PRZEGLĄD STATYSTYCZNY
R. LIX – ZESZYT 4 – 2012

MICHAŁ BERNARD PIETRZAK, MIROSŁAWA ŻUREK, STANISŁAW MATUSIK, JUSTYNA WILK

APPLICATION OF STRUCTURAL EQUATION MODELING FOR ANALYSING INTERNAL MIGRATION PHENOMENA IN POLAND

1. INTRODUCTION

Migration flows, or population movement to a temporary stay or permanent residence, is a natural occurrence that existed at every stage of human history. They result in a change in the size and structure of human resources both at the origin and the destination regions. The changes are not only demographic, but also have a social and economic character.

It must be noted that analysing the migration phenomenon is not an easy task. The existing research mostly consists in conducting indicator analyses, however, it appears not to be satisfactory due to the complicated nature of this occurrence. Migration moves occur in regional space. The size and direction of a population flow, its determining factors, including strength and direction of their influence and the role of a geographical factor, need to be considered.

The reasons for migration include a chain of circumstances and conditions which may be economic, social, political, cultural, environmental, etc. A typical example of population movement in Poland is connected with social and economic aspects, such as, for instance, taking up employment or education. Econometric gravity model plays a significant role in explaining the specificity of migration phenomenon. An important problem in analysing the migration flows with the use of the gravity model is the presence of a statistic correlation of explanatory variables. The correlation is mostly the result of high level aggregation of the analysed data, especially at a regional (*e.g.*, NUTS 2, NUTS 3) or higher level of territorial division.

The objective of the article is to propose another approach in analysis the migration phenomenon with the use of the Structural Equation Modeling (*SEM*) methodology and to test its usefulness in empirical research. Factor analysis and an econometric gravity model were applied within SEM. Based on this, we examined the determinants of the migration flows between subregions (NUTS 3) in Poland in the years 2008-2010.

The first part of the article describes the proposed research approach, the construction and interpretation of the gravity model as well as the *SEM* methodology. The second part is focused on the description of accepted research procedure and the scope of the research. The last part of this elaboration presents the results of the gravity model estimation based on the *SEM* methodology and some conclusions concerning the impact of the social and economic situation on interregional migrations.

2. RESEARCH APPROACH

The subject literature includes abundant works undertaking the problem of internal migrations and their determinants (Todaro, 1980; Lucas, 1997; Kupiszewski, Durham and Rees, 1999; Holzer, 2003; White, Lindstrom, 2006; Ghatak, Mulhern and Watson, 2008).

Econometric modeling plays a special role in research explaining the migration phenomenon understood in the context of social and economic conditions. The presented analyses usually consist in examining the interdependency between a given explanatory variable (*e.g.*, GDP, or the unemployment level) and the dependent variable represented by a specific migration coefficient (*e.g.*, net migration coefficient). Such an approach, however, is simplified and at the same time is an incomplete presentation of the migration phenomenon due to the following:

– migrations are a phenomenon characterised by flows,
– the cause of migration is usually some concrete situation dependent on a number of conditions,
– migrations occur in territorial space, therefore, it is significant to consider the geographic factor.

Considering migrations merely within an ratio approach is insufficient, since they occur in territorial space. Therefore, they should be considered as a flow from one region (the origin) to another region (the destination). However, the assumption is that every region, depending on the point of reference, may function as the origin region or the destination region. Besides, the migration phenomenon occurs in territorial space. Also, the role of the geographic factor must be taken into account (the geographical distance of the territorial units or the degree of their neighbourhood) especially when internal migrations are the subject of research.

The gravity model (spatial interaction model) is especially useful for migration research. The issues of the migration phenomenon and the use of gravity models were discussed by such authors as, for instance: Tinbergen (1962), Pöyhönen (1963), Chojnicki (1966), Anderson (1979), Grabiński et al. (1988), Sen, Smith (1995), Frankel, Stein and Weil (1995), Helliwell (1997), Deardorff (1998), Egger (2002), Baltagi, Egger and Pfaffermayr (2003), Anderson, van Wincoop (2004), Roy (2004), Serlenga, Shin (2007), Dańska-Borsiak (2007, 2008), Faustino, Leitão (2007), LeSage, Pace (2009), Kabir, Salim (2010), Lewandowska-Gwarda, Antczak(2010), Pietrzak, Wilk and Matusik (2012).

For years various conditions fostering migrations have been distinguished including political, economic, social, cultural, environmental and other (for example, see Pietrzak, Wilk and Matusik, 2012). At present in Poland, it is in the time of a free market economy, economic growth and development, the social and economic aspect as well as factors related to regional development are growing in importance. It must be noted, however, that the causes and the determinants are complex in nature. An element of the actualities should be considered as a reference point, such as, for instance, the current situation on the labour market, or the material situation in households and not a variable *sensu stricto*. It is necessary, therefore, to search for such solutions that would enable us to consider those aspects that cannot be measured directly. One of them is the application of synthetic measures constructed based on a set of observable primary variables with the use of factor analysis methods or taxonomic methods.

A vital problem in the analysis of migration flows based on econometric gravity models is the occurrence of statistical correlation of explanatory variables (see, for instance, Pietrzak, Wilk and Matusik, 2012). This correlation most frequently results from a high degree aggregation of the data analysed, in particular, at a regional level (e.g., NUTS 2, NUTS 3) or a higher one and from natural connection of the phenomena under scrutiny. In such a situation it becomes necessary to leave in the model only uncorrelated variables or to develop separate models for each explanatory variable. However, the first solution leads to loss of data due to the need to eliminate some variables, the other – within which each variable is treated separately – may result in over-interpretation of the results obtained. Yet another solution consists in taking so-called composite variables for explanatory variables. These composite variables are indicated by means of principal component analysis (see Welfe, 2009). The difficulty, however, is the uncertainty about the results obtained for estimated factors and their correct economic interpretation. For that reason it is proposed to apply confirmatory factor analysis within which the construction of the factor (a composite variable, a synthetic measure) is based on the researcher's knowledge and experience. It allows all the assumed complex variables to be considered in the construction of a factor and results in the number of factors assumed by the researcher.

Numerous researchers seek universal determinants of migration. In fact, however, the factors pulling inhabitants (*i.e.*, the pull factors) to a given region (the destination region) may differ from the factors pushing inhabitants (*i.e.*, the push factors) from the origin region, or they may be affecting but to a different degree. For instance, various pieces of research prove that a high level of economic development fosters both increased migration inflows and migration outflows, however, the difference is in the strength of the impact. This is yet another argument for the legitimacy of the application of the gravity model for research purposes. The model makes it possible to measure the strength of the impact of the examined explanatory variables on the push and pull effects in migration flows.

Due to the complex and multidimensional nature of the migration problems, an attempt was made to formulate an approach that would allow the specificity of the phenomenon to be treated in a complex way and to position it in the context of the economic situation. A solution proposal is to apply the *SEM* (*Structural Equation Modeling*) methodology with the consideration of confirmatory factor analysis and the econometric gravity model. The assumptions and procedures of modeling SEM can be found in the world literature (Goldberger, 1972; Jöreskog, Sörbom, 1979; Bollen, 1989; Kaplan, 2000; Pearl, 2000; Byrne, 2010; Kline, 2010) and in Polish (Gatnar, 2003; Strzała, 2006; Osińska, 2008; Matusik, 2008, Konarski, 2011, Osińska, Pietrzak and Żurek, 2011a; Osińska, Pietrzak and Żurek, 2011b).

The proposed research procedure entails tow steps. The first step includes a confirmatory factor analysis. Based on the set of primary explanatory variables, two composite variables are constructed (latent variables, factors) describing the economic and social situation in the origin and destination regions. In the second step a casual gravity model is developed in which interregional migration flows constitute a dependent variable and its realizations are the size of flow from the origin region to the destination region. The explanatory variables, in turn, are the two latent variables derived in the first step as well as the geographical distance between the regions.

The proposed approach also encompasses two other essential issues. It must be noted that empirical research discussed in the subject literature refers most frequently to a selected time period and the values of the dependent variable representing the migration phenomenon, originate from the same period as the values of the explanatory variables. However, advances in regional developments, changes in the economic situation or in living standards become visible after longer time intervals. Migrations, in turn, are a consequence of the occurrence of some event (*e.g.*, deterioration of the labour market) and once started take on a form of a lengthy process. For that reason the proposed approach considers migration flows in a determined time period and the causes of their occurrence are sought for in the preceding time periods.

Moreover, the research conducted using gravity models are most frequently based on the data characterised by a high degree of aggregation, *e.g.*, they encompass units of NUTS 2 or higher. Adopting such an approach it is possible to explain numerous phenomena occurring on a smaller scale. Lowering a territorial level gives a wider view of the migration phenomenon, *e.g.*, the NUTS 3 level allows metropolitan areas to seized. Polish subregions include such subregions whose borderlines are convergent with the areas of the largest Polish cities possessing the poviat rights (the examples include the cities of Warsaw and Wroclaw). The approach proposed promotes considering the lowest possible territorial units. Due to the fact that statistical data on migration flows are compiled based on the current and previous places of residence (legal domicile) and because major economic indicators (including GDP) are calculated at a regional level, the proposed reference point is the NUTS 3 regions.

3. CONSTRUCTION OF THE GRAVITY MODEL

The gravity model is frequently applied in researching the scope of a flows analysis resulting from the interaction of regions. The issue of the flows appears in such cases as, for instance, regional and international trade, transportation economics, or population migration research[1]. The model describes dependencies between the size of flow of a

[1] See Lewandowska-Gwarda and Antczak (2010).

given economic category and the conditions that influence fundamentally the formation of this phenomenon.

The flow applies to two regions, an origin region and a destination region. For that reason the processes that characterize both regions and the distance between them may be considered as explanatory variables. A physical distance is most frequently taken into account, however, it is also possible to identify such distances as economic, social and other.[2]

A typical gravity model[3] takes the following form (see Sen, Smith, 1996; LeSage, Pace, 2009)

$$\bar{Y} = \beta_o \cdot \bar{x}_{o1}^{\beta_{o1}} \cdot \bar{x}_{d1}^{\beta_{d1}} \cdot ... \cdot \bar{x}_{ok}^{\beta_{ok}} \bar{x}_{dk}^{\beta_{dk}} \cdot \bar{d}^{-\gamma} \cdot e^{\bar{\varepsilon}}. \tag{1}$$

After a logarithmic linearization the gravity model can be described using the formula below

$$\bar{Y}^* = \beta_0^* + X_o\bar{\beta}_o + X_d\bar{\beta}_d - \gamma\bar{d}^* + \bar{\varepsilon}, \tag{2}$$

$$X_o = [\ln \bar{x}_{o1},\ \ln \bar{x}_{o2}, ...,\ \ln \bar{x}_{ok}], \quad X_d = [\ln \bar{x}_{d1},\ \ln \bar{x}_{d2}, ...,\ \ln \bar{x}_{dk}], \tag{3}$$

$$\bar{\beta}_o = [\beta_{o1},\ \beta_{o2}, ...,\ \beta_{ok}]^{'}, \quad \bar{\beta}_d = [\beta_{d1},\ \beta_{d2}, ...,\ \beta_{dk}]^{'}, \tag{4}$$

where: $\bar{Y}$ – vector of flow values between regions, $\bar{Y}^* = \ln \bar{Y}$,

X_o, X_d – matrices of explanatory variables values for the origin regions and destination regions respectively,

$\bar{d}$ – vector containing distances for pairs of regions,

– constant,

$\bar{\beta}_o, \bar{\beta}_d \quad \gamma$ – structural parameters of the model,

$\bar{\varepsilon}$ – vector of disturbances.

A square matrix of flows $Y_{[nxn]}$ is a starting point during the gravity model construction process. The size of the matrix is $n \times n$, where n columns refers to origin regions and n rows refers to destination regions. Each element y_{ij} of the matrix means a value of flow from the origin region i into the destination region j ($i, j = 1, 2, ..., n$, n – number of regions). On the main diagonal there are intraregional flows for the region. They can be either included or excluded depending on the purpose of research.

[2] The problem of choosing the distance between regions was discussed by Grabiński (1991).

[3] In a wider range the construction of the model can also differentiate interregional and intraregional flows as well as occurring spatial effects (for details see LeSage, Pace, 2009).

The $Y_{[nxn]}$ matrix is transformed to the destination-centric ordering vector $\bar{Y}$ that takes the following form

$$\bar{Y} = \begin{bmatrix} y_{11} \\ . \\ y_{n1} \\ . \\ y_{1n} \\ . \\ y_{nn} \end{bmatrix} = \begin{bmatrix} y_{1} \\ . \\ y_{n} \\ . \\ y_{n^2-n+1} \\ . \\ y_{n^2} \end{bmatrix}_{[n^2x1]} . \quad (5)$$

Destination-centric ordering means that all flows for the primary destination region are being arranged first, next for the second and subsequently for all other regions (see LeSage, Pace, 2009). For the vector of a dependent variable created in that way the matrices of explanatory variables X_o, X_d and a vector containing distances $\bar{d}$ are developed in such a way that the indices match the indices of the vector $\bar{Y}$. The vector $\bar{d}$ containing distances between regions is being indexed in the same way as the vector $\bar{Y}$. For the set of explanatory variables two matrices are being created: the matrix X_o containing values of explanatory variables in origin regions and the matrix X_d containing values in destination regions. We write down the matrices X_o and X_d as

$$X_o = \begin{bmatrix} \ln x_{11o} & \ln x_{12o} & . & \ln x_{1ko} \\ . & . & . & . \\ \ln x_{n1o} & \ln x_{n2o} & . & \ln x_{nko} \\ . & . & . & . \\ \ln x_{l1o} & \ln x_{l2o} & . & \ln x_{lko} \\ . & . & . & . \\ \ln x_{n^21o} & \ln x_{n^22o} & . & \ln x_{n^2ko} \end{bmatrix}_{[n^2xk]} , \; X_d = \begin{bmatrix} \ln x_{11d} & \ln x_{12d} & . & \ln x_{1kd} \\ . & . & . & . \\ \ln x_{n1d} & \ln x_{n2d} & . & \ln x_{nkd} \\ . & . & . & . \\ \ln x_{l1d} & \ln x_{l2d} & . & \ln x_{lkd} \\ . & . & . & . \\ \ln x_{n^21d} & \ln x_{n^22d} & . & \ln x_{n^2kd} \end{bmatrix}_{[n^2xk]} , \quad (6)$$

where for any element x_{ij} the index i denotes the number of a region and the index j is the number of an explanatory process, $l = n^2 - n+1$, and k represents the number of explanatory variables.

4. CONCEPT OF THE SEM MODEL

The *SEM* (*Structural Equation Modeling)* model is the effect of a merger between confirmatory factor analysis and the causal models built on its base. The application of the confirmatory factor analysis[4] in an econometric model allows the latent variables

[4] More on confirmatory factor analysis see, for instance, Kim, Mueller (1978), Brown (2006), Hair et al. (1998).

to be considered. These variables may constitute a form of reducing description in the case when the number of variables is high. However, relevant to the research objective set in the present work, it was intended to derive a factor which in a synthetic way[5] would be representing a certain complex phenomenon (criterion) that is not subject to direct measures, such as, for example, the economic situation in a given region.

The variables are not measurable in a direct way as they represent a set of observed variables strongly correlated with each other. They are hypothetic conceptual structures (*e.g.* the socio-economic situation of a region) and they are being treated as potential determinants that impact the economic category that is under consideration.

The SEM model includes the internal model that describes causal dependencies and the external model that is being used to measure endogenous and exogenous latent variables (see Bollen, 1989; Kaplan, 2000; Pearl, 2000; Gatnar, 2003; Osińska, 2008; Konarski, 2011). In other words, the internal model allows causal dependencies between the endogenous and exogenous variables to be specified, including the influence of the latent variables. It takes the following form

$$\bar{\eta} = B\bar{\eta} + \Gamma\bar{\xi} + \bar{\zeta}, \tag{7}$$

where: $\bar{\eta}$ – vector of endogenous latent variables,
$\bar{\xi}$ – vector of exogenous latent variables,
B – matrix of regression coefficients at endogenous variables,
Γ – matrix of coefficients at exogenous variables,
$\bar{\zeta}$ – vector of disturbances.

The external model that is described as the model of measurement is a representation of results of the confirmatory factor analysis that allows for calculation of factor loadings for the variables that form the latent variable (factor). The model is given as

$$\bar{Y} = \Pi_y \,\bar{\eta} + \bar{\varepsilon}, \tag{8}$$

$$\bar{X} = \Pi_x \,\bar{\xi} + \bar{\delta}, \tag{9}$$

where: $\bar{Y}$ – the vector of observed endogenous variables,
$\bar{X}$ – the vector of observed exogenous variables,
Π_y, Π_x – matrices of factor loadings,
$\bar{\varepsilon}, \bar{\delta}$ – vectors of measurement errors.

5. THE SEM MODEL IN THE ANALYSIS OF INTERREGIONAL MIGRATIONS IN POLAND

The objective of the conducted study is to analyze the impact of socio-economic situation on the phenomenon of internal migrations in Poland. The adopted units of

[5] More on constructing synthetic measures see, e.g., *Handbook...* (2008).

reference were the areas of 66 Polish sub-regions (NUTS 3)[6]. The study was performed for the period of 2008-2010. The accepted time interval coincides with the period of global financial and economic crisis which was of significant influence on the slowdown of economic growth in Poland and the stabilization of migration trends. The study focused on inter-regional[7] migrations related to permanent residence.

The concept of structural equations models was applied in the study. The construction of *SEM* model covered the establishment of econometric gravity model based on the results of confirmatory factor analysis. The aggregate size of migration flows between sub-regions, in the period of 2008-2010, was adopted as dependent variable of the gravity model. The variable values illustrate the inflow volume from the origin region (previous place of permanent residence) to the destination region (present place of permanent residence). The factors illustrating socio-economic situation (latent variables), in the origin regions and the destination ones as well as the geographical distance between the centroids of sub-regions, were accepted as explanatory variables in the discussed model.

In Poland, relatively large regional differences in terms of socio-economic situation are observed. It is manifested by considerable disparities between the level of productivity, the scale of entrepreneurship and the tendency to invest, which determine regional economic potential, the absorption capacity of labour market and the related social problems, including unemployment, and also the level of salaries which affects the local population living standards and the possibility of the offered goods and services consumption.

These phenomena are of complex nature which results in difficulties – and in fact makes it impossible - to describe and measure them from a holistic perspective. Therefore, in the framework of the conducted study, a set of variables was prepared to be used to illustrate the socio-economic situation referring to sub-regions. The data collected cover 2008 which is considered to be the time when Poland began to feel the effects of the global economic crisis.

Table 1 presents the final set of variables which were obtained after conducting substantive and formal selection, having in mind that the factor responsible variables have to present certain qualities, such as, among others, strong statistical correlation.

---

[6] In 2008, due to substantive incorrectness of the division of Poland into 45 sub-regions (see: Bąk et al., 2009), their number was increased up to 66 sub-regions. The previous division resulted in the fact that the separated territorial units, owing to significant disproportions, among others in the population number, were incomparable and could not be referred to as units of the same level in the territorial division.

[7] The population flow within sub-regions, which at the same time are the metropolitan cities (e.g. Warsaw, Wrocław) are not considered as migrations, since they do not involve crossing administrative borders of a territorial unit, but only the movement of population between a particular city quarters. In such cases the value of intra-regional migration flow is 0. This situation results in considerable difficulties in the structure and interpretation of the gravity model. For this reason the hereby paper does not discuss migrations within the sub-regions (intra-regional migrations).

Table 1.

Set of the observed variables applied in the analysis of internal migrations

Variables	Description	Accepted measurement unit
X_{1o}, X_{1d}	Gross Domestic Product *per capita*	1 000 PLN
X_{2o}, X_{2d}	Number of entities in the national economy per one hundred citizens	unit
X_{3o}, X_{3d}	Investment outlays on fixed assets *per capita*	100 PLN
X_{4o}, X_{4d}	Fixed assets in enterprises *per capita*	100 PLN
X_{5o}, X_{5d}	Registered unemployment rate	%
X_{6o}, X_{6d}	Average monthly gross payment in the national economy	1 000 PLN

Source: Authors' own calculations.

Including the latent factor for the origin and destination regions, the internal model, referred to as the structural model, has been stated as follows

$$\bar{Y}^* = \beta_0^* + \beta_o\bar{\eta}_o + \beta_d\bar{\eta}_d - \gamma\bar{d}^* + \bar{\zeta}, \tag{10}$$

where: $\bar{Y}^* = \ln \bar{Y}$ ($\bar{Y}$ is the vector of flows),

β_o, β_d – regression coefficients at latent variables,

$\bar{d}$– vector containing distances for pairs of regions,

$\bar{\eta}_o, \bar{\eta}_d$ – latent variables, identified based on the X_o, X_d matrices,

ξ – vector of disturbances.

The external model was described as below

$$X_o = \pi_o\eta_o + \varepsilon_o, \tag{11}$$

$$X_d = \pi_d\eta_d + \varepsilon_d, \tag{12}$$

$$X_o = \begin{bmatrix} x_{1o} \\ x_{2o} \\ x_{3o} \\ x_{4o} \\ x_{5o} \\ x_{6o} \end{bmatrix} = \begin{bmatrix} \ln X_{1o} \\ \ln X_{2o} \\ \ln X_{3o} \\ \ln X_{4o} \\ \ln X_{5o} \\ \ln X_{6o} \end{bmatrix}, \quad X_o = \begin{bmatrix} x_{1d} \\ x_{2d} \\ x_{3d} \\ x_{4d} \\ x_{5d} \\ x_{6d} \end{bmatrix} = \begin{bmatrix} \ln X_{1d} \\ \ln X_{2d} \\ \ln X_{3d} \\ \ln X_{4d} \\ \ln X_{5d} \\ \ln X_{6d} \end{bmatrix}, \tag{13}$$

where: π_o, π_d – factor loadings,

$\varepsilon_o, \varepsilon_d$ – measurement error.

In the case of the registered unemployment rate, a ratio scaled transformation of the X_{5o} and X_{5d} variables has been performed (from a destimulant into a stimulant), so all of the observed variables being parts of the factor were to affect in the same direction. Then, with the confirmatory factor analysis two latent variables (η_o and

η_d) have been extracted for the origin and the destination regions respectively. They reflect the social and economic situation of the subregions. For both factors the sets of observed variables are being characterized by the Cronbach's Alpha at an acceptable level which are 0.90 and 0.91 respectively. This provides coherence and a correct level of description of a complex form of latent variables (see Cronbach, 1951).

Both of the η_o and η_d factors and also a geographical distance have been considered as the explanatory variables in the gravity model that has been described as a structural equations model. The model shown in Figure 1 is also a graphical interpretation of equations 10-12. The assumed model has been estimated by using of a maximum likelihood method, while its quality and level of adjustment have been evaluated by means of the measures proposed by the subject literature.

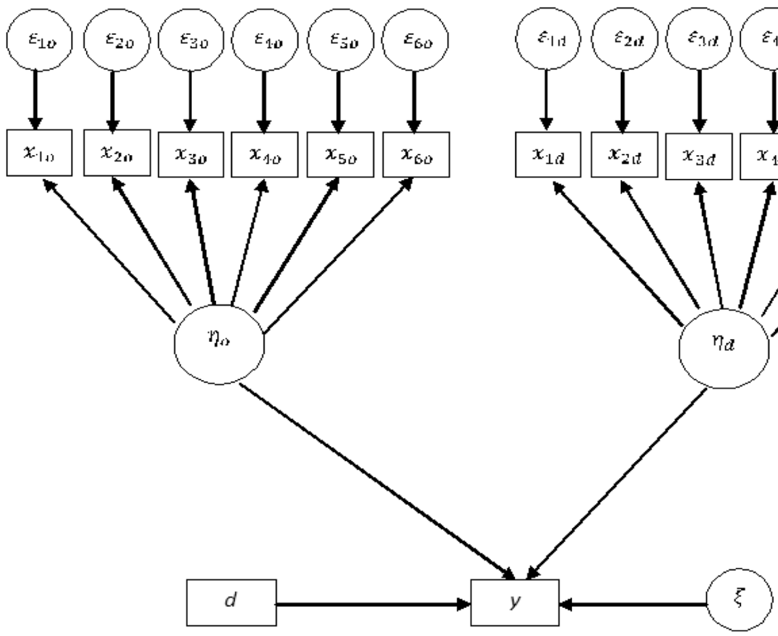

Figure 1. Hypothetical gravity model for interregional migrations in Poland
y– dependent variable (value of interregional migration flows from origin regions into destination regions), x_{mo}, x_{md} – primary variables that create the factors, for the origin and destination regions respectively, n_o, n_d – latent variables (factors) that represent the socio-economic situation of regions, d – geographical distance.

Source: elaborated by the authors

Table 2 presents the results of the confirmatory factor analysis made for the considered gravity model. The first column contains dependencies between the observed variables and the latent variable for the confirmatory factor analysis. The second column contains names of the parameter for a specific relation. In the third one its estimates are shown, in the next – standardized values that are also factor loadings and the last of the columns presents the p value that allows the statistical significance of the parameters to be evaluated. The results from the external model indicate that all factor loadings are statistically significant. All of the 'standardised' values contained in Table 2 are higher than 0.7. This confirms a proper identification of the latent variables.

Table 2.

Results of the confirmatory factor analysis

Dependencies	Parameter	Estimate	Standardized value	p-value
$x_{1o} \leftarrow \eta_o$	Π_{1o}	0.306	0.999	<0.001
$x_{2o} \leftarrow \eta_o$	Π_{2o}	0.178	0.783	<0.001
$x_{3o} \leftarrow \eta_o$	Π_{3o}	0.455	0.910	<0.001
$x_{4o} \leftarrow \eta_o$	Π_{4o}	0.337	0.891	<0.001
$x_{5o} \leftarrow \eta_o$	Π_{5o}	0.111	0.840	<0.001
$x_{6o} \leftarrow \eta_o$	Π_{6o}	0.111	0.882	<0.001
$x_{1d} \leftarrow \eta_d$	Π_{1d}	0.305	0.998	<0.001
$x_{2d} \leftarrow \eta_d$	Π_{2d}	0.178	0.783	<0.001
$x_{3d} \leftarrow \eta_d$	Π_{3d}	0.456	0.912	<0.001
$x_{4d} \leftarrow \eta_d$	Π_{4d}	0.338	0.892	<0.001
$x_{5d} \leftarrow \eta_d$	Π_{5d}	0.111	0.840	<0.001
$x_{6d} \leftarrow \eta_d$	Π_{6d}	0.111	0.883	<0.001

Source: Authors' own calculations.

Table 3 shows the results of the internal model. The arrangement of columns in the table is exactly the same as in Table 2. The third column is a key column, because it contains the estimates of the parameters of the analysed model that indicate the strength and direction in the case of the influence of specific factors on the dependent variable, which represents the value of the migration flows[8]. All the parameters of the internal model are statistically significant. This indicates the influence of the geographical distance and the social and economic situation in regions for intensity of migrations. The value of the IFI coefficient at the level of 0.937 indicates a good adjustment of the model to the input data. The only reason of concern may be the value of the RMSEA indicator at the level of 0.114 that is higher than the suggested 0.08. However, in the case of this kind of macro-economic data, the value of this index can be assessed as admissible, which does not disqualify the model construction and its interpretation.

[8] It was assumed that for the both factors the variation equals one, the received evaluations of the parameters can be used to estimate the strength of the influence of the factors. Identical conclusions about the influence strength will be achieved based on the analysis of standardized parameters evaluation.

Table 3.

Results of the estimation of the internal model

Causal dependencies	Parameter	Estimate	Standardized value	p-value
$y \leftarrow \eta_o$	β_o	0.117	0.095	<0.001
$y \leftarrow \eta_d$	β_d	0.327	0.267	<0.001
$y \leftarrow d$	γ	-1.394	-0.686	<0.001
Quality assessment of model	IFI	0.937	RMSEA	0.114

Source: Authors' own calculations.

6. INFLUENCE OF THE SOCIAL AND ECONOMIC SITUATION ON MIGRATION FLOWS

As compared to other OECD countries, the population of Poland is featured by a relatively low scale of territorial mobility. This situation has further deepened during the global economic crisis during which a significant decrease in domestic migration intensity was observed. Therefore one may assume that, in this period, the socio-economic situation can play a particularly important role in the development of inter-regional migration, just as the problem of geographical distance.

The results of the estimation that are shown in Tables 2 and 3 prove that both the geographical distance and the social and economic situation of a given area are significant determinants that describe the size and directions of the internal migrations in Poland. A negative sign for a distance parameter means decreasing the migration flow level between subregions with as it grows. Population most often moves into the subregions that are not so far away from their primary places of residence. It can, therefore, be assumed that the most intensive migration movements are occurring between the sub-regions characterized by a relatively high degree of neighbourhood[9].

However, positive evaluations of the latent variables indicate that improvement of the social and economic situation has impact either on increasing the push effect from origin regions and the pull effect of destination regions. Comparing the standardized values of parameters $\beta_o = 0.095$ and $\beta_d = 0.267$, a much bigger influence of the pulling than pushing factors can be noted.

An especially interesting fact is the influence, in the form of pushing from origin regions, which indicates that a higher migration outflow characterizes regions that are better developed socially and economically. It is, therefore, a low level of a social and economic development that is the factor that slows down the migration movements, especially during the economic crisis that is being considered in the research.

In the context of the gravity model's interpretation, there is an important difference between parameter estimates of the origin and destination regions. Flows between the regions have a two-way character, due to that fact the two chosen areas are both origin

[9] The two areas are adjacent in n-th order if the need arises to cross at least n administrative borders in order to move from one to the other (see: Lewandowska-Gwarda, Olejnik, 2010). Direct neighbourhood means that two territorial units (e.g. sub-regions) share a common administrative border.

and destination at the same time. Therefore, a positive difference of evaluations means a higher level of flows towards the region with a higher value of the explanatory variables and, in the case of migration, its positive balance. On the other hand, a negative difference of evaluations means a higher level of flows towards the region with a lover value of the explanatory variables.

A positive difference of the parameters β_d and β_o equals 0.210 which means in this case that existing migration moves will lead to positive net migration for the most developed regions. The regions with a better social and economic situation will have higher migration inflows of population than outflows. An exactly opposite situation will occur in less developed regions.

7. CONCLUSIONS

The objective of the paper was to analyse the impact of the socio-economic situation on the intensity of interregional (NUTS 3) migrations in Poland in the time period 2008-2010. The article contains the analysis of dependencies for the migration phenomenon that had been conducted with the use of a spatial gravity model that was defined as a part of Structural Equation Modeling procedure.

The interregional migrations have been researched where the latent variables (for origin regions as well as for destination regions) that describe the socio-economic situation of subregions and their geographical distance were assumed as explanatory variables. The set of observed primary variables, which are indicators of the latent variables, included the value of GDP *per capita*, fixed assets and investments *per capita*, the number of economic entities, the registered unemployment ratio, and wages and salaries level. Based on the estimation results of the model taken, the evaluation of the influence of latent variables on a direction and intensity of migration flows was also taken into account. Moreover, the role of the distances between the regions was considered.

The achieved results have confirmed the influence of the socio-economic situation and the geographical distance on the interregional migration flows in Poland. Also, the interpretation of the parameters obtained indicated that subregions with a relatively good economic situation are the areas where population inflows and this situation will contribute to increasing their positive net migration.

Due to the fact that mainly citizens with a proper potential migrate into better developed subregions in a socio-economic sense, it can be presumed that under-developed regions constitute a kind of a ‘demographic base’ for better developed subregions. Such a situation may impact less developed subregions by leaving them with fewer valuable human resources. This is one of the factors that consolidate an economic and territorial divergence in Poland.

On the other hand, a greater tendency for migration oriented travelling is demonstrated by the residents of regions featured by a relatively more favourable socio-economic situation than that of the weaker ones. It can be assumed that the sub-regions which are referred to as cities with district rights, or which cover townships, are characterized by a better situation. Positive economic conditions of a particular region are translated into a more favourable material and financial situation of its inhabitants. For this reason, they may perceive the decision to change the place of residence as less risky than those living in the regions lagging behind.

The issue of geographical distance is of significance too. The relatively low territorial mobility of Polish people is manifested not only in the low intensity of migration flows, but also in the coverage of relatively short distances. Therefore, the relationship between the flow volume and the distance is negative. One can also put forward the assumption that a high level of sub-regional adjacency is likely to enhance mutual migration flows.

The research approached applied enabled us to outline migration processes in Poland and reveal potential determinants of migration flows, which may constitute a reference point for further research in this area. Incorporating intra-regional migration flows in further research persist an open problem. Particularly difficult, in this respect, is considering sub-regions, which are also the largest Polish cities with the rights of a district. Flows within these sub-regions are not related to crossing a territorial unit administrative borders and, in accordance with the Central Statistical Office methodology, they are not referred to as migration phenomena. In this case, one of the acceptable solution is to exclude in the scope of the study these types of sub-regions, but this may distort the due analyses results and cause difficulties in their interpretation.

One of the subsequent research directions could become the analysis of suburbanization phenomenon, the determination of relationship between the intensity of domestic migrations (especially intra-regional flows) and the volume of international migration, as well as an attempt to apply the suggested approach with using of Structural Equation Modeling in the study of international migration determinants.

Michał Bernard Pietrzak, Mirosława Żurek – Nicolaus Copernicus University in Toruń
Stanisław Matusik – University School of Physical Education, Cracow
Justyna Wilk – Wroclaw University of Economics

LITERATURA

[1] Anderson J.E., (1979), *A Theoretical Foundation for the Gravity Model*, American Economic Review, 69 (1), 106-116.

[2] Anderson J.E., Van Wincoop E., (2004), *Trade Costs*, Journal of Economic Literature, 42(3), 691-751.

[3] Bąk A., Chmielewski R., Piotrowska M., Szymborska A., Ślesik M. (2009), *Zróżnicowania społeczno-gospodarcze w Polsce. Porównanie podziału na 45 i 66 podregionów*, Przegląd Regionalny nr 3, Stan wdrażania programów operacyjnych realizowanych w latach 2007-2013. Perspektywa regionalna, Ministerstwo Rozwoju Regionalnego, Warszawa.

[4] Baltagi B.H., Egger P., Pfaffermayr M., (2003), *A Generalized Design for Bilateral Trade Flow Models*, Economics Letters, 80, 391-397.

[5] Bollen K.A., (1989), *Structural Equations with Latent Variables*, Wiley.
[6] Brown T.A., (2006), *Confirmatory Factor Analysis for Applied Research*, Guilford Press.
[7] Byrne B.M., (2010), *Structural Equation Modeling with AMOS: Basic Concepts*, Applications and Programming (2nd ed.), Routledge/Taylor & Francis, New York.
[8] Chojnicki Z., (1966), *Zastosowanie modeli grawitacji i potencjału w badaniach przestrzenno-ekonomicznych*, PWN, Warszawa.
[9] Cronbach L.J., (1951), *Coefficient Alpha and the Internal Structure of Tests*, Psychometrika, 16 (3), 297-334.
[10] Dańska-Borsiak B., (2007), *Migracje międzywojewódzkie ludności a działalność badawczo-rozwojowa w województwach (zastosowanie modeli grawitacji)*, Wiadomości Statystyczne no. 5 (552), 53-66.
[11] Dańska-Borsiak B., (2008), *Zróżnicowanie poziomu rozwoju gospodarczego województw w Polsce a wielkość migracji międzywojewódzkich*, Taksonomia 15. Klasyfikacja i analiza danych – teoria i zastosowania, PN UE we Wrocławiu no. 7(1207), 364-370.
[12] Deardorff A., (1998), *Determinants of Bilateral Trade: Does Gravity Work in a Neoclassical World? NBER Chapters*, The Regionalization of the World Economy, 7-32, National Bureau of Economic Research, Washington DC.
[13] Egger P., (2002), *An Econometric View on the Estimation of Gravity Models and the Calculation of Trade Potentials*, World Economy, 25, 297-312.
[14] Faustino H., Leitão N.C., (2007), *Intra-Industry Trade: A Static and Dynamic Panel Data Analysis*, International Advances in Economic Research, 13 (3), 313-333.
[15] Frankel J., Stein E., Wei S.J., (1995), *Trading Blocs and the Americas: The Natural, the Unnatural and the Super-Natural*, Journal of Development Economics 47, 61-95.
[16] Gatnar E., (2003), *Statystyczne modele struktury przyczynowej zjawisk ekonomicznych*, Wydawnictwo Akademii Ekonomicznej w Katowicach, Katowice.
[17] Ghatak S., Mulhern A., Watson J., (2008), *Inter-regional Migration in Transition Economies. The Case of Poland*, Review of Development Economics, 12(1), Oxford, p. 209-222.
[18] Goldberger A.S., (1972), *Structural Equation Models in the Social Sciences*, Econometrica, 40(2), 979-1001.
[19] Grabiński T., (1991), *Modele generalizacji przestrzennej*, In: Zeliaś A. (Ed.), (1991), Ekonometria przestrzenna, PWE, Warszawa.
[20] Grabiński T., Malina A., Wydymus S., Zeliaś A., (1988), *Metody statystyki międzynarodowej*, PWE, Warszawa.
[21] Hair J.F., Anderson R.E., Tatham R.L., Black W.C., (1998), *Multivariate Data Analysis with Readings*, Macmillan Publishing Company, New York.
[22] Handbook on Constructing Composite Indicators. Methodology and User Guide, OECD, Paryż 2008.
[23] Helliwell J., (1997), *National Borders, Trade and Migration*, Pacific Economic Review 2, 165-185.
[24] Holzer J.Z., (2003), *Demografia*, PWE, Warszawa.
[25] Jöreskog K.G., Sörbom D., (1979), *Advances in Factor Analysis and Structural Equation Models*, New York: University Press of America.
[26] Kabir M., Salim R., (2010), *Can Gravity Model Explain BIMSTEC'S Trade?*, Journal of Economic Integration , 25(1), 144-166.
[27] Kaplan D., (2000), *Structural Equation Modeling: Foundations and Extensions*, Sage Publications.
[28] Kim J.O., Mueller C.W., (1978), *Factor Analysis. Statistical Methods and Practical Issues*, Sage, Beverly Hills.
[29] Kline R.B., (2010), *Principles and Practice of Structural Equation Modeling*, Guilford Press.
[30] Konarski R., (2011), *Modele równań strukturalnych. Teoria i praktyka*, PWN, Warszawa.

[31] Kupiszewski M., Durham H., Rees P., (1999), *Internal Migration and Regional Population Dynamics in Europe: Poland Case Study*, W: P. Rees, M. Kupiszewski, Internal Migration and Regional Population Dynamics in Europe: A Synthesis, Collection Demography, Council of Europe, Strasbourg.
[32] LeSage J.P., Pace R.K., (2009), *Introduction to Spatial Econometrics*, CRC Press, New York.
[33] Lewandowska-Gwarda K., Antczak E., (2010), *Wprowadzenie do przestrzennych analiz ekonomicznych*, In: Suchecki B. (ed.), Ekonometria przestrzenna, Wydawnictwo C. H. Beck, Warszawa.
[34] Lucas R., (1997), *Internal Migration in Developing Countries*, In: M.R. Rosenzweig, 0. Stark (Ed.), Handbook of Population and Family Economics, Elsevier Science B.V, Amsterdam, s. 721-798.
[35] Matusik S., (2008), *Zastosowanie SEM w analizie czynników określających poziom rozwoju społeczno-gospodarczego na przykładzie gmin woj. małopolskiego*, In: Z. Zieliński (Ed.), Współczesne trendy w ekonometrii, Wyższa Szkoła Informatyki i Ekonomii TWP, Olsztyn.
[36] Osińska M., (2008), *Ekonometryczna analiza zależności przyczynowych*, Uniwersytet Mikołaja Kopernika, Toruń.
[37] Osińska M., Pietrzak M.B., Żurek M., (2011a), *Ocena wpływu czynników behawioralnych i rynkowych na postawy inwestorów indywidualnych na polskim rynku kapitałowym*, Przegląd Statystyczny, 58 (3-4).
[38] Osińska M., Pietrzak M.B., Żurek M., (2011b), *Wykorzystanie modeli równań strukturalnych do opisu psychologicznych mechanizmów podejmowania decyzji na rynku kapitałowym*, Acta Universitatis Nicolai Copernici – Ekonomia, 42 (403), 7-21.
[39] Pearl J., (2000), *Causality. Models*, Reasoning and Inference, Cambridge.
[40] Pietrzak M., Wilk J., Matusik S., (2012), *Gravity Model as the Tool for Internal Migration Analysis in Poland in 2004-2010*, Quantitative Methods for Modelling and Forecasting Economic Processes, Wydawnictwo UE w Krakowie, Kraków, in print.
[41] Pöyhönen P., (1963), *A Tentative Model for the Volume of Trade between Countries*, Weltwirtschaftliches Archiv, 90 (1), 93-99.
[42] Roy J.R., (2004), *Spatial Interaction Modeling: A Regional Science Context*, Berlin: Springer-Verlag.
[43] Sen A., Smith T.E., (1995), *Gravity Models of Spatial Interaction Behavior*, Springer, Berlin-Heilderberg-New York.
[44] Serlenga L., Shin Y., (2007), *Gravity Models of Intra-EU Trade: Application of the Hausman-Taylor Estimation in Heterogeneous Panels with Unobserved Common Time- Specific Factors*, Journal of Applied Econometrics, 22, 361-81.
[45] Strzała K., (2006), *Modelowanie równań strukturalnych: koncepcja i podstawowe pojęcia*, Prace i Materiały Wydziału Zarządzania Uniwersytetu Gdańskiego, 2, 121-130.
[46] Suchecki B., Lewandowska-Gwarda K., Olejnik A., (2010), *Wprowadzenie do przestrzennych analiz ekonomicznych*, In: Suchecki B. (Ed.), Ekonometria przestrzenna, Wydawnictwo C. H. Beck, Warszawa.
[47] Tinbergen J., (1962), *Shaping the World Economy*, The Twentieth Century Fund Inc., New York.
[48] Todaro M., (1980), *Internal Migration in Developing Countries*. A survey, In: R.A. Easterlin, Population and Economic Change in Developing Countries, University of Chicago Press, Chicago, 361-402.
[49] Welfe A., (2009), *Ekonometria. Metody i ich zastosowanie*, Polskie Wydawnictwo Ekonomiczne, Warszawa.
[50] White M.J., Lindstrom D.P., (2006), *Internal Migration*, In: D.L. Poston, M. Micklin (Eds.), Handbook of population, Springer, Berlin-Heilderberg.

WYKORZYSTANIE MODELU RÓWNAŃ STRUKTURALNYCH DO OPISU MIGRACJI WEWNĘTRZNYCH W POLSCE

S t r e s z c z e n i e

Migracje kształtują wielkość i strukturę zasobów ludzkich w regionach, mają więc wymiar nie tylko demograficzny, ale także społeczno-gospodarczy. Prowadzone badania obejmujące najczęściej analizę wskaźnikową, są często niewystarczające ze względu na przestrzenno-czasowy i wielowymiarowy charakter tego zjawiska.

Celem artykułu jest propozycja zastosowania w analizie migracji metodologii *SEM* z uwzględnieniem modelu grawitacji. Takie podejście pozwala opisać sytuację społeczno-ekonomiczną regionów z wykorzystaniem zmiennych skorelowanych, a także określić jej wpływ na przepływy migracyjne.

W pierwszej części artykułu omówiona jest konstrukcja i interpretacja modelu grawitacji oraz koncepcja modelu *SEM*. W drugiej części omówiono przyjętą procedurę badawczą oraz zakres badań. W ostatniej części zaprezentowano wyniki prowadzonej analizy. Określono siłę i kierunek oddziaływania poziomu rozwoju społeczno-ekonomicznego na przepływy migracyjne między podregionami (NUTS 3) w Polsce w latach 2008-2010.

Słowa kluczowe: migracje wewnętrzne, model grawitacji, modelowanie *SEM*

APPLICATION OF STRUCTURAL EQUATION MODELING FOR ANALYSING INTERNAL MIGRATION PHENOMENA IN POLAND

A b s t r a c t

Migrations form the size and structure of human resources in regions. For that reason they have not only a demographic but also a social and economic character. The existing research is often not enough due to a spatial, temporary and also a multidimensional character of this occurrence.

The subject of this paper is to apply Structural Equation Modeling (*SEM*) concept for analysing the population migration phenomena considering the gravity model. This approach allows describing the social and economic situation in regions considering a set of correlated variables and it also allows determining its influence on migration flows.

The first part of the article describes the construction and interpretation of the gravity model and the concept of the *SEM* model. In the second part the adopted research procedure and the research scope are presented. In the last part, the results of the investigation are explained. The strength and direction of the impact of the socio-economic development on the migration flows between subregions (NUTS 3) in Poland in the years 2008-2010 are discussed.

Key words: internal migrations, gravity model, Structural Equation Modeling