ESTIMATING SIZE AND STRUCTURE OF YOUGH ROMANIAN MIGRANTS BASED ON THE GRAVITY MODELS

Univ. Prof. Ph. D. Mariana BALAN
„Athenaeum" University – Bucharest
dr.mariana.balan@gmail.com

PhD. Rodica PERCIUN
National Institute for Economic Research - Academy of Sciences of Moldova

Abstract

Within the specialised literature are found an impressive number of models for studying migration/mobility of labour force both at national and international level. Several of them give a description of either international migration compared to the domestic one, or of the inter-regional migration from Europe or other areas of the world.

Among the multiple analysis methods and techniques for studying migration/mobility of labour force are counted: the Markov Chains method, the Harris-Todaro Model with its variants and extensions, the models for studying the economic impact of labour force mobility in an enlarged Europe, the models for studying the mobility of researchers, the models for studying labour force mobility by wages' flexibility, the models for characterising the impact of labour force mobility on some macroeconomic indicators and others.

From the multitude of the models used within the specialised literature for studying the migration phenomenon, in this paper are presented the outcomes of applying the gravity model for the study of Romanian emigration.

Key-words: youth migration, skilled youths, gravity model
JEL Classification: C13, C32, F22, J13, R23

1. Introduction

International migration is a complex, multidimensional phenomenon which makes its modelling and forecast to involve the analysis/use of methods developed by various disciplines: demography, geography, economy, statistics, sociology, political sciences, or even advanced theoretical mathematics and physics models.

Because migration theories are much too fragmented, the only option for obtaining credible analyses and forecasts about population flows is the use of both determinist and probabilistic methods and models that can be potentially interconnected.

In the neoclassical approach at macro level, migration is a consequence of the economic opportunities differences between regions or countries, in particular of the differentiated earnings. Migration is seen as an optimum allocation of labour force in regions with the highest productivity, leading to wages' levelling in the absence of migration costs.

Some studies, based on the Harris-Todaro model attempt to identify some positive influence in the probability of an individual taking the decision to migrate. Other studies have highlighted that individuals are attracted to regions with jobs' deficit.

The impact of immigration depends also on the skills' level of the migrants.

Migrants are not a random sample from the population of a country. This aspect already played an important role in the brain-drain debate regarding the emigration of individuals with higher education and the negative consequences upon the country of origin. The self-selection question of migrants returned within the performance analyses of the migrants on the labour force market from numerous host-countries.

The vast majority of high-skilled migrants are voluntary ones. Because their skills and competencies are on demand in the country of destination (and, increasingly more, the immigration policies tend to attract them) the overwhelming majority of them are legal immigrants. They are very often, but not always, permanent migrants.

In the case of high-skilled migrants, the country of destination receives the human capital as a contribution. This country had a "brain" gain, while the country of origin suffered on exodus of the "brains". It should be mentioned that, usually, the discussion is about a "brain" gain, respectively a "brain" exodus only if the number of migrants is considerable, as well as in the case when the equilibrium between high-skilled immigrants and emigrants is markedly disproportionate. The situations characterised by a relative equilibrium are very often called situations of "brain" exchange.

The small number of reliable and comparable papers at international level regarding high-skilled migration determined various national and international bodies to evaluate the size and effects of the migration of this special group. Additionally, high-skilled migration does not necessarily lead to brain gain in the country of destination.

The literature deficit about professional mobility was used as motivation for undertaking such analyses for several decades (Shaw, 1978; Parrado et al., 2007) these remaining to their largest extent very few as compared with the specialised literature about the geographic mobility. One of the reasons is due to the fact that "*it is very difficult to quantitatively define and measure an occupation"* (Shaw, 1978: 703).

The standardisation of occupational classification was set up in the vast majority of developed countries, but the structure and the relations between a high number of occupational classifications are not yet identified or known.

2. Brief presentation of the gravity model

Based on an analogy between physics and Newton's idea about the gravity force, the gravity model used in economy was developed by Lowry (1966) and thereafter it was extended by various economic variables for representing the push and pull effects on individuals.

Sometimes, the gravity model is used not only for describing the types of geographic mobility based on physical distances, but also for quantifying the similarities between pairs of occupations in a changing matrix of jobs.

Within the classic gravity model for analysing migration, a flow from the location i to the location j is regarded as being proportional with the population of the country of origin and of the country of destination, and in reverse proportion to the distance between the two locations:

$$M_{ij} = a \cdot P_i \cdot P_j \cdot f(d_{ij}) \quad (1)$$

where: a is a constant associated to mobility as a whole;

P_i and P_j represent the population of the country of origin, respectively of the country of destination;

$f(d_{ij})$ a reverse function of the distance between the two locations.

Because in space $d_{ij} = d_{ji}$, it results that net migration between the two locations is equal to zero. Consequently, the gravity model in this formulation can be interpreted as describing random flows that take place in the absence of net migration. In the empirical analysis of the geographic mobility, where net migration appears, the gravity model is usually expanded with the generalising attraction factor, so that the regional economic variables can be incorporated into the migration equation as determinant factors (Molho, 1986):

$$M_{ij} = a \cdot P_i \cdot P_j \cdot f(d_{ij}) \times A_i B_j \quad (2)$$

where: A_i is the "push" factor from the country of origin and B_j is the "pull" factor of the country of destination.

Both *A* and *B* have positive values and are functions with explanatory variables for each location. For maintaining the consistency with the initial equation (1), both variables should be calibrated.

In time, various variants of the original gravity model were realised with the purpose of describing as good as possible the evolution of the considered economic phenomenon.

Thus, in the paper "The Deregulation of People Flows in China: Did the Structure of Migration Change?", Shuming Bao, Örn B. Bodvarsson, Jack W. Hou and Yaohui Zhao have developed several versions of the traditional gravity model for internal migration applied to the study of inter-provincial migration.

The uniqueness of this version consists in including provincial investments and the control of migrants' networks, but also of other various controls with respect to the economic, political and social characteristics of the province. As dependent variable is regarded the logarithm of the gross inter-provincial emigration rate ($\log(M_{ij})$), computed as the emigration volume from the province *i* to the province *j* divided by total inter-provincial migration from the province *i*:

$$\begin{aligned}\log M_{ij} &= \alpha_0 + \alpha_1 \log D_{ij} + \alpha_2 \log NETWORK + \alpha_3 \log FDI_{ji} + \alpha_4 \log FAI_{ji} + \\ &+ \alpha_5 (ogFDI_{ji})(\log FAI_{ji}) + \alpha_6 \log Y_{ji} + \alpha_7 \log E_i + \alpha_8 \log E_j + \alpha_9 \log U_i + \\ &+ \alpha_{10} \log U_j + \alpha_{11} \log MANEMP_{ji} + \alpha_{12} \log URBAN_{ji} + \\ &+ \alpha_{13} \log MINORITY_{ji} + \alpha_{14} \log WARM_{ji} + \varepsilon_{ij}\end{aligned} \quad (3)$$

where:

D_{ij} = the distance on the railroad (in km) between the capital of province *i* and the one of province *j*;

NETWORK = the size of the migrants' community with residence in the province *j* who migrated from province *i,* measured as rate of past migration flows;

FDI_{ji} = the ratio between foreign direct investments per capita in province *j* and of foreign direct investments per capita in province *i*;

FAI_{ji} = the ratio between real domestic investments of the fixed assets per capita in province *j* and the real domestic investments of fixed assets per capita in province *i*;

Y_{ji} = the ratio between the real per capita income in province *j* and the real income per capita in province *i*;

E_i ;E_j = the education level in the province *i* and, respectively, *j*;

U_i ; U_j = the unemployment rates during the week preceding the enforcement of the census in the province i, and respectively *j*;

$MANEMP_{ji}$ = the ratio between the labour force employment weight in the productive sector in the province *j* depending on the labour force employment weight in the productive sector in the province *i*;

$URBAN_{ji}$ = the ratio between the urban population weight in the province *j*, and the weight of urban population weight of the province *i*;

$MINORITY_{ji}$ = the ratio between the weight of the minority population weight from the province *j* and the weight of the minority population in the province *i*;

$WARM_{ji}$ = the ratio between the average yearly temperature in the capital of province *j* and the average yearly temperature in the capital of the province *i*;

ε_{ij} = residual value/error

Hypotheses of the model:

- the distance on the railroad and the migration rate are assumed to be in reverse proportion; the longer the distance, the more direct migration costs (for instance, the train tariff or the bus one, food stuff prices, accommodation on the road and on arrival), as well as the indirect migration costs (for instance, diminishment of incomes due to time loss before taking a new job, as well as the psychic costs of migration);
- the migration rate from the province *i* to province *j* is positive in relationship to the NETWORK variable. The higher the size of the migrants' community in the destination area, the lower the migration costs because there will be more information available to the area of origin regarding labour force employment, business opportunities, dwellings, schools, leisure opportunities, etc. Additionally, the lower will be also the psychic migration costs, because a large migrant community in the province of destination shall be a source of greater comfort, security and familiarity for those considering migration;
- the migration rate could be positively or negatively correlated with any type of investment expenditures in the province of destination as compared to the one of origin. On one hand, the increase of investments in the area of destination would generate a higher labour force demand and, implicitly the "pull" of migration to the province. In turn, higher investments in the province of origin would diminish the incentives to migrate, so that several attractive opportunities emerge on the local labour market. If the investments in the province of origin are higher and directed

towards infrastructure expenditures, especially roads, highways and public transportation, then the migration costs are diminished and thus, the migration to the province of destination would be encouraged;

- the relationship between the weight of labour force employment in the destination province in industry and the one in the province of origin is included in the model as a control mix for industry. The relationship between the labour force employment within the industry from the province of destination and the one of emigration should be positive. As result, the provinces with relatively large labour markets in the field of industry could attract more immigrants against the provinces where large agricultural sectors were developed;
- the provinces that are relatively more urbanised, have various jobs available, different living standards, etc., all influencing migration flows. For instance, the proportion of skilled jobs in more provinces with several urban zones is usually higher, which might encourage immigration more, because such jobs tend to be more attractive for potential migrants;

Within the realised estimates, the authors have postulated that the effects of minority migration could be either positive or negative:

i) this variable could impose the general political conditions of the province, for instance in provinces with a higher weight of minority population there could be several political divisions than in other provinces, and these can influence the migration models;
ii) there are several economic motives by which the weight of the minority population could influence migration: the provinces with a relatively higher weight of the minority population lack several of the basic industries, services and, as result, entrepreneurial migrants intending to start various businesses could find in these provinces profitable locations for starting-up their businesses (Bao, Hou and Shi (2006). On the other hand, migrants with superior qualification could be less interested to migrate to provinces with higher minority shares, because they can perceive these provinces as areas with limited possibilities of higher qualifications in the labour force employment.

- the migration rates are positively correlated to the real incomes from the province of destination (Y_{ji});
- the migration rate is assumed as positively correlated to the education level in the province of destination (E_j) because the existence of better educated labour force there usually means a wider distribution of employment opportunities for skilled individuals. Nevertheless, by using the same type of argument, a higher education level in the province of origin (E_i) is assumed to be in reverse proportion to the migration rate.

• at a higher unemployment rate in the province of destination (u_j) a discouragement of migration is expected, and at a higher rate of unemployment in the province of origin (u_i) it is expected an encouragement of migration;

• the yearly average temperature in the province of destination ($T_{ji}i$) is included as control on the differences in facilities. It is assumed that immigrants prefer warmer provinces.

Another version of the gravity model was the one realised by McFadden (1973). Introducing the chain notion in the random utility model variant of McFadden, in which the filter-decision of those living in location *i*, to opt for a living place in the location j , with $j = \overline{1, J}$ possibilities, was based on the expected utility of living in a certain location, for a period of time. The utility deriving from such a move can be written under the form $U_{ij}^{n}(Y_j, A_j, C_{i,j}, \mu_j^n)$, where Y_j indicate the incomes in location *j*, A_j are the facilities in *j*, $C_{j,i}$ are the relocation costs from the location *i* in *j*, and μ_j^n is the individual random earning. The probability P_{ij} to move from the region *j* in region *i* is:

$$P_{ij} = \Pr[U_{ij} = \max(U_{i1}, U_{i2}, ..., U_{iJ})] \quad (4)$$

McFadden's result indicates that if U_{ij} is linear in the argumentation logarithms, and the logarithm of the original stochastic disturbance is independent and identical with the Weibull distribution, then the probability to move from one place to the other is given by the relationship:

$$\mathrm{P_{ij}} = \frac{exp(\mathrm{U_{ij}})}{\sum_{\mathrm{k=1}}^{\mathrm{J}} exp(\mathrm{U_{ik}})} = \frac{exp(\alpha_1 \, ln\, \mathrm{A_j} + \alpha_2 \, ln\, \mathrm{Y_j} + \alpha_{31} \, ln\, \mathrm{C_{ij}})}{\sum_{\mathrm{k=1}}^{\mathrm{J}} exp(\alpha_1 \, ln\, \mathrm{A_k} + \alpha_2 \, ln\, \mathrm{Y_k} + \alpha_{31} \, ln\, \mathrm{C_{ik}})} \quad (5)$$

This expression of the migration probability provides an easier formula for the probability of moving from *i* to *j*, as compared with the one of remaining in *i*, a linear function depending on the differences of local characteristics:

$$\ln\left(\frac{P_{ij}}{P_{ii}}\right) = \alpha_1(\ln A_i - \ln A_j) + \alpha_2(\ln Y_i - \ln Y_j) + \alpha_3 C_{ij} \quad (6)$$

3. The study of the Romanian migration evolution with the help of the gravity model

For studying the Romanian migration was used a simplified gravity model, because the data provided by the Eurostat statistics, by the National Institute of Statistics from Romania (the Tempo-online databank), or by the Statistical Yearbooks of some EU countries where the Romanian migration is significant don't provide for complete data regarding all exogenous variables of the gravity model. As result, recourse was made to analysing the Romanian migration flows only in some EU countries, as well as the EU-27 level in the period 1995-2014.

The exogenous variables taken into account within the model were:

a. the GDP of each considered *i* country;
b. the employment rate corresponding to each *i* country;
c. the unemployment rate corresponding to each *i* country;
d. the volume of foreign direct investments of each *i* country;
e. the number of Romanian emigrants in each considered country and at EU-27 level;
f. the road (distance) between Romania and the various destinations of the Romanian migrants.

The description of the Romanian migration dependency with respect to **Spain** on the variables enumerated by means of the gravity model is presented in Table 1 from Annex 1.

The dependency degree of the endogenous variable on all independent variables of 98.25% indicates the fact that these are those which have a particular impact on the size of the migration phenomenon of Romanians to Spain.

Because the data sample has a volume lower than 30 observations, the *variable Student* $t_{\nu,\alpha}$ was used, the values of which were taken over from the table of the *Student* distribution, (depending on the value set for α and the number of liberty degrees ν). The analysis of the computed values for each of the determined estimators indicates for all that $t_{calc}>t_{theoretical}$ (a significance threshold of 5%), thus, these being significantly different from zero. Also the probabilities corresponding to each estimator support the idea that these can be accepted as being unbiased, consistent and efficient for the considered model.

The Durbin-Watson statistics, which is one of the most known procedures used for identifying the first rank self-correlation of errors within the linear regression models, has the value of 2.72, indicating the fact that the errors are not self-correlated.

For **Italy**, the real data about the number of Romanian migrants are unavailable in all international statistics. Therefore, the results obtained by applying the gravity model for studying the evolution of Romanian migration led to the results from Table 2 from Annex 1.

The dependency of Romanians' migration on all independent variables is of 74% which indicates the need of using also other exogenous variables in the gravity model.

The analysis of the computed values for each of the determined estimators indicates the fact that only for part of the estimators $t_{calc}>t_{theoretical}$ (these being thus significantly different from zero), while the others have lower values than the theoretical value, but for the majority this value being over 1.

The Durbin-Watson statistics has the value of 2.44, which indicates the fact that the errors are not self-correlated.

If the total number of Romanian migrants into the **European Union** is taken into account, then the use of the gravity model leads to obtaining the estimators presented in Table 3 from Annex 1.

The exogenous variables influence the evolution of the number of migrants at EU-27 level to a share of 91.5%, with a confidence threshold of 5%.

The comparison of the *Student* $t_{\nu,\alpha}$ *variable*, the values of which were taken over from the table of the *Student* distribution, with the computed ones, highlights for almost all estimators that the computed values are higher than the theoretical ones. This fact allows for stating that the estimators of the model are significantly different from zero.

The Durbin-Watson statistics which is one of the most known procedures used for identifying first rank self-correlation of the errors within the linear regression models has the value of 2.59, which leads to the conclusion that the errors are not self-correlated.

In conclusion, the use of the gravity model in estimating the number of migrants in general, and for the Romanians in particular may lead to significant results from the econometric viewpoint.

The variables taken into account in these estimates describe to a very good percent the variation in the number of Romanian migrants in each of the analysed countries.

Yet, these estimates can be substantially improved if, i) data were available for a longer period of time; ii) the information about the number of Romanian migrants would be available with more details; iii) also other exogenous variables would be introduced, for instance expenditures for R&D and I; iv) also qualitative variables would be introduced, which very often have a higher importance in taking the decision to migrate; v) "education and vocational training" would be taken into account as exogenous variables, etc.

ACKNOWLEDGMENTS

The work of Mariana BALAN is elaborated and published under the auspices of the Institute for the Research of the Quality of Life Romanian Academy, as part of the project co-financed by the European Union by the Sectorial Operational Programme Human Resources Development 2007−2013, within the Project Pluri-and Inter-disciplinarity in Doctoral and Post-Doctoral Programmes Project Code POSDRU/159/1.5/S/ 141086.

Selective Bibliography

1 Koser K., (2009), *The Impact of Financial Crises on International Migration: Lessons Learned*, IOM Migration Research Series, No. 37, Geneva, www.iom.int;

2 Lowry I., (1966), *Migration and metropolitan growth: two analytical models*, Chandler Publishing, Company, San Francisco;

3 McFadden, D. (1973), *Conditional Logit Analysis of Qualitative Choice Behavior*, in P. Zarembka, ed., Frontiers in Econometrics, Academic Press, New York;

4 Molho I., (1986), *Theories of migration: a review*, Scottish Journal of Political Economy, vol. 33;

5 Parrado E. A., Cerrutti. Marcela, (2007), *The Remittances of Parguayan Migrants to Argentina: Their Prevalence, Amount, and Utilization,* Integration & Trade (27):21;

6 Ratha D., Mohapatra S., Silwal A., (2009), *Migration and Development Brief,* Development Prospects Group, World Bank, www. siteresources.worldbank.org;

7 Shaw R.P., (1978), *Migration Theory and Fact: A Review and Bibliography of Current Lit*erature, Bibliography, Series No. 5. Regional Science Research Institute, Philadelphia, Pennsylvania;

8 Shuming Bao, Örn B. Bodvarsson, Jack W. Hou şi Yaohui Zhao, (2009). *The Deregulation of People Flows in China: Did the Structure of Migration Change?*, https://www.aeaweb.org/assa/.../retrieve.php;

*** *Anuarul Statistic al României 2007-2014*, Institutul Naţional de Statistică;

*** *International Migration Outlook, 2014*, OECD, www.oecd.org;

*** Immigration in the EU, Source: Eurostat 10/6/2015, www.ec.europa.eu/.../statistics... / Migration_and_migr...;

*** Statistics Eurostat, www.eurostat.eu;

*** *World Migration Report 2013,* Migrant Well-being and development, International Organization for Migration, http://www. publications.iom.int

Annex 1

Table 1

Dependent Variable: EMIGRSPANIA
Method: Least Squares
Sample(adjusted): 1995 2014
EMIGRSPANIA=C(1)+C(2)*PIBSP(-2)+C(3)*PIBROM+C(4)*RSOMAJSP+ +C(5)*RSOMAJRO+C(6)*ROCUPSP(-1)+C(7)*ROCUPROM+ + C(8)*DIST+C(9)*INVROMSP

	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	-21193.17	3015.838	-7.183357	0.0018
C(2)	0.529036	0.109806	4.398702	0.0108
C(3)	-0.564656	0.223916	2.296178	0.0480
C(4)	-60.29115	18.20178	-3.091431	0.0332
C(5)	189.5165	35.36077	5.267334	0.0058
C(6)	31.05385	41.45997	-1.672083	0.0438
C(7)	-212.2456	30.75947	7.322641	0.0019
C(8)	-408.2148	127.9892	-3.094232	0.0301
C(9)	-23517.34	6179.462	-4.110924	0.0135

R-squared	0.982518	Mean dependent var	302.1538
Adjusted R-squared	0.960553	S.D. dependent var	164.8579
S.E. of regression	31.91210	Akaike info criterion	9.969181
Sum squared resid	4170.977	Schwarz criterion	10.36030
Log likelihood	-56.79968	F-statistic	39.55644
Durbin-Watson stat	2.728524	Prob(F-statistic)	0.001520

Table 2

Dependent Variable: EMIGRITALIA
Method: Least Squares
Sample(adjusted): 1995 2010
EMIGRITALIA=C(1)+C(2)*PIBIT1+C(3)*PIBROM1+C(4)*ROCUPIT+ +C(5)*RSOMAJIT(-1)+C(6)*RSOMAJRO(-1)+C(7)*DIST+C(8)*INVROMIT(-1)

	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	-70924.30	29109.32	-2.409352	0.0428
C(2)	1.003226	0.843199	-1.101553	0.1748
C(3)	-3.049011	2.110208	1.444886	0.1086
C(4)	984.2977	352.7620	2.445918	0.0382
C(5)	-2105.978	926.3909	2.137953	0.0665
C(6)	324.7379	182.2585	-1.745110	0.1046
C(7)	-665.5840	615.6644	1.106152	0.2831
C(8)	2530.577	7109.025	0.305374	0.5345

R-squared	0.740703	Mean dependent var	1841.857
Adjusted R-squared	0.449857	S.D. dependent var	680.3415
S.E. of regression	500.0132	Akaike info criterion	15.56271
Sum squared resid	1500079.	Schwarz criterion	15.92788
Log likelihood	-100.9389	F-statistic	2.581098
Durbin-Watson stat	2.447208	Prob(F-statistic)	0.134212

Table 3

Dependent Variable: EMIGR
Method: Least Squares
Date: 04/22/12 Time: 12:59
Sample(adjusted): 1995 2010
EMIGR=C(1)+C(2)*PIBUE1+C(3)*PIBROM1+C(4)*RSOMAJUE+C(5)*RSOMAJRO
+C(6)*ROCUPUE(-3) +C(7)*ROCUPROM+C(8)*DIST

	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	-572480.5	186266.3	-3.073452	0.0372
C(2)	8.100699	2.751530	2.944071	0.0422
C(3)	-12.49417	6.346712	1.968605	0.1204
C(4)	-4957.699	1106.996	4.478514	0.0110
C(5)	1014.022	844.0468	1.201382	0.2959
C(6)	6914.966	2504.218	2.761328	0.0508
C(7)	-1703.000	435.2528	3.912669	0.0174
C(8)	-651.0388	3274.179	-0.198840	0.8521

R-squared	0.914846	Mean dependent var	11635.67
Adjusted R-squared	0.768577	S.D. dependent var	2842.375
S.E. of regression	1367.367	Akaike info criterion	17.51388
Sum squared resid	7478765.	Schwarz criterion	17.83715
Log likelihood	-97.08329	F-statistic	6.218851
Durbin-Watson stat	2.592207	Prob(F-statistic)	0.048314