} is the number of people from state i enumerated in state j.

The variable P_i stands for the size of population in state i on the date of the census.

Employment opportunities at the destination, designated as E_j, is measured by the size of population engaged in secondary and tertiary industries, combined. Because ordinarily people are more attracted to jobs in these two industries than in agriculture, the use of the size of the labour force in the two industries to represent E_j appears to be based upon reasonable assumptions. Therefore, *ceteris paribus,* the bigger the combined labour force in the manufacturing and service industries in region j, the greater is the "pull" that region exerts upon migration from region i.

The distance variable, D_{ij}, unless otherwise indicated, is measured in miles between state capitals via the shortest motor route. It is therefore assumed that the state capitals serve as the "centre of gravity" of the state – which is a reasonable assumption as the population of the state tends to be heavily concentrated at state capitals. The use of state capitals as starting and destination points thus collapses all regions into mere points in space.

Admittedly the use of state capitals as starting and destination points oversimplifies the picture somewhat. This is a familiar problem in studying the space economy using an econometric approach.

Although the use of state capitals as starting and destination points substantially reduces the number of points covered, it should be recognised that with n points in space, one could obtain $n(n-1)$ "flows". For example, with four points, one could get $4(4-1) = 12$ "flows". With eleven state capitals, we have $11 \times 10 = 110$ observations on the M_{ij}'s.

Model IV, however, tries to avoid this over-simplication of regarding regions as points. The variable Y takes into account of the possibility that state i and state j may share a common boundary *and* be linked to one another by a road or railway or both. Thus, if state i and state j share a common boundary but not linked by a road or railway, (e.g. Perak and Kelantan) the variable Y_{ij} takes the value 0 in the same way as if these two states did not at all share a common boundary. The D_{ij}'s for Perlis-Kelantan and Kedah-Kelantan were measured by the rail distance through Thailand rather than the road distance through Pahang.

With eleven states in Peninsular Malaysia, there will be $11(11-1) = 110$ observations for the estimation of the parameters of the above four models. There are thus ample degrees of freedom in the statistical estimation.

RESULTS

Table I below gives the results of regression analyses of the four models formulated earlier.

Equation I which employs three explanatory variables to explain gross migration, shows that these variables explain nearly 50 percent of the variation in the observed dependent variable. All the regression coefficients are significant at the 1 percent level and their signs are also consistent with the theoretical restrictions imposed earlier. The coefficient of multiple determination (R^2) of 0.49 is statistically significant at the 1 per cent level as indicated by the F-statistic.

The second equation (Model II) includes a distance squared term, besides the original explanatory variables of Model I. The D_{ij}^2 term may indicate that after a certain distance is travelled, increments of distance become less important. The impact of distance upon migration may be seen by taking the partial derivative of Model II with respect to the distance variable:

$$\frac{\partial M_{ij}}{\partial D_{ij}} = -92.65 + 0.248\, D_{ij}$$

The impact of distance, therefore, depends upon the distance already travelled. Where greater distances are travelled, further increments in distance do not affect migration very much. The D_{ij}^2 term is statistically significant at the one percent level. The D_{ij} term remains significant (1 percent) with the expected negative sign. The explanatory power of Model II is increased to 0.57 and this is significant at the one percent level of probability.

Model III is an attempt to improve the statistical fit whereby certain interaction terms are included. As seen in Table 1, all the interaction terms are significant although the P_i variable is significant only at the 25 percent level. The R^2 is further improved to $R^2 = 0.70$.

Model IV is estimated in natural logarithms. All regression coefficients in Model IV are significant at the one percent level, including the dummy variable. The four independent variables in Model IV together explain 87.6 percent of the variation in the dependent variable. Moreover the signs of the regression coefficients are all consistent with a priori expectation and economic theory. Model IV is shown below and is singled out as the "best" model in this study:

$$\ln M_{ij} = -4.096 + \underset{(13.859)}{1.099 \ln P_i} + \underset{(13.733)}{0.851 \ln E_j}$$

$$\underset{(-7.569)}{-0.909 \ln D_{ij}} + \underset{5.884)}{1.063\, Y_{ij}} + \xi_{ij}$$

$$(R^2 = 0.876)$$

(Figures in parentheses are t-statistics)

The regression coefficients in Model IV represent elasticities. The coefficient of $\ln P_i$ of 1.099 indicates that, *ceteris paribus*, an increase of 10 percent in the population in region i is associated with about 11 percent increase in the number of migrants moving from region i to region j.

With respect to the destination, an increase of 10 percent in employment opportunities in region j, *ceteris paribus*, is associated with an 8.5 percent increase in the number of people migrating from region i to region j.

The variable D_{ij} may be regarded as a proxy for financial costs of moving from i to j. The elasticity figure of -0.909 indicates that, *ceteris paribus*, a 10 percent increase in the distance between state capitals is associated with a 9 percent reduction in the number of people migrating between the two points.

Preliminary studies revealed very high erythrocyte concentration and counting by the standard method was impossible. A dilution method as described below was adopted. 0.02 ml. of blood (using a haemoglobin pipette) was transferred to a test-tube containing 0.78 ml of normal saline (1/40 dilution) and 0.5 ml of the mixture was transferred to a test-tube containing 0.5 ml of normal saline to give a 1/80 dilution. Using a red blood cell diluting pipette blood (1/80 dilution) was drawn to the 0.5 mark and Hayem's fluid drawn to the 101 mark. The total count was estimated by using a Improved Neubauer haemocytometer.

Total r.b.c. $=$ no of r.b.c. counted $\times$ 80 $\times$ 10,000

The accuracy of this modified method was tested. A sample of goat's blood was subjected to both the normal procedure and the modified method for r.b.c. count and found to be not significantly different ($t = 0.858$, d.f. $= 8$).

The packed cell volume (PCV) was determined by the micro-haematocrit centrifuge method and haemoglobin concentration (Hb) by the cyanmethaemoglobin method. Thin blood smears stained with Wright's stain were used for differential counts and for measurement of erythrocyte size. The erythrocytes were measured by means of a 10 $\times$ ocular micrometer and a 100 $\times$ oil immersion objective.

RESULTS AND DISCUSSION

The range and mean of haematological results are presented in Table 1 and comparative data on ox, goat and mouse-deer are summarised in Table 2.

Erythrocytes

The results indicate that Mouse-deer have higher erythrocyte counts ($153.9 \times 10^6/mm^3$) than ox ($7 \times 10^6/mm^3$) and goat ($13 \times 10^6/mm^3$). The erythrocyte diameter (2.3 μ) was smaller than ox (5.9μ) and goat (3.7μ). The mean corpuscular volume (MCV) was extremely small ($3.5 \mu m^3$). The haemoglobin (16.4 g/100 ml) and packed cell volume (55.1%) were higher than in ox and goat. The mean corpuscular haemoglobin (MCH) was extremely low (1.1 pg), but the mean haemoglobin concentration (MCHC) was slightly lower (29.8%) than ox (31%) and goat (32%).

The total erythrocyte count is the highest known of any known ungulate. The erythrocyte diameter (2.3 μ) measured from dry blood smear preparations was similar (2.1μ) to that reported by Gulliver (1870) but differed (1.5μ) from that of Dukes (1963), derived from tissue preparations. There is an evolutionary trend to reduce the size and to increase the number of erythrocytes for better transport of gases (Wintrobe, 1961). The erythrocytes of Mouse-deer therefore appear to present more evolved mechanisms for transport of gases. The extremely low MCV ($3.5 \mu m^3$) is the smallest of all the known ungulates. The PCV and Hb values were higher than the other ungulates. Wintrobe (1961) reported that the more active species had high haemoglobin values and elevated haemoglobin may be common in wild animals.

Dunaway and Lovell (1965) reported a relation between r.b.c. count, mean corpuscular volume and size of the mammals. Generally the number of erythrocytes per mm^3 of blood is inversely related to the body-weights of the species in each of the families. However, Jordan (1938) has suggested that in general erythrocyte count, size and shape are related to the metabolism of the animal. Wilson and Hoskins (1975) have suggested that high haemoglobin levels in red kangaroos *Megaleia rufa* may reflect their capacity for sustained muscular activity over open plains. The small size of the animal and high erythrocyte count may reflect a high metabolic rate while high haemoglobin values may reflect that the lesser Mouse-deer, which is a very agile animal, has the capacity for sustained muscular activity.

Leucocytes

The total leucocyte count ($8.1 \times 10^3/mm^3$) was similar to ox ($8.00 \times 10^3/mm^3$) but lower than goat ($9.00 \times 10^3/mm^3$). The mean differential counts were: neutrophils 22.2%, lymphocytes 70.4%, monocytes 5.6%, eosinophils 2.5% and basophils 0%. No basophils were recorded in this study series although they had been encountered on other occasions.

The leucocyte values were similar to those in ox and goat with the lymphocyte and monocyte tending to occur in higher percentages and neutrophil and eosinophil in lower percentages.

ACKNOWLEDGEMENTS

The technical assistance of Mr. Ong Yew Fai, Mr. Fakri and Miss Kartini is gratefully acknowledged. This work was supported by a research grant from the Faculty of Veterinary Medicine, University Pertanian Malaysia.

TABLE 1

Composite Haematological Data of 10 Mouse-deers

Observations	Range	Mean	SD	No of Samples
Erythrocytes ($\times 10^6/mm^3$)	123–216	153.9	22.1	20
Erythrocyte size (μ)	1.42–3.3	2.3	0.6	42
Hb (g/100 ml)	15.6–18.5	16.4	0.9	10
PCV (%)	48–67	55.1	5.5	10
Sedimentation (24 hours)	NO SEDIMENTATION			10
MCV (μm^3)	3.1–4.2	3.5	0.3	10
MCH (pg)	0.95–1.24	1.1	0.1	10
MCHC (%)	27.6–32.5	29.8	1.4	10
Leucocytes ($\times 10^3/mm^3$)	5.1–11.2	8.1	2.0	10
Neutrophils (%)	19–31	21.2	3.7	10
Lymphocytes (%)	62–74	70.4	3.9	10
Monocytes (%)	5–6.5	5.6	0.5	10
Eosinophils (%)	1.5–4	2.5	0.9	10
Basophils (%)	0	0	0	10

TABLE 2

Comparative haematological data of Ox, Goat and Mouse-deer

Observations	Ox*	Goat*	Mouse-deer
Erythrocytes ($\times 10^6/mm^3$)	7	13	153.9
Erythrocyte size (μ)	5.9	3.7	2.3
Hb g/100 ml	11	11	16.4
PCV (%)	34	28	55.1
Sedimentation (24 hours)	2.2–4	2–2.5	No sedimentation
MCV (μm^3)	52	27	3.5
MCH (pg)	14	6	1.1
MCHC %	31	32	29.8
Leucocytes ($\times 10^3/mm^3$)	8.0	9.0	8.1
Neutrophils %	28.5	36	21.2
Lymphocytes %	58	56	70.4
Monocytes %	4	2.5	5.6
Eosinophils %	9	5	2.5
Basophils %	0.5	0.5	0

* Values from Veterinary Clinical Pathology by E. H. Coles (1974).

REFERENCES

ANONYMOUS (1971): Institute for Medical Research, Annual Report. Kuala Lumpur.

COLES, E. H. (1974): Veterinary Clinical Pathology. Philadelphia and London. W.B. Saunders Co.

DUKES, K. L. (1963): Erythrocyte diameter in *Tragulus javanicus,* the Chevrotain or Mouse-deer. *Anat. Rec.* **147**: 239–241.

DUNAWAY, P. B. and LOVELL, L. L. (1965): Taxonomic relation of erythrocyte count, mean corpuscular volume and body. *Nature.* **205**: 481–484.

GULLIVER, G. (1870): On the size of the red corpuscles of the blood of Moschus, Tragulus, Orvetropus, Ailurus, and some other Mammalia, with historical notices. *Proc. Zool. Soc. London* pp 92–99.
Cited by Dukes, K. L. (1963): Erythrocyte diameter in *Tragulus javanicus,* the Chevrotain or Mouse-deer. *Anat. Rec.* **147**: 239–241.

HARRISON, J. (1966): An Introduction to Mammals of Singapore and Malaya. Malayan Nature Society. Singapore Branch. Singapore. Tien Wah Press.

JORDAN, H. E. (1938): Handbook of Haematology, 2. New York, Paul B. Hoeber, Inc.
Cited by Dunaway, P. B. and Lovell, L. L. (1965): Taxonomic relation of erythrocyte count, mean corpuscular volume and body. *Nature.* **205**: 481–484.

MEDWAY, Lord (1978): The wild Mammals of Malaya and Off Shore Islands including Singapore. Kuala Lumpur and London, Oxford Univ. Press.

SHEILD, J. (1971): A seasonal change in blood cell volume of the Rottnest Island Quokka, *Setonix brachyurus. J. Zool.* **165**: 343-354.

WILSON, G. R. and HOSKINS, L. (1975): Haematology and blood chemistry of the red kangaroo *Megaleia rufa* in captivity. *Aust. Vet. J.* **51**: 146–149.

WINTROBE, M. M. (1961): Clinical Haematology. (5th Edn.). Philadelphia. Lea and Febiger.

(Received 13 March 1979)