ACTA FACULTATIS RERUM NATURALIUM UNIVERSITATIS COMENIANAE
Geographica Supplementum No 2/I, 1999, pp. 11-22

COMPARING THE APPROACHES TO MIGRATION FLOW PROJECTIONS: THE CASE OF INTERREGIONAL MIGRATION IN SLOVAKIA

Anton Bezák

Institute of Geography, Slovak Academy of Sciences, Bratislava, Slovak Republic

Abstract: The aim of this paper is to assess the performance of four migration projection models by using them to generate short-term forecasts of migration flows between 12 functional regions in Slovakia. The results indicate that the best forecasts are provided by either growth factor models or conditional probability models. On the contrary, spatial interaction models and movement rates models do not generate particularly accurate projections.

Key words: migration flow projections, interregional migration in Slovakia, movement rates models, conditional probability models, growth factor models, spatial interaction models.

1. INTRODUCTION

During the past two or three decades a marked drop of fertility, low and relatively stabilized level of mortality and gradual decline of interregional differences in natural population change have become an inherent feature of population development in most European countries. An immediate consequence of this trend is that migration has emerged as the most important component of population dynamics. Regional and local population change and distribution are now to a large extent affected by internal migration. It is not surprising, therefore, that migration begins to play an increasingly important role in population projections at both the local and regional levels.

As Willekens and Baydar (1986) point out, a dominant feature of classical approach to the integration of migration into population projection models is that only net migration, that is, inmigration minus outmigration, is taken into account. More recently, the net migration component has been decomposed into gross inflows and outflows. In the multiregional system, these gross flows are further disaggregated by region of origin and region of destination. The result is a set of region-to-region migration flows.

In geographic and demographic literature there are very few studies examining the performance of various approaches to migration flow projections. The aim of this paper is to explore the effectiveness of four projection models by using them to generate short-term forecasts, which can be compared against observed flows. Migration data from the current registration of population in Slovakia are utilized for this purpose, and the year 1993 has been chosen as the base period for producing forecasts of migration flows between 12 functional regions during 1996. These forecasts are then compared with observed flows for the target period using three goodness-of-fit statistics in order to evaluate the performance of alternative models.

The remainder of this paper is organized in four sections which in turn describe the projection models to be tested, discuss the data set and spatial units used in the analysis, present the results of empirical tests and consider some implications for further research.

2. MIGRATION FLOW PROJECTION MODELS

There exists in the literature a wide variety of migration models that can be classified on the basis of the definition and measurement of the migration variable, level of aggregation, structural form, calibration methodology, and purpose for which the model is intended (Clark 1982, Stillwell and Congdon 1991). Most of them have been constructed to provide descriptions and explanations of the historical processes of population redistribution at the macro and micro levels. On the contrary, the development and use of migration models as tools for projection and forecasting is less well advanced. Given the importance of the migration component in population change, it becomes essential to explore the forecasting performance of the various models that have been used or are available for use in migration flow projections and examine some choices about the way in which projections are constructed on the basis of these models.

In the attempt to assess the performance of different migration projections, four alternative types of migration models, recommended by Stillwell (1986, 1991), were selected for comparison. Although the incorporation of non-demographic data is feasible within the framework of several models, no attempt is made in any of these models to identify and include explicitly any economic, social, political or other factors, which are known to affect interregional migration. Each model requires a matrix of interregional migration flows for a historical base period. In addition, further information on regional population size, total outmigration and inmigration flows, as well as on the overall level of migration in the multiregional system is needed for a projection period in the form of independent forecasts. As these forecasts can be produced using a variety of different methods, several versions of each projection model can be considered.

The first type of models to be tested is a **movement rates model**, which simply applies historical interregional rates to initial populations of the projection period. The model has the form

$$\hat{M}_{ij} = m_{ij}\,\hat{P}_i$$

where $\hat{M}_{ij}$ is the projected migration flow from region i to region j, m_{ij} is the migration rate observed for a historical period, and $\hat{P}_i$ is the population of region i at the beginning of the projection period[1]. Historical migration rates, m_{ij}, are computed by dividing each migration flow from region i to region j, M_{ij}, by the origin population at the beginning of the period, P_i. Flows for the projection period are determined under the assumption that the historical rates remain constant. As Stillwell (1986) points out, the results obtained by this model can be used as a standard for comparing projections constructed by other models.

It should be mentioned, however, that the movement rates model requires an independent projection of the population of each region at the beginning of the projection period. There is now a large number of regional projection models that can be employed to generate population forecasts (cf. Openshaw and van der Knaap 1983). On the basis of previous experience with regional projection models in Slovakia (cf. Bezák and Holická 1995), a simple geometric growth model was chosen to forecast the initial regional populations for 1996. The model assumes that population change will occur at a constant percéntage rate over time. Note that the rates for individual regions were estimated by calculating the average percentage increase in the 1991-1993 period.

The second type of migration model is based on the conceptual decomposition of the migration flow into a level, generation and distribution component (van der Knaap and Sleeger 1984, Willekens and Baydar 1986)[2]. The overall level of migration in the multiregional system for a projection period is assumed to be known, and each flow from region i to region j is estimated by applying two probabilities derived from data for a historical period. A **conditional probability projection model** takes the form

$$\hat{M}_{ij} = \hat{L}\,po_i\,pm_{ij}$$

where, $\hat{L}$ is the overall level of migration for a projection period, measured by the total number of interregional moves that occur in the system, po_i is the probability of migration occurring from region i defined as

$$po_i = \frac{\sum_{j=1}^{n} M_{ij}}{\sum_{i=1}^{n}\sum_{j=1}^{n} M_{ij}}$$

[1] The symbol ^ indicates that the variable is associated with a projection period.

[2] For an application of this idea to historical analysis of interregional migration flows in Slovakia see Bezák (1999)

and pm_{ij} is the probability of migration to region j, given that the move originated from region i, defined as

$$pm_{ij} = \frac{M_{ij}}{\sum_{j=1}^{n} M_{ij}} .$$

It is assumed that both these probabilities remain stable over time.

The forecast of the overall migration level for the projection period depends on the extent of the data time-series available. In this study, the annual data for the 1986-1993 period were used and the projection of the migration level for 1996 was based on a quadratic polynomial trend.

If outmigration and inmigration totals, and , are projected independently, growth factor models can be used to distribute these totals on the basis of the migration distribution for the historical period. A **doubly constrained growth factor model** can be written as

$$\hat{M}_{ij} = \hat{A}_i \, \hat{B}_j \, g_{ij} \, M_{ij} .$$

The growth factor element, g_{ij}, is defined as the product of the ratio of projection to historical period total outmigration from region i and the ratio of the projection to historical period total inmigration to region j, that is

$$g_{ij} = \frac{\hat{O}_i \, \hat{D}_j}{O_i \, D_j} .$$

The balancing factors, $\hat{A}_i$ and $\hat{B}_j$, defined as

$$\hat{A}_i = \frac{\hat{O}_i}{\sum_{j=1}^{n} \hat{B}_j \, g_{ij} \, M_{ij}}$$

$$\hat{B}_j = \frac{\hat{D}_j}{\sum_{i=1}^{n} \hat{A}_i \, g_{ij} \, M_{ij}}$$

are introduced to ensure that

$$\hat{O}_i = \sum_{j=1}^{n} \hat{M}_{ij}$$

$$\hat{D}_j = \sum_{i=1}^{n} \hat{M}_{ij}$$

and internal consistency in the projected migration matrix is achieved.

The independent projection of the total outmigration and inmigration flows, $\hat{O}_i$ and $\hat{D}_j$, were obtained in three different ways suggested by Stillwell (1986). In the first version of the growth factor model, historical outmigration and inmigration rates were applied directly to projection period populations as derived within the context of the movement rates model. In the second version these rates were first adjusted using a multiplier reflecting the estimated changes in the general level of mobility during the 1993-1996 period and then applied to the projection period populations.

The third version of the model is based on the assumption that the overall migration level has been projected independently (as within the conditional probability model). The overall total is then disaggregated into regional outflow and inflow totals using historical probabilities po_i and pd_j, defined analogously as in the case of the conditional probability model. Clearly, all these probabilities are assumed to be stable in time.

Spatial interaction models can also be used for the purpose of generating flow distribution from independently projected regional outmigration and inmigration totals, $\hat{O}_i$ and $\hat{D}_j$. Unlike the growth factor model in which the effect of distance is assumed implicit in the historical migration matrix, spatial interaction models contain an explicit functional relationship between migration and distance. Evidently, any one of the family of spatial interaction models (Wilson 1971) can be selected for the projection of migration flows. A **doubly constrained spatial interaction model** with a power distance function can be formulated in this context as

$$\hat{M}_{ij} = \hat{A}_i \; \hat{B}_j \; \hat{O}_i \; \hat{D}_j \, c_{ij}^{-\beta}$$

where c_{ij} is distance between region i and region j and β is a distance decay parameter, which can be interpreted as a measure of the general propensity to migrate over distance.

The balancing factors, $\hat{A}_i$ and $\hat{B}_j$, defined as

$$\hat{A}_i = \left[\sum_{j=1}^{n} \hat{B}_j \; \hat{D}_j \, c_{ij}^{-\beta} \right]^{-1}$$

$$\hat{B}_j = \left[\sum_{i=1}^{n} \hat{A}_i \; \hat{O}_i \, c_{ij}^{-\beta} \right]^{-1}$$

are introduced again to ensure that the constraints

$$\hat{O}_i = \sum_{j=1}^{n} \hat{M}_{ij}$$

$$\hat{D}_j = \sum_{i=1}^{n} \hat{M}_{ij}$$

are satisfied.

A necessary prerequisite to projection on the basis of the spatial interaction model is the calibration of parameter β for the historical base period. In addition, independent forecasts of the total outmigration and inmigration flows for the projection period are required prior to interregional flow projection. In this study, these forecasts were produced in the same three alternative ways as in the case of the growth factor model.

Concluding this section, one relevant point must be mentioned. As noted, all the models outlined above require independent estimates of some exogenous variables for a projection period. Consequently, each of them has two sources of forecast error: one caused by errors in projecting exogenous variables and the other caused by errors in flow projection itself. In order to discriminate between the two sources of forecast errors, a separate version of each model was considered in which observed values of exogenous variables for the projection period were used.

3. DATA SOURCES AND STUDY REGIONS

The models presented in the preceding section were evaluated using data on migration between functional regions in Slovakia. The primary data used in this study are the data on the total number of persons leaving a given administrative district for another district, which are reported annually in electronic form by the Statistical Office of Slovak Republic. Note that these data are counts of moves rather than of transitions. If a person makes several moves across district boundaries during a given year, it appears in the data set as many times as the person moves.

Defining study area units, the 38 old administrative districts in Slovakia were aggregated into 12 functional regions depicted in Figure 1. These regions were defined on the basis of some previous studies by this author devoted to functional regionalization of Slovakia (cf. Bezák 1991a, 1991b). Note that each of them exceeds 300 000 in population. The main reason here was one of avoiding the problems of sparse matrices, which arise with data for single-year migration flows between districts. Consequently, the 38 × 38 primary matrices of migration flows between districts were transformed to the 12 × 12 matrices of interregional flows.

As mentioned in the introductory section, the year 1993 has been chosen as the base period for producing forecasts of migration flows during 1996. Therefore, two migration matrices were used in the analysis. One of them contains data for 1993 and forms the historical migration matrix. The second one contains data for 1996, which are required to compare the degree of agreement between the projected and observed flows in the target period. As migration within functional regions is not considered, all diagonal elements in the two migration matrices were set to zero.

The last point to be discussed in this section refers to the distance matrix required for calibrating the spatial interaction model. Distances between functional regions were measured in principle as the road distances between the largest cities of the regions. It should be noted, however, that in the case of four regions (Lower Váh, Nitra, Liptov-Orava-Turiec, and Gemer-Novohrad) the preference was given to cities with a central position. As the result, the following cities were taken into account in computing the interregional distances: Bratislava, Piešťany, Topoľčany, Nové Zámky, Žilina, Ružomberok, Banská Bystrica, Rimavská Sobota, Košice, Poprad, Prešov, and Michalovce.

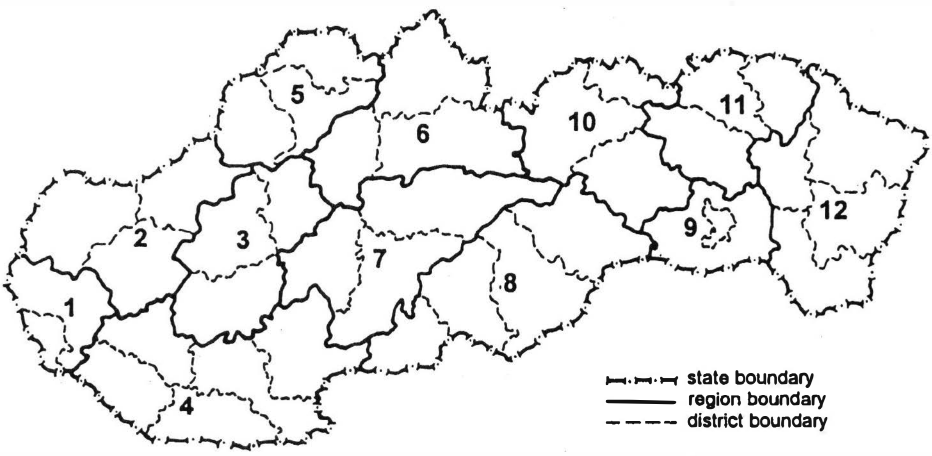

Fig. 1 The 12 functional regions used in the study
1. Bratislava metropolitan region (constituent districts: Bratislava, Bratislava vidiek), 2. Lower Váh region (Senica, Trenčín, Trnava), 3. Nitra region (Nitra, Prievidza, Topoľčany), 4. Danubian region (Dunajská Streda, Galanta, Komárno, Levice, Nové Zámky), 5. Middle Váh region (Čadca, Považská Bystrica, Žilina), 6. Liptov-Orava-Turiec region (Dolný Kubín, Liptovský Mikuláš, Martin), 7. Zvolen region (Banská Bystrica, Zvolen, Žiar nad Hronom), 8. Gemer-Novohrad region (Lučenec, Rimavská Sobota, Rožňava, Veľký Krtíš), 9. Košice metropolitan region (Košice, Košice vidiek), 10. Spiš region (Poprad, Spišská Nová Ves, Stará Ľubovňa), 11. Šariš region (Bardejov, Prešov, Svidník), 12. Zemplín region (Humenné, Michalovce, Trebišov, Vranov nad Topľou)

4. EMPIRICAL RESULTS

The forecast accuracy of the models considered can be measured in a number of different ways and there are many different criteria, which could be employed. In this study, three goodness-of-fit statistics were used to measure the degree of agreement between projected and observed flows in the target period.

The Mean Absolute Difference (MAD) is the sum of the absolute deviations between the observed and projected flows divided by the sum of the observed flows in the system and expressed as a percentage. The statistic has a lower limit of zero indicating perfectly accurate projections but its upper limit is variable. The Index Of Dissimilarity (IOD) is calculated as the sum of the deviations between the observed and projected proportions of total migration in each cell of the migration matrix. The statistic has a minimum of zero when projections are perfect and a maximum of 100 in the reverse case. The final statistic is the coefficient of determination (R2) measuring the proportion of total variation in the matrix of observed flows, which is statistically explained by the matrix of projected flows. The statistic R2 ranges between 0 and 1. Zero indicates no correspondence between the observed and projected flows, one indicates perfect correspondence. A detailed discussion of various statistics for comparing observed and projected spatial interaction matrices can be found in Knudsen and Fotheringham (1986).

Summary goodness-of-fit statistics associated with each of the models considered are presented in Table 1, disaggregated by alternative versions of the particular models. To obtain a general indication of how well the various models perform, it is useful to start with the subset of projections in which observed values of endogenous variables for the projection period are input. This set of results shows the degree of forecast accuracy, which is attributable to the model itself and eliminates forecast errors caused by independent estimation of endogenous variables.

As one might expect, the goodness-of-fit of the models depends on the amount of information that is transferred from the historical period and incorporated into the model. Using the movement rates model as a standard for comparing the other projections, it is evident that the growth factor model generates the best projections, where the historical migration is simply adjusted to comply with new row and column totals. Sufficiently accurate forecasts can also be produced by the conditional probability model, in which the observed total level of movement is disaggregated by applying two historical probabilities. On the contrary, the poorest projections are provided by the doubly constrained spatial interaction model, in which information carried forward from the historical period is reduced to one parameter.

As regards the differences in forecast accuracy among alternative specifications of the same model, several interesting conclusions can be drawn from Table 1. Negligible differences were found in the case of both the movement rates and conditional probability models. As shown in Table 1, the projections based on estimated regional populations or level of movement tend to be only marginally less accurate than those incorporating observed values of these variables. Note that these small differences are due to the use of very accurate population and migration level forecasts for the projection period.

On the other side, the substantial differences among different model specifications can be identified for the other two models, using the MAD as the fit statistic. When estimated gross flows are incorporated into the growth factor model, the mean absolute difference between the observed and projected flows increases from 11 to 17-24 %. This result illustrates the proportion of the mean deviation, which is caused by errors in the gross flow estimation and confirms the need to develop more sophisticated methods for this independent projection. A similar decline in the forecast accuracy can also be found for the spatial interaction model. It should be noted, however, that in this case most of the deviation is accounted for by the distribution model itself.

Table 1 Goodness-of-fit statistics for selected migration projection models

Type of model	MAD	IOD	R^2
MOVEMENT RATES MODEL			
using projected regional populations	24,15	10,04	0,9210
using observed regional populations	23,97	10,04	0,9212
CONDITIONAL PROBABILITY MODEL			
using projected migration level	16,93	8,11	0,9388
using observed migration level	16,22	8,11	0,9388
GROWTH FACTOR MODEL			
using rate-based gross flow projections	24,18	8,04	0,9349
using rate-based gross flow projections adjusted for changes in the migration level	17,16	8,04	0,9401
using probability-based gross flow projections	16,93	8,11	0,9388
using observed gross flows	10,7	5,35	0,9837
SPATIAL INTERACTION MODEL			
using rate-based gross flow projections	35,07	15,13	0,7953
using rate-based gross flow projections adjusted for changes in the migration level	31,1	15,13	0,7953
using probability-based gross flow projections	30,93	15,17	0,7937
using observed gross flows	28,33	14,17	0,8604

Note: gross flows = total outmigration flows and total inmigration flows

Furthermore, it is clear that both projections with rate-based gross flow forecasts can be considerably improved by the adjustment procedure reflecting the estimated changes in the overall level of mobility. It is especially true of the growth factor model, but a small improvement is also found in the case of the spatial interaction model. Finally, there are no apparent differences between rate-based and probability-based methods of forecasting gross flows, provided that the historical rates were adjusted for estimated changes in the overall mobility prior to flow projection.

To conclude this section, some comments on goodness-of-fit statistics used in this study are worth mentioning. The most accurate statistic for comparing the performance of projection models appears to be the mean absolute difference. Although the MAD is a rather crude measure of the deviation between the observed and projected flows, it has the advantage of being relatively easy to interpret. The other two statistics - index of dissimilarity and coefficient of determination - seem to be rather poor for the purpose of assessing the accuracy of the alternative models. From Table 1, it is clear that both these statistics are relatively insensitive to variations in model specification. In addition, the coefficient of determination yields artificially high values even when the mean absolute difference exceeds 20 %.

5. CONCLUSION

In this study, an attempt has been made to evaluate the performance of four migration models by using them to produce short-term projections of interregional migration flows among the 12 functional regions in Slovakia. The results obtained indicate that the spatial interaction model does not provide particularly accurate short-term projections of migration flows. Although projections generated by the other three models are superior, all models generally perform less accurately than has been expected. It should be noted, however, that the proportion of error attributable to the gross flow estimation is still high in the case of the growth factor model. An approach, which distributes gross flows on the basis of the pattern of historical migration, has, therefore, the considerable potential to produce more accurate projections. Evidently, further work is needed not only to develop and test alternative projection models but also to elaborate more sophisticated methods of gross flow estimation.

Acknowledgements

Research on this article was supported, in part, by the Scientific Grant Agency of the Ministry of Education of the Slovak Republic and the Slovak Academy of Sciences under grant 2/4066/99.

References

Bezák, A. (1991a): Migračné toky a regionálna štruktúra Slovenska: hierarchická regionalizácia (Migration flows and the regional structure of Slovakia: a hierarchical regionalization). *Geografický časopis*, 43, 193-202.

Bezák, A. (1991b): Migračné toky a regionálna štruktúra Slovenska: nehierarchická regionalizácia (Migration flows and the regional structure of Slovakia: a non-hierarchical regionalization). *Geografický časopis*, 43, 265-275.

Bezák, A. - Holická, A. (1995): Komparatívna analýza extrapolačných modelov pre regionálne populačné projekcie (A comparative analysis of some extrapolation models for regional population projections). *Geografický časopis*, 47, 233-246.

Bezák, A. (1999): Stabilita interregionálnych migrácií na Slovensku v rokoch 1984-1995 (Stability of interregional migration in Slovakia, 1984-95). *Pedagogická fakulta Masarykovy Univerzity v Brně, Geografie*, 11, 9-13.

Clark, W.A.V. (1982): Recent research on migration and mobility: a review and interpretation. *Progress in Planning*, 18, 1-56.

Knaap, G.A. van der - Sleegers, W.F. (1984): Structural analysis of interregional and intraregional migration patterns. In Heide, H. ter, Willekens, F.J., eds. *Demographic research and spatial policy: the Dutch experience*. London (Academic Press), pp. 113-131.

Knudsen, D.C. - Fotheringham, A.S. (1986): Matrix comparison, goodness-of fit, and spatial interaction modeling. *International Regional Science Review*, 10, 127-147.

Openshaw, S. - Knaap, G.A. van der (1983): Small area population forecasting: some experience with British models. *Tijdschrift voor Economische en Sociale Geografie*, 74, 291-304.

Stillwell, J. (1986): The analysis and projection of interregional migration in the United Kingdom. In Woods, R., Rees, P., eds. *Population structures and models: developments in spatial demography*. London (Allen and Unwin), pp. 160-202.

Stillwell, J. (1991): Spatial interaction models and the propensity to migrate over distance. In Stillwell, J., Congdon, P., eds. *Migration models: macro and micro approaches*. London (Belhaven Press), pp. 34-56.

Stillwell, J. - Congdon, P., eds. (1991): *Migration models: macro and micro approaches*. London (Belhaven Press).

Willekens, F. - Baydar, N. (1986): Forecasting place-to-place migration with generalized linear models. In Woods, R., Rees, P., eds. *Population structures and models: developments in spatial demography*. London (Allen and Unwin), pp. 203-244.

Wilson, A.G. (1971): A family of spatial interaction models, and associated developments. *Environment and Planning*, 3, 1-32.

Resume

Porovnanie metód projekcie migračných tokov (interregionálne migrácie na Slovensku ako príklad)

V tejto štúdii sme sa pokúsili zhodnotiť účinnosť štyroch migračných modelov, ktoré sa dajú využiť pri projekciách interregionálnych migračných tokov. Prvým z testovaných modelov je *model jednoduchých mier*, v ktorom sa predpovede migračných tokov získajú tak, že sa historické miery migrácie aplikujú na odhadnuté veľkosti populácií východiskových regiónov na začiatku projekčného obdobia. V *modeli podmienených pravdepodobností* sa odhadnutá celková úroveň migrácie v projekčnom období dezagreguje na jednotlivé migračné toky na základe dvoch pravdepodobností odvodených z historických dát. Ak sú známe nezávislé odhady veličín udávajúcich celkový počet emigrácií a imigrácií za každý región v projekčnom období, potom pomocou *modelu rastových koeficientov* sa tieto veličiny dezagregujú na migračné toky na základe charakteru rozloženia tokov v historickom období. Z predpokladu nezávislej projekcie celkového počtu emigrácií a imigrácií za každý región vychádza aj *model priestorovej interakcie*, v ktorom sa však rozhodujúca úloha prisudzuje vplyvu vzdialenosti na migráciu.

Kvalitu jednotlivých modelov sme zhodnotili tak, že na základe migračných dát z roku 1993 sme zostrojili predpovede migračných tokov medzi 12 regiónmi na Slovensku v roku 1996, ktoré sme potom porovnali so skutočnými tokmi pomocou troch štatistík zhody. Keďže všetky modely vyžadujú nezávislé odhady hodnôt exogénnych premenných v projekčnom období, uvažovali sme niekoľko alternatívnych verzií každého modelu v závislosti od spôsobu špecifikácie týchto premenných. Z výsledkov empirických testov, zhrnutých v tab. 1, vyplýva, že presnosť predpovedí generovaných uvažovanými modelmi nie je vo všeobecnosti vysoká. Najlepšie predpovede produkuje model rastových koeficientov a model podmienených pravdepodobností, podstatne horšie predpovede sa získali prostredníctvom modelu jednoduchých mier a najmä pomocou interakčného modelu. Výsledky ďalej potvrdzujú, že presnosť projekcie migračných tokov závisí nielen od účinnosti vlastného modelu, ale aj od množstva

informácie týkajúcej sa projekčného obdobia, ktorá sa inkorporuje do modelu v podobe viac alebo menej kvalitného odhadu hodnôt exogénnych premenných. V súvislosti s tým možno za perspektívny projekčný model pokladať model rastových koeficientov, v rámci ktorého existuje možnosť značnej redukcie chýb predpovede aplikáciou dokonalejších metód odhadu exogénnych premenných.