Thailand Statistician
January 2011; 9(1) : 1-20
http://statassoc.or.th
Contributed paper

Regression Analyses of Migration with an Application to Selected Provinces in Thailand

Amonpun Makmeesup, Elvin J. Moore* and Utomporn Phalavonk
Department of Mathematics, Faculty of Applied Science, King Monkut's University of Technology North Bangkok, Bangkok 10800, Thailand.
* Author for correspondence; e-mail: ejm@kmutnb.ac.th, elvinmoo@gmail.com

Received: 6 April 2010
Accepted: 10 November 2010

Abstract

Mathematical models of migration are developed that are suitable for describing migration in Thailand of people in four different occupation classes. Nine main types of human migration models and the data that they require are reviewed. However, data is available from the National Statistical Office of Thailand for only three of these models: Gravity, Expanded Gravity, and Neo-Classical/Labor Flow Models. The available Thai data does not fit any of these models in their original form. A linear regression analysis on logarithmically transformed data is carried out to find the variables that are most important for describing migration for provinces from four main regions: Central (Bangkok), North (Chiang Mai), North-East (Nakhon Ratchasima) and South (Songkhla). We analyze movement out from, and movement into, these four provinces for all other provinces for which data is available. The most important variables for moving out of Bangkok are unemployment (or labor force) in the destination province and distance to that province. For moving into Bangkok only labor force in the province of origin is important. For Chiang Mai, wage rate is the most important with distance being less important for moving out and cost of living or wage rate are the most important for moving in. For Nakhon Ratchasima, wage rate and labor force are important for moving out and labor force and distance for moving in. For Songkhla, labor force or wage rate are the most important for moving out. The results for moving in to Songkhla are not reliable. The results show that the reasons for migration for the four occupation classes can be different and that the reasons for moving in and moving out of each of the four provinces can also be different.

Keywords: expanded gravity model, gravity model, migration in Thailand, neo-classical/labor flow model.

1. Introduction

Migration is a permanent movement of people from one region to another. It is an important problem for many countries. It has advantages and disadvantages for the people and the countries. Migration models try to explain the movements that occur and the reasons people have for moving. People have many different reasons for moving. Migration models usually look at different classes of people, e.g., workers, professionals, male, female, young, old etc. Migration is important in Thailand and we are interested in looking at suitable models for Thailand. Many migration models try to develop utility functions which are a measure of the value people put on living in a particular region.

The purpose of this study is to develop a human migration model that is suitable for selected provinces in Thailand for classes of migrants whose main reason for moving is given as “jobs”. We use data collected by The National Statistical Office of Thailand on migration between the 76 provinces in Thailand in the year 2006 and develop a mathematical model that fits available data for movement in and out of four provinces that are representative of the four main regions of Thailand: 1) Central (Bangkok), 2) North (Chiang Mai), 3) North-East (Nakhon Ratchasima) and 4) South (Songkhla).

An outline of this paper is as follows. In section 2, we give a review of nine migration models and select models for which the required data is available in Thailand. In section 3, we test three models for which data is available to see if they are suitable for explaining migration in Thailand. In section 4, we use linear regression on logarithmically transformed data to find variables that are important for migration for each of the four regions and four occupation classes and to develop models that fit the data. In section 5, we summarize the results and give a discussion and suggestions for development of improved models.

2. Survey of Migration Models

2.1 Gravity Model

Zipf [1] developed a model called the "gravity model of migration" for explaining migration between two towns or cities. The gravity model of migration is based upon the idea that as the size of one or both of the towns increases, there will be an increase in movement between them, and as the distance between two towns increases the movement between them will decrease. This phenomenon is called distance decay. The gravity model for migration between a region *i* and a region *j* is given by:

$$M_{ij} = k\frac{P_i \cdot P_j}{D_{ij}}. \quad (1)$$

M_{ij} = amount of migration between region *i* and *j*

P_i = population of region *i*

P_j = population of region *j*

D_{ij} = shortest distance between region *i* and *j* following a highway

k is a constant that reflects the relative weighting of the constituent variables.

2.2 Expanded Gravity Models

Lowry [2] developed a modified gravity model that included economic variables in addition to population to determine the attractiveness of a region. In 1967 Rogers [3] developed a modified form of Lowry's model which he called the "expanded gravity model". He applied this model to migration flow in California. The model took the form:

$$M_{ij} = k\left[\frac{U_i}{U_j} \cdot \frac{WS_j}{WS_i} \cdot \frac{LF_i \cdot LF_j}{D_{ij}}\right]. \quad (2)$$

M_{ij} = number of migrants going from location *i* to location *j*

U_i = civilian unemployment rate at location *i*

LF_i , LF_j = the available labor force, i.e., people of working age at *i* and *j*

WS_i, WS_j = wages and salaries at *i* and *j*

D_{ij} = shortest highway distance between *i* and *j*

k is a constant that reflects the relative weighting of the constituent variables.

2.3 Intervening Opportunities Model

In 1949, Strodbeck [4] stated the law of intervening opportunities in the form

$$y = k\frac{\Delta x}{x}. \tag{3}$$

y = expected number of people who migrate from one place to a given distance band around this place

Δx = number of opportunities in distance band

x = number of intervening opportunities between origin and center of distance band

In 1960 Stouffer [5] improved the intervening opportunities model because he believed it was not flexible enough as it only measured migration from one place to a distance band around this place. His new model was of the form:

$$y = k\frac{X_O \cdot X_1}{X_B \cdot X_C}. \tag{4}$$

y = number of people who move from city 1 to city 2

X_0 = total number of people who move out from city 1

X_1 = opportunities in city 2 measured from total number of people who move in

X_B = intervening opportunities between city 1 and city 2 measured from total number of people who move in to a circle that has diameter equal to distance from city 2 to city 1

X_C = migrants who may compete for opportunities in city 2 measured from number of total people who move out from all cities in a circle that has center at city 2 and a radius equal to distance from city 2 to city 1.

2.4 Neo-classical/labor-flow approach Model

Harris and Todaro [6] suggested that a person decides to migrate based on the expected income differences between two areas, e.g., rural and urban areas. This model is called Neo-classical/labor-flow approach model. In this model, workers will move from rural to urban areas if:

$$w_r < \frac{l_e}{l_{us}} w_u . \tag{5}$$

Conversely, urban to rural migration will occur if

$$w_r > \frac{l_e}{l_{us}} w_u \,, \qquad (6)$$

where

w_r = wage rate in the rural agricultural sector.

l_e = total number of jobs available in the urban sector.

l_{us} = total number of job seekers, employed and unemployed, in the urban sector.

w_u = wage rate in the urban sector, which could possibly be set by government with a minimum wage law.

2.5 Human Capital Model

Dierx [7] said migration is an investment decision in which individuals calculate their present discounted value of expected returns in every location. Utility functions in this model are a function of age, location-specific human capital, non-specific human capital and rent per unit of capital.

2.6 Interrelationship Model

Hägerstrand [8] assumes that when people move the number of contacts or interrelationships between people in the original location and the destination location is important. For example, if a migrant has a friend or relative to stay with in the new location then they can stay with them on their first arrival in the new location. So friends and relatives are important for flow. In estimating the number of contacts Hagerstrand uses the amount of movement between origin and destination in the past.

2.7 Household Migration Models

In the 1970s, household migration models became quite popular. In these models a family decides whether to move or not based on net family gain from moving from one location to another location. Utility functions in this model are a function of age, children, husband’s education, wife’s education, wife’s labor force participation, migration history, in-migration, home ownership, family income, commuting, unemployment experience, area unemployment rate, size of municipality, share of agriculture and share of industry [9].

2.8 Network Models

Kalashnikov et al [10] presented a new type of model called a Networks model. They said that there is evidence of strong flows between certain parts of countries and not others that are not explained by the neo-classical, human capital and household migration models. The authors explained that migrants from one country to another often cluster in specific locations in the new country. This proves the importance of networks that link new migrants by ties of ethnicity, kinship and friendship to earlier migrants. Kalashnikov et al [10] found that their model gave a reasonable fit to the actual numbers of migrants of selected classes between three cities in Mexico in the period 1980-2000.

2.9 Example of Migration in Thailand

Prasartkul [11] has estimated the numbers of people who migrate into Nakhon Pathom in the years 1980 to 1990, 1997 and 1998. He used the following methodology :

1. Assume rate of change of population in province = rate of change of population in country,
2. Calculate number of emigrants in province from:
   number of population in province - number of population in province $\times$ rate of change of population in country,
3. Add number of births in a province,
4. Subtract number of deaths in a province.

3. Model Fitting with Migration Data of Thailand

The available data in Thailand that are expected to be of importance for building a migration model for Thailand are income, occupation, cost of living, employment rate, distance, population in each area, age, marital status and children. The three models for which data are available from the National Statistical Office of Thailand that we can select to test are: gravity model, expanded gravity model, and neo-classical/labor-flow approach model. To test the gravity model we first modify it as follows:

$$M_{ij} = k\frac{P_i^{b_1} \cdot P_j^{b_2}}{D_{ij}^{b_3}}, \qquad (7)$$

where b_1, b_2, b_3 are constants that should be equal to 1 if the gravity model is correct. We then take the logarithm of Eq.7 and carry out a linear least-squares regression analysis on the logarithmically transformed Eq. 8

$$\log \hat{M}_{ij} = \log k + b_1 \log P_i + b_2 \log P_j - b_3 \log D_i , \qquad (8)$$

where $\log \hat{M}_{ij}$ is the expected value of $\log M_{ij}$. If the gravity model is suitable, the values of the coefficients should be $b_1 = b_2 = b_3 = 1$. If the conditions are not satisfied, then the variable corresponding to the coefficient of largest value can be regarded as the most important variable for predicting migration.

Table 1. Regression results for the gravity model in selected provinces in Thailand.

Variable	Coefficient	95% Confidence interval		Statistics
Constant	-7.3137	-14.5542	-0.0732	R^2=0.4609 F=10.2585 P=0.0001 SE=0.6038
log P_i	0.7784*	0.4829	1.0739	
log P_j	0.3352	-0.0101	0.6806	
log D_{ij}	-0.0231	-0.2319	0.1857	

The results of the regression estimation are shown in Table 1 for movements between Bangkok and 10 other provinces, Chonburi and 10 other provinces, Samut Prakan and 10 other provinces, and Nakhon Ratchasima and 10 other provinces. The results show that the gravity model does not give a good fit to Thai data because the values of the coefficients do not agree with the model, there are large 95% confidence intervals for each coefficient and $R^2 = 0.4609$. However, the data do suggest that the distance is not an important factor as the value of the log D_{ij} coefficient (b_3 = -0.023) is close to zero. Also, the data suggest that the populations of the origin and destination are of importance as the expected values of the coefficients are not close to zero, with the origin population (b_1 = 0.778) being more important than the destination population (b_2 =0.335). However, both populations are of less importance than in the gravity model in which all coefficient values are 1. Note that for both b_2 and b_3, the 95% confidence intervals include both positive and negative values, where a positive value means that migration increases as the value of the corresponding variable increases and a negative value means that migration decreases as the value of the variable increase. From the results we can conclude that the gravity model does not fit the Thai data.

Next we tested Thai data on the expanded gravity model using a similar method. We found that this model also does not fit the Thai data as the values of the coefficients estimated from linear regression did not agree with the values in the model.

Finally, we tested the neo-classical/labor-flow approach model. We found that this model successfully predicted net movements between 25 of the 40 pairs of cities that we tested. When we selected data for the single occupation class professional, technical and administrative workers it correctly predicted the direction of movement for 3 out of the 6 pairs of cities. Therefore, this model fails to predict the correct directions of movements for many pairs of cities. In addition, this model is a utility function type model of the form: if utility at location i is greater than utility at location j, then people will move from i to j, i.e., it predicts that migration should be in one direction. However, in the Thai data, there are usually large movements in both directions and it is difficult to explain these movements with a single utility function.

We can conclude that, in their original form, these three models are not suitable for fitting Thai data. However they do suggest variables that may be of importance in building a suitable model. The gravity model results show that the distance between provinces does not appear to be an important factor in migration decisions. However, the populations of the cities could be of importance. The expanded gravity model results show that wage rages at provinces which migrants leave and the provinces to which they move are important factors. The distance appears to be of some importance, but much less than the model assumes. The neo-classical/labor-flow approach model suggests that wage rates, labor force and unemployment are important variables that should be included in a model of people who migrate for "job reasons".

4. Migration Model for Selected Provinces in Thailand

In this section we describe models that we have developed that give reasonable fits to the data available for the year 2006 for selected provinces in Thailand. We selected 4 provinces for detailed analysis, Bangkok, Chiang Mai, Nakhon Ratchasima and Songkhla as these are typical provinces in the Central, Northern, North-Eastern and Southern regions of Thailand, respectively. As the 3 models that we consider were developed to explain migration of people who move to obtain employment, we will examine the Thai data on people who give “job reasons” as their main reason for migration.

In developing the migration models, we have found it necessary to group the migrants into four different occupation classes rather than the 10 classes of the WHO classification [12]. The four classes are: 1) professional, technical and administrative workers, 2) clerical, sales and services worker, 3) production workers (people who work in factory, machine and plant) and 4) skilled (e.g. farmer, fishermen). The main reason is that the available wage and salary data is grouped into the 4 classes shown. We use this data to build migration models. We will look in detail at the variables that appear in the expanded gravity model and the neo-classical/labor flow model.

The original form of the expanded gravity model is given in Eq. 2. We have already stated in section 3 that Eq. 2 does not give a good fit for Thai data. We consider a modified version of the form:

$$M_{ij} = k\left[\frac{U_i^{b_4}}{U_j^{b_5}} \cdot \frac{WS_j^{b_7}}{WS_i^{b_6}} \cdot \frac{LF_i^{b_1}}{D_{ij}^{b_3} LF_j^{b_2}}\right]. \tag{9}$$

In Eq. 9, we use U_i, U_j as the total unemployment at locations *i* and *j* rather than as the unemployment rates that were used in the expanded gravity model. We now take logarithms of Eq. 9 for the purposes of estimation via linear least-squares regression analysis. Eq. 9 then takes the following form:

$$\begin{aligned}\log \hat{M}_{ij} &= \log k + b_1 \log LF_i + b_2 \log LF_j - b_3 \log D_{ij} \\ &\quad + b_4 \log U_i - b_5 \log U_j - b_6 \log WS_i + b_7 \log WS_j\end{aligned} \tag{10}$$

We have used Minitab for the linear least-squares regression. Note: For convenience, in the remainder of this paper the expected value $\hat{M}_{ij}$ of the response variable obtained from the linear regression will be denoted simply by M_{ij}.

As costs of living (C_i, C_j) might also be important variables, we include these in the model. As stated above, data is available for four occupation classes and therefore we develop separate models for each of these classes. We begin by analyzing the data for Bangkok. We also modify the expanded gravity model by analyzing migration out of Bangkok and into Bangkok separately as we want to determine if people moving in and out move for different reasons.

4.1 Linear Regression for Moving out from Bangkok

Data is available for migration out from Bangkok into 51 provinces. An example of the data that we use is given in Table 2. For migration out, the variables LF_i ,U_i and WS_i in Eq.9 refer to the location of origin (Bangkok) and are therefore constant. Eq. 10 reduces to the form:

$$\log M_{ij} = \log k + b_2 \log LF_j - b_3 \log D_{ij} - b_5 \log U_j + b_7 \log WS_j + b_9 \log C_{j.} \quad (11)$$

Table 2. Example of data of movement out of Bangkok for occupation class professional, technical and administrative workers for 28 of the total number of 51 provinces included in the data analysis.

Province	LF_j	U_j	D_{ij}	WS_j	C_j	M_{ij}
Nonthaburi	710613	9285	20	47670	37777	1142
Pathum Thani	442839	8338	45	46973	34291	438
Sing Buri	131075	2146	142	37997	27615	89
Chai Nat	208748	2886	194	40587	36006	12
Saraburi	38698	830	107	34425	26152	188
Chon Buri	640477	7197	81	33086	27918	881
Rayong	311600	4000	179	34641	29149	1034
Chanthaburi	327330	2620	245	34141	30841	207
Chachoengsao	356547	5273	82	31873	29662	63
Prachin Buri	228671	4310	136	33464	27240	44
Nakhon Ratchasima	1447676	34815	259	32064	21588	1040
Buri Ram	787259	16991	383	31053	20684	816
Surin	725400	7100	426	30667	26006	1206
Yasothon	299641	2373	531	27630	22448	17
Am Nat Charoen	306538	4270	585	36502	27991	40
Nong Bua Lam Phu	392152	6395	577	24934	22948	260
Khon Kaen	837159	16465	449	27016	20118	588
Ratchaburi	495764	7795	100	41501	30952	75
Kanchanaburi	479313	6687	128	33634	28388	245
Suphan Buri	478030	3637	100	32603	24608	11
Samut Sakhon	348169	3000	36	35164	23872	904

Table 2. (continued) Example of data of movement out of Bangkok for occupation class professional, technical and administrative workers for 28 of the total number of 51 provinces included in the data analysis.

Province	LF_j	U_j	D_{ij}	WS_j	C_j	M_{ij}
Phetchaburi	245463	3126	123	36265	27468	75
Prachuap Khiri Khan	269328	1922	281	34450	30162	92
Krabi	218229	3611	814	28151	25566	27
Songkhla	739474	12723	950	36650	30893	284
Satun	144225	1844	973	26647	19275	27
Phatthalung	975300	5900	840	34695	37033	12

Source: The National Statistical Office of Thailand year 2006

$LF_{i,j}$ = the labor forces in locations *i* and *j* $U_{i,j}$ = unemployment in *i* and *j*

$D_{i,j}$ = shortest highway distance between *i* and *j*

$WS_{i,j}$ = wages and salaries in *i* and *j* $C_{i,j}$ = cost of living in *i* and *j*

$M_{i,j}$ = number of migrants going from *i* to *j*

In our model, the variables log *k* (constant), $\log LF_j, \log U_j, \log D_{ij}, \log WS_j$ and $\log C_j$ are the predictor variables and, as noted above, $\log M_{ij}$ means the expected value of the response variable. The first step in a regression analysis is to determine if there is correlation between the variables in the model [13]. Table 3 shows the results obtained from Minitab for correlations between the variables in Eq.11 for the first occupation class of professional, technical and administrative workers. We look for predictor variables that are correlated with the response variable, but not with each other. Correlation between two predictor variables shows that the variables are not independent and only one should be included in a final model. In Table 3 predictor variables are regarded as independent and not correlated if the p-value > 0.05. Therefore predictor variables with p-value < 0.05 are regarded as correlated and not independent. Only one of each correlated pair can be included in a regression model. Therefore, only one of $\log LF_j$ and $\log U_j$ and one of $\log WS_j$ and $\log C_j$ can be included as independent predictors. From the $\log M_{ij}$ row we see that $\log U_j$ has better correlation with $\log M_{ij}$ than $\log LF_j$. We therefore choose $\log U_j$. Both

$\log WS_j$ and $\log C_j$ are not correlated with $\log M_{ij}$ and therefore we do not include them in the model. We therefore use $\log U_j$ and $\log D_{ij}$ as predictor variables.

Table 3. Correlations between variables for movement out of Bangkok for occupation class professional, technical and administrative workers.

Correlations: log_LF$_j$, log_U$_j$, log_D$_{ij}$, log_WS$_j$, log_C$_j$, log_M$_{ij}$

	log_LF$_j$	log_U$_j$	log_D$_{ij}$	log_WS$_j$	log_C$_j$
log_U$_j$	0.793				
	0.000*				
log_D$_{ij}$	0.019	0.021			
	0.899	0.888			
log_WS$_j$	0.022	0.039	-0.390		
	0.883	0.788	0.006*		
log_C$_j$	0.019	-0.043	-0.227	0.800	
	0.899	0.770	0.117	0.000*	
log_M$_{ij}$	0.449	0.480	-0.308	0.105	-0.037
	0.001	0.000**	0.031	0.473	0.802

Cell Contents: Pearson correlation
P-Value

We next test if a constant term should be included by analyzing correlations between predictor variables and response variable with a constant term included. Table 4 shows correlations between predictor variables and response variable for movement out of Bangkok for occupation class professional, technical and administrative workers. The columns in the table are the coefficients in Eq. 11, the standard error for each coefficient, the t-values and the p-values for the hypotheses that the coefficients in the regression formula are zero. A p-value > 0.05 shows that the hypothesis that the coefficient is zero must be accepted. The table shows that the constant term should not be included in the model because p-value > 0.05. The table also shows that unemployment (U_j) and distance (D_{ij}) are important for migration because the p-value of both variables is less than 0.05. The regression equations are of the form:

$$\log M_{ij} = 0.92 \log U_j - 0.515 \log D_{ij} \quad \text{or} \quad M_{ij} = \frac{U_j^{0.92}}{D_{ij}^{0.515}} . \qquad (12)$$

Table 4. Regression analysis for movement out of Bangkok for occupation class professional, technical and administrative workers.

Regression Analysis: log_M_{ij} versus log_U_j, log_D_{ij}

Predictor	Coef	SE Coef	T	P
Constant	1.977	5.798	0.34	0.735
log_LF_j	0.1154	0.5200	0.22	0.826
log_U_j	0.8488	0.4632	1.83	0.074
log_D_{ij}	-0.4711	0.2306	-2.04	0.048
log_WS_j	1.136	1.928	0.59	0.559
log_C_j	-1.718	1.674	-1.03	0.311

Predictor	Coef	SE Coef	T	P
Noconstant				
log_U_j	0.9199	0.1338	6.88	0.000
log_D_{ij}	-0.5151	0.1926	-2.67	0.010

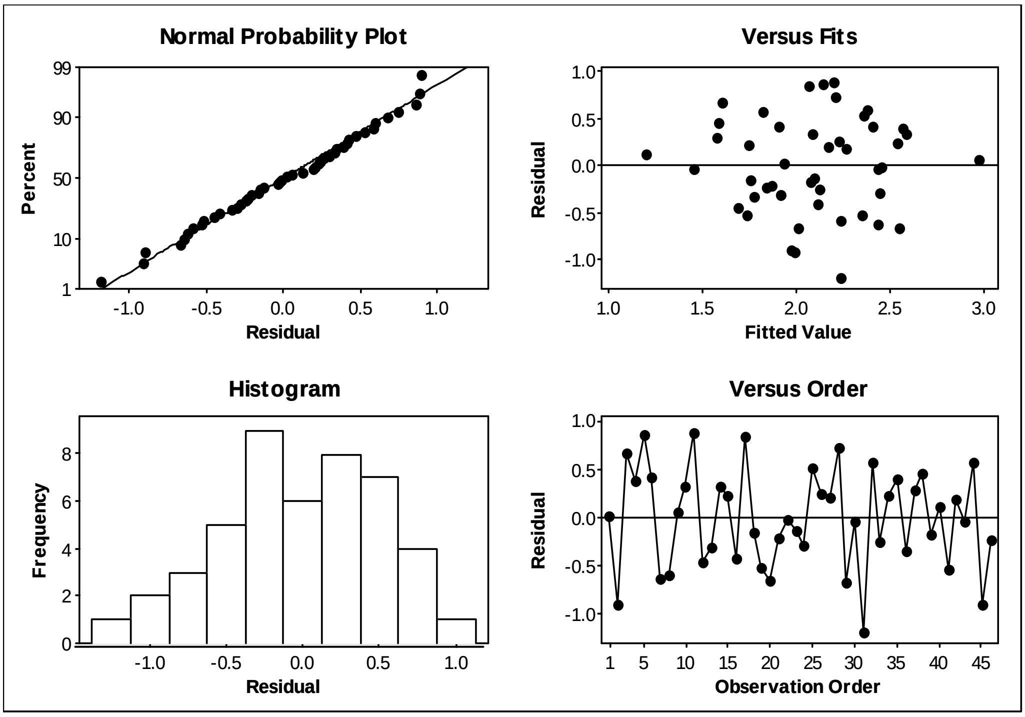
Analysis of Variance

Source	DF	SS	MS	F	P
Regression	2	205.49	102.74	388.70	0.000
Residual Error	44	11.63	0.26		
Total	46	217.12			

We can find the coefficient of determination (R^2) from $\frac{SSR}{SST}$, where *SSR* is the sum of squares due to regression and *SST* is total sum of squares. From Table 4, the value of R^2 = 0.95 shows that mean unemployment (U_j) and distance (D_{ij}) can describe migration (M_{ij}) out of Bangkok for professional, technical and administrative workers with a reliability of 95%.

The results shown above are based on an ordinary least-squares (OLS) analysis. The final and necessary step is to check the errors in the model using residual diagnostic tools. The OLS analysis assumes that the errors: 1) Exhibit constant variance, 2) are normally distributed and 3) are independent from each other. These assumptions can be checked by looking at the residual plots in Figure 1. The normal probability plot and the histogram show that the residuals have a normal distribution. We see directly that assumption (2) appears to be met as the p-value for the Anderson Darling test (0.882) for normality is greater than 0.05. Hence we are unable to reject the null hypothesis of normality. The fitted value plot shows that the residual variations are constant so assumption (1) is proved. The rapid changes in the order plot show that there is no correlation in the residuals and they are independent of each other and the Durbin-Watson statistic value of 2.21 is greater than 1.62 (d_U) and so assumption (3) is correct. From the form of the four graphs in Figure 1, we can say that the assumptions on the behavior of the residuals in an OLS analysis are satisfied and therefore that the regression model can satisfactorily describe the data [13].

Residual Plots for log_M$_{ij}$



Durbin-Watson statistic = 2.21042, P-value = 0.882 (Anderson–Darling test)

FIGURE 1. Residual plots for movement out of Bangkok for occupation class professional, technical and administrative workers.

The regression analysis for migration out from and into Bangkok, Chiang Mai, Nakhon Ratchasima and Songkhla for the 4 occupation classes and for all classes combined follows a similar process to that given for the class 1 professional, technical and administrative workers who move out of Bangkok. Table 5 contains a summary of the results from the linear regression fits for all cases considered.

Table 5. A comparison of regression equations for each occupation class for movement out of and movement into 4 provinces.

Class		**Bangkok**	**Chiang Mai**	**Nakhon Ratchasima**	**Songkhla**
1*	Out	$M_{ij} = \frac{U_j^{0.92}}{D_{ij}^{0.515}}$	$M_{ij} = WS_j^{0.567}$	$M_{ij} = WS_j^{0.538}$	$M_{ij} = LF_j^{0.474}$
		R^2 =0.95 / S**	R^2 =0.96 / S**	R^2 =0.95 / S**	R^2 =0.98 / S**
1*	In	$M_{ij} = LF_i^{0.458}$	$M_{ij} = C_i^{0.631}$	$M_{ij} = \frac{LF_i^{1.23}}{D_{ij}^{1.71}}$	$M_{ij} = C_i^{0.758}$
		R^2 =0.99 / S**	R^2 =0.99 / S**	R^2 =0.999 / U**	R^2 =0.999/F**
2*	Out	$M_{ij} = \frac{U_j^{0.88}}{D_{ij}^{0.471}}$	$M_{ij} = WS_j^{0.589}$	$M_{ij} = WS_j^{0.537}$	$M_{ij} = LF_j^{0.492}$
		R^2 =0.94 / S**	R^2 =0.96 / S**	R^2 =0.95 / S**	R^2 =0.97 / S**
	In	$M_{ij} = LF_i^{0.445}$	$M_{ij} = WS_i^{0.646}$	$M_{ij} = \frac{LF_i^{1.22}}{D_{ij}^{1.76}}$	$M_{ij} = C_i^{0.724}$
		R^2 =0.99 / S**	R^2 =0.99 / S**	R^2 =0.999 / U**	R^2 =0.999/F**

Table 5. (continued) A comparison of regression equations for each occupation class for movement out of and movement into 4 provinces.

Class		Bangkok	Chiang Mai	Nakhon Ratchasima	Songkhla
3*	Out	$M_{ij} = \frac{U_j^{0.825}}{D_{ij}^{0.524}}$	$M_{ij} = \frac{WS_j^{1.15}}{D_{ij}^{1.02}}$	$M_{ij} = WS_j^{0.624}$	$M_{ij} = LF_j^{0.34}$
		R^2 =0.92 / S**	R^2 =0.97 / S**	R^2 =0.96 / S**	R^2 =0.95 / S**
	In	$M_{ij} = LF_i^{0.4}$	$M_{ij} = C_i^{0.553}$	$M_{ij} = \frac{LF_i^{1.24}}{D_{ij}^{1.68}}$	$M_{ij} = C_i^{0.622}$
		R^2 =0.99 / S**	R^2 =0.98 / S**	R^2 =0.999 / U**	R^2 =0.998/F**
4*	Out	$M_{ij} = \frac{LF_i^{0.805}}{C_j^{0.916}}$	Not enough data	$M_{ij} = LF_j^{0.331}$	$M_{ij} = WS_j^{0.514}$
		R^2 =0.80 / S**		R^2 =0.93 /**	R^2 =0.97 / S**
	In	$M_{ij} = LF_i^{0.55}$	Not enough data	$M_{ij} = \frac{LF_i^{1.16}}{D_{ij}^{1.72}}$	Not enough data
		R^2 =0.99 / S**		R^2 =0.999 / U**	
5*	Out	$M_{ij} = \frac{U_j^{0.824}}{D_{ij}^{0.499}}$	$M_{ij} = \frac{WS_j}{D_{ij}^{0.744}}$	$M_{ij} = \frac{WS_j^{0.843}}{D_{ij}^{0.52}}$	$M_{ij} = WS_j^{0.541}$
		R^2 =0.88 / U**	R^2 =0.96 / S**	R^2 =0.94 / U**	R^2 =0.95 / S**
	In	$M_{ij} = LF_i^{0.454}$	$M_{ij} = \frac{C_i^{0.979}}{D_{ij}^{0.596}}$	$M_{ij} = \frac{LF_i^{1.2}}{D_{ij}^{1.68}}$	$M_{ij} = WS_i^{0.686}$
		R^2 =0.98 / U**	R^2 =0.99 / S**	R^2 =0.99 / U**	R^2 =0.99 / S**

Note: * Occupation class:

1) professional, technical and administrative workers,
2) clerical, sales and services worker
3) production workers
4) skilled (e.g. farmer, fishermen)
5) combined 4 classes of occupation

** Results of residual data test for using ordinary least squares (OLS) method for linear regression

S = satisfied, U = unsatisfied, F= not enough data for test

5. Results and Discussion

Further details on the data and results given in this paper can be found in reference [14].

The most important variables for movement between Thai provinces are labor force and wage rate, with distance also being important in some cases.

Bangkok:

- The most important variables for moving out are unemployment (or labor force) and distance. The only important variable for moving in is labor force in the original province. The data show that the movement is approximately proportional to the square root of the labor force.

Chiang Mai:

- The most important variable for moving out is wage rate, with distance sometimes also being important.
- The only important variable for moving in is cost of living (or wage rate). For Chiang Mai the labor force is not important either for moving out or moving in.

Nakhon Ratchasima:

- The most important variable for moving out for occupation classes 1-3 is wage rate. For occupation class 4, the most important variable is labor force.
- The most important variables for moving in are labor force and distance. However, these results do not pass the required residual tests and cannot be relied on. The powers of the variables in the equations appear much higher (greater than 1), than any of the powers in the results for other provinces.

- The results for Nakhon Ratchasima for moving out pass the coefficient of determination (R^2) test for reliability and residuals test. For moving in the coefficient of determination (R^2) test is satisfied, but the Durbin-Watson statistics show that the residuals are not independent. However, the tables for correlation (not shown in this paper) between the predictor variables and between the response and predictor variables show that the variables mentioned above are important variables for explaining the trend of the data.

Songkhla:

- The most important variable for moving out is labor force for occupation classes 1-3 and wage rate for class 4.
- The most important variable for moving in is cost of living (or wage rate). However, these results are only for 3 provinces.

The distance is important for moving out of Bangkok and for occupation class 3 for Chiang Mai. The results also suggest distance may be important for Nakhon Ratchasima, but these results are not reliable.

The distance does not appear to be important for moving in to Bangkok or into any other large province. This result agrees with the suggestion in the Ravenstein model [15] that for workers moving into big cities distance is not important, whereas for moving into smaller cities or villages distance is important.

For each province, the results for occupation classes 1-3 are usually similar to each other, but are different for each province. For each province, the class 4 (skill) results for move out are either different from the other 3 classes or there is not enough data for the class.

The results shown under “5” in Table 1 are combined results for all 4 occupation classes. The results are usually different from the results for the individual classes or are shown to be unreliable by the statistical tests. This suggests that conditions of salary, labor force, unemployment, and cost of living for migrants in the different occupation classes can be sufficiently different that combining the data gives unreliable results. Migrants in the different occupation classes can also have different reasons for moving.

In this paper we have developed models for people in four occupation classes who move for job reasons into and out of Bangkok, Chiang Mai, Nakhon Ratchasima and Songkhla. The National Statistical Office of Thailand collects data on people who move for job reasons between all 76 provinces in Thailand. The method of linear regression of logarithmically transformed data could be applied to study migration of people for job reasons between other provinces in Thailand. The National Statistical Office of Thailand also collects data on variables such as age, marital status and children. However, this data is not available for people in the different occupation classes. Our results have shown that people in the different occupation classes can have different reasons for moving and therefore these variables cannot be included easily in models based on occupation class.

In this paper, we have studied only three of the nine different migration models. For the remaining six models, the required migration data is not available in Thailand. Collection of some or all of this migration data would mean that migration in Thailand could be analyzed using the mathematical models that have been successfully applied to study migration in other countries. We have only used linear regression of logarithmically transformed data to analyze the migration data for the year 2006. In order to build a satisfactory model of migration it will be necessary to look at data for more than one year and to use alternative methods of statistical analysis. A study of migration changes over time is important for forecasting population movements so that the government and other organizations can plan for the future. Some of the alternative statistical methods that could be used to analyze migration data include: Time series, forecasting methods, Bayesian statistics, Markov chains and stochastic models [16].

Acknowledgements

The authors would like to thank Assistant Professor Dr. Winai Bodhisuwan and Assistant Professor Dr. Tidadeaw Mayureesawan for invaluable suggestions regarding the methods of statistical analysis used in this paper. The authors would also like to thank the anonymous referees for helpful suggestions.

References

[1] Zipf, G.K., *Human Behavior and the Principle of Least-Effort*. Addison-Wesley, 1949.

[2] Lowry, I.S., *A Model of Metropolis* RAND Memorandum 4025-RC, 1964.

[3] Rogers, A., A Regression Analysis of Interregional Migration in California. *Review of Economics and Statistics*, 1967; 49: 262-267.

[4] Strodbeck, F., Equal opportunity intervals: a contribution to the method of intervening opportunity analysis, *American Sociological Review*, 1949; 14: 490-7.

[5] Stouffer, S., Intervening opportunities and competing migrants, *Journal of Regional Science*, 1960; 2: 1-26.

[6] Harris, J.R. and Todaro M.P., Migration, Unemployment and Development: A Two Sector Analysis, *American Economic Review*, 1970; 60: 126-142.

[7] Dierx, A.H., Estimation of a Human Capital Model of Migration. The Annals of Regional Science, Springer, November, 1988; 22: 99-110.

[8] Hägerstrand, T., Innovationsförloppet ur korologisk synpunkt, C.W.K Gleerup, Lund, Sweden, 1953. (English translation: Innovation Diffusion As a Spatial Process, University of Chicago Press, Chicago, 1967)

[9] Nivalainen, S., Determinants of family migration: short moves vs. long moves, *Journal of Population Economics,* 2004; 17: 157-175.

[10] Kalashnikov, V.V., Kalashnykova, N.I., Rojas, R.L., Munos, M.M., Uranga, C., Rojas, A.L., Numerical experimentation with a human migration model*. European Journal of Operational Research*, 2008; 189: 208-229.

[11] Prasartkul, P., *Educational Foundations for Population*. The Institute for Population and Social Research, Mahidol University, 2000.

[12] International Standard Classification of Occupations *(ISCO-88)*. International Labor Office. Geneva. 1990.

[13] Montgomery, DC., Peck, E.A., Viming, G.G., *Introduction to Linear Regression Analysis*. John Wiley & Sons, New York, 2007.

[14] Makmeesup, A., *Mathematical Models of Migration with an Application to Selected Provinces in Thailand*, Master of Applied Mathematics thesis, King Mongkut's University of Technology North Bangkok, 2010.

[15] Ravenstein, E., The Laws of Migration, *Journal of the Statistical Society of London*, 1885; 48:167-235

[16] Raymer, J. and Willekens, F*. International Migration in Europe*, John Wiley & Sons Ltd, London, 2008.