ECONOMETRIC ANALYSIS OF INTERNAL MIGRATION IN RUSSIA[1]

EKONOMETRIJSKA ANALIZA UNUTRAŠNJIH MIGRACIJA U RUSIJI

VAKULENKO E. S.,
Russia, National Research University Higher School of Economics,
Department of Mathematical Economics and Econometrics,
e-mail: lena-vakulenko@yandex.ru,

MKRTCHYAN N. V.,
Russia, National Research University Higher School of Economics,
Institute of Demography,
e-mail: nmkrtchyan@hse.ru,

FURMANOV K. K.,
Russia, National Research University Higher School of Economics,
Department of Mathematical Economics and Econometrics,
e-mail: furmach@menja.net,

Abstract: *Using panel data for the 2001-2008 period we estimate the gravity model of migration between Russian regions. We show that though the migration flows have been quite stable, their determinants have changed substantially. Our special attention is drawn to the role of distance between the regions. Our special attention is drawn to the role of distance between regions as one of the main factors of migration. Dividing pairs of regions into nine groups according to the distance between regional centers, we estimate the model for each group separately. We find out that social and economic factors are affecting migration mainly between nearby regions. Yet the intensity of flows between distant (>500 km) areas is almost uncorrelated with indicators of social and economic development of regions.*
Keywords: *migration, gravity model, panel data, Russian regions.*

Apstrakt: *Korišćenjem tabelarnih podataka za period 2001.-2008.god., obavili smo procijenu gravitacionog modela migracija između regiona Rusije. Iako su tokovi migracija prilično stabilni, u ovom radu pokazujemo da je došlo do značajnih promjena njihovih determinanti. Posebnu pažnju smo posvetili ulozi udaljenosti između datih regiona kao jednog od glavnih faktora migracija. Podjelom parova regiona na devet grupa, na osnovu razdaljine između regionalnih centara, procijenili smo dati model za svaku grupu pojedinačno. Otkrili smo da socijalni i ekonomski faktori utiču na migracije uglavnom između susjednih regiona. Sa druge strane, intezitet tokova između udaljenih područja (>500 km) gotovo da nema veze sa indikatorima društvenog i ekonomskog razvoja regiona.*
Keywords: *migracije, gravitacioni model, tabelarni podaci, regioni Rusije*

JEL Classification: J 61; C 23;
Original Scientific Paper, Recived: July 01, 2011

1. Introduction

Migration is a key consequence of social and economic changes and, besides, one of regional inequality evaluation criteria. Social and economic indicators in Russian regions differ dramatically, which theoretically should stimulate interregional migration. However, despite the fact that development rates vary significantly between regions, migration trends remain stable.

The chart below (see fig.1) shows the decrease in Russia's officially recorded internal migration rate during the 1990s and its stabilization from

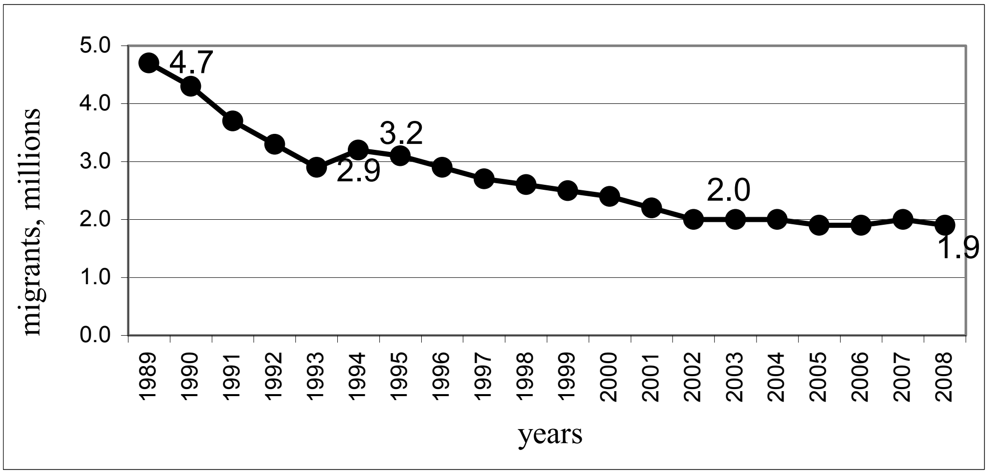
[1] „Authors would like to thank Inna Kosmachiova for indispensable help in preparing this article".

2002 on. However, it should be noted that in the end of 1995 the population registration system was notably modified, therefore comparability between before and after data is hardly possible. Nevertheless, as seen in the chart, no drastic migration rate changes occurred in this period. This is probably due to the fact that the adoption of the new registration system and the rearrangement of statistics services actually took several years. This results in scarcity of comparable data and unreliability of the research based on data gathered in the mid-to-late 90s.

Despite the growing number of publications on econometric analysis of migration, there seems to be poor progress in this field. In our research, an attempt is made to improve the migration model by means of flexible specification and segmentation of the data analyzed into homogenous groups. Considering the mentioned above problem of data inconsistency, the 1990s data are not analyzed, and thus the research is based on the more reliable 2001-2008 data.

The paper proceeds as follows. Section 2 contains a short review of the previous studies in the field. Section 3 describes the dataset. In section 4 we present our econometric model. Section 5 describes the obtained results, and section 6 draws the conclusions.

Fig. 1. Internal migration in Russia, 1989-2008.



2. Previous Research

The article by Denisenko (1994) was one of the first econometric studies of migration in post-Soviet Russia. It explored the relevance of classical economic theory principles to transition economy. According to its results, 1990s migration flows didn't fully (if at all) conform to classical patterns. The author suggests this might be due to poor statistical data.

Brown (1997) cites in her study that higher wages and higher rate of housing privatization increase both in-migration and out-migration. She also shows that migration doesn't smooth out inter-regional inequality since population flows mostly occur between prosperous regions.

Japanese economist K. Kumo (2003) studies how determinants of migration change during transition from a centrally planned economy to a market economy. He shows the impact of state economic incentives on internal migration in Russia in 1980s and the growing influence of social and economic factors after transition to the market economy.

L. Korel and I. Korel in their study (1999) aimed at: a) regions classification based on typical migration processes; b) search of key determinants of migration for transition economy in Russia. Their regression analysis of cross-sectional mid-90's data showed that average income, housing prices and geographic location are key factors shaping migration, while unemployment rate remains an insignificant factor. While Gerber (2000) pointed at some downsides of this study and put its results in doubt due to: simultaneity bias (endogeneity) and duplication of some regions (*autonomous okrugs* counted up as part of other regions).

Meanwhile, in his own studies Gerber made a step forward moving on from cross-sectional data analysis to panel data analysis. His results for panel dataset of net migration in Russian regions from

1993 to 1997 showed that labor market conditions shape migration in Russia to the same extent as they do in market economies. Unfavorable economic conditions in a region make people move to regions with higher real wages, lower unemployment rate and lower rate of unprofitable business. This result remains valid with control variables applied – i.e. provision of public goods, including housing; region's crime rate; urbanization level and region's geography.

Gerber's method, however, was criticized by Andrienko and Guriev (2004) who pointed to the fact that Gerber's random effects model is of limited application due to highly probable correlation with regressors. Therefore, giving preference to the fixed effects approach, the authors developed a modified gravity model of migration in Russia using panel data on region-to-region flows for the period of 1992-1999.

Their results showed that real income positively affects migration inflows to destination regions and negatively affects migration outflows from origin regions. However, due to positive effects of income level on outflows there are financial constraints for migrants from poor regions. The study also shows the negative impact of distance between regions and the positive effect of region population on migration flows, which is essentially key prerequisite of the gravity model. Besides, unemployment has negative effect for destination regions. It is worth noting that according to the authors, unobserved factors (or "individual effect") should also be taken into account.

The subsequent study by Gerber (2006) emphasizes the importance of dynamic effects of labor market (changes in wages and unemployment rate). Having analyzed panel data on net migration for the period from 1993 to 2002, Gerber showed the positive effects of higher real wages and the negative impact of unemployment on net migration flows. As to dynamic effects, real wages growth proved to increase net migration flows, while unemployment dynamics had no independent effects. It's worth noting that since 1996 the negative impact of unemployment reduced and the wages positive effects increased. According to Gerber, this reflects attainment of migration equilibrium with respect to unemployment, but not to wages.

Guriev and Andrienko also went on with their research (Guriev, Andrienko, 2006). This time they analyze the period from 1992 to 2003, including migration flows to and from the CIS countries and Germany, using the dynamic gravity model (with the lag of migration flow among explanatory variables). One of the method's innovations was the attempt to control the correlation between region-to-region migration flows. The results of their research show that gravity model is applicable to Russian data.

We would like to note another study by Japanese researcher Kumo (2006). He analyzed interregional migration data using not panel but cross-sectional data for 2003. Kumo's findings comply with results expected from application of gravity model with respect to distance and population. However, this study leaves some questions unanswered concerning region income interpretation and some other regional factors.

And lastly we would like to note a study by Oschepkov (Oschepkov, 2008) analyzing interregional migration in the period from 1990 to 2006. The author's results of gravity model application comply with other studies made earlier. They show that migration flows between regions are affected not only by wages and unemployment but also by their growth rates. Besides, the author notes that the downward migration trend is not related to decrease in migration factors effects, quite the opposite, the impact of standard factors on migration flows increases, which disagrees with the author's initial idea. Nevertheless, according to the author, liquidity constraints diminish, which may cause increase in unregistered (informal) migration rate.

The problem common for the studies based on panel data is the inconsistency of migration data due to the adoption of the new registration system and the drawn-out rearrangement of statistics services. In case of cross-sectional data analysis the issue of concern is the bias that can be caused by the unobserved heterogeneity between regions.

3. Data

To carry out our analysis we used data from the following sources:

- The data on social and economic situation in Russian regions for the period from 2001 to 2007 annually published by the Federal State Statistics Service (Russian Regions 2001, 2009).
- Interregional migration data for the period from 2001 to 2008 developed but not published by the Federal State Statistics Service. These data are collected for four migration types separately in terms of urban/rural residence.

However in our research, when estimating migration flows we make no distinctions between types of residence and suppose that migration flow from one region to another is the total number of migrants moving from town to town, from town to village, from village to town and from village to village.

- The matrix of distances between regional centers (capitals). The distance was calculated according to Tatevosov's method (Tatevosov, 1973), on the basis of railway mileage (the distance being calculated on the basis of railway mileage between regional centers, and in case of absence of railroads – on the basis of distance of flight to the nearest regional center).

We excluded from our panel the migration flows to or from the Republic of Ingushetia, the Republic of Chechnya and the Chukotka Autonomous Okrug due to insufficient data, as well as 9 autonomous okrugs (the Nenets Autonomous Okrug, the Komi-Permyatski Autonomous Okrug, the Khanty-Mansijsk Autonomous Okrug, the Yamalo-Nenets Autonomous Okrug, the Taimyr (the Dolgano-Nenets) Autonomous Okrug, the Evenk Autonomous Okrug, the Ust-Ordyn Buryat Autonomous Okrug, the Aginsk Buryat Autonomous Okrug, the Koryak Autonomous Okrug) which are administrative parts of the following regions: the Republic of Daghestan, the Republic of Kabardino-Balkaria, the Republic of Kalmykia and the Republic of North Ossetia-Alania – because the data available for those regions are unreliable. Thereby, our research covers 73 regions. Besides, we only used the data on region-to-region migration — assuming that migration within regions is a subject of other research with other data involved.

As already mentioned in the Introduction, there was a decrease in Russia's internal migration rate during the 1990s (see fig. 1). By the beginning of the period analyzed the internal migration rate stabilized at the level of 2 million migrants annually. What follows is a brief description of the trends in region-to-region flows, revealed by the analysis of migration gain/loss coefficients (migration gain coefficient is calculated by taking the difference between the number of in-migrants and out-migrants divided by the average annual population in the region (in thousands).

Table 1 represents the regions which have shown the highest migration gain and loss during the period from 2002 to 2008.

Table 1. Regions Leading in Migration Gain/Loss, 2002-2008.

Regions – Migration Gain Leaders	*Average Migration Gain/Loss Coefficient, 2002-2008*	*Population growth between 2002 and 2008/*	*Regions – Migration Loss Leaders*	*Average Migration Gain/Loss Coefficient, 2002-2008*	*Population growth between 2002 and 2008*
The Moscow Region	7.13	1.45%	The Chukotka Autonomous Okrug	-18.25	-5.66%
The Leningrad Region	6.07	-2.10%	The Magadan Region	-16.92	-10.44%
Moscow	4.25	1.17%	The Murmansk Region	-7.09	-5.28%
The Belgorod Region	3.89	0.86%	The Komi Republic	-6.94	-5.61%
The Krasnodar Territory	3.11	0.43%	The Kamchatka Region	-6.89	-3.91%
Saint-Petersburg	2.97	-1.59%	The Republic of Kalmykia	-6.48	-2.74%
			The Sakhalin Region	-6.23	-5.69%

Apparently, the key migration inflow leaders are Moscow, Saint-Petersburg, the Moscow region and the Leningrad region. These regions have the highest standards of living. Analysis of net migration in Russian federal districts in 2008 shows that the Central and the North-East Federal Districts only had migration surplus – that is due to the long-range attractive force of their "capitals". All other federal districts demonstrate migration loss in favor of the Central Federal District. The East-to-West migration (the so called "West Drift" (Mkrtchyan, 2005) and North-to-South migration are the directions of migration flows that remain unchanged for the last two decades.

Are these directions of migration flows to be attributed to recolonization of Russia's vast territories, to the tendency of moving to the more developed part of the country with more favorable climate, or are they shaped by social and economic factors, which can be interpreted by means of econometric calculations? In this article we make an attempt to answer these and some other questions.

4. Method of Analysis

We follow one of the most common approaches in econometric analysis of migration and use the so-called gravity model. The model takes its name from the well-known counterpart in physics: Newton's law of universal gravitation. In the beginning, gravity models used in social sciences assumed that some kind of attraction force between regions or countries is proportional to their "masses" (measured by population or income) and inversely proportional to the squared distance between regions. The gravity approach was used in 1929 by W.J. Reilly in his analysis of retail trade; in 1948 J.Q. Stewart proposed it as a tool for demographic research (Paas, 2000, pp. 9-10). Since then, the two main fields of gravity models' application are trade and migration analysis. The detailed review of gravity approach, its history and applications can be found in (Paas, 2000).

The first models were far from perfect and during the following decades researchers modified and extended the gravity equation by introducing new explanatory variables and relaxing the proportionality assumption. In fact, nowadays the term "gravity model" refers to almost every regression model of bilateral flows between regions or countries.

As our research is based on the panel data, we can take into account unobserved heterogeneity between regions, or individual effect of regional pairs. Individual effect absorbs all the factors, whose influence on migration flows is constant over time. As the distance between regions is constant, it is a part of the individual effect. That leads to the problem of identifiability: the effect of distance cannot be separated from the individual effect without making some additional assumptions (Verbeek, 2000, pp.321-322). However, as it is shown further, distance has the role different from that of other explanatory variables. It is worth mentioning that the individual effect refers not to a certain region but to a pair of regions, i.e. regions of origin and destination.

We use the following regression equation as the basis for our analysis:

$$\ln M_{i,j,t} = \alpha_{i,j,t} + \beta_t{}' Y_{i,t-1} + \gamma_t{}' Y_{j,t-1} + \varepsilon_{i,j,t} \qquad (1)$$

Here $M_{i,j,t}$ is the migration flow from region i to region j in the year t,

$Y_{i,t-1}$ is the vector containing characteristics of the region of origin in the year t-1,

$Y_{j,t-1}$ is the vector containing characteristics of the region of destination in the year t-1,

$\alpha_{i,j,t}$ is the individual effect of the regional pair (i, j) in the year t,

β_t and γ_t are the coefficient vectors corresponding to regional characteristics $Y_{i,t-1}$ and $Y_{j,t-1}$.

All explanatory variables are taken in logs. The complete list of variables is presented in Appendix A.

A problem that is likely to occur when estimating the gravity model of migration is endogeneity. Let us consider such an explanatory variable as, for example, unemployment rate. Of course, it can affect migration. Potential migrants may avoid regions with high unemployment rate because they are likely to face difficulties in finding a job there. But unemployment rate itself can be affected by migration. And as it is determined, to a certain extent, by the migration flow, it is correlated with random error in the regression equation. Such correlation is called endogeneity and leads to bias and inconsistency of estimates. The similar argument can be applied to other explanatory variables.

In order to avoid endogeneity, we include lagged regressors in equation (1), which is represented by "t-1" index. That's the method used in (Andrienko, Guriev, 2006) and (Gerber, 2006). A rationale for using lagged regressors is that explanatory variables in years t and t-1 are supposed to be highly correlated, so that the migration flow can be explained by lagged values. But migration in year t is not likely to affect regressors in the previous year, so the problem of endogeneity does not occur.

It's worth mentioning that the coefficient vectors β_t and γ_t and the individual effect $\alpha_{i,j,t}$ depend on time (*t*). A preliminary analysis based on short, two-year panels has shown that the coefficients were not constant over the analyzed period. To consider this without including too many variables in the model we assume that the coefficients have linear trends. Thus, vectors β_t and γ_t could be written in the following form: $\beta_t = \beta_0 + \eta t$, $\gamma_t = \gamma_0 + \lambda t$ where η and λ are the vectors containing slopes of the coefficient trends. Also, the individual effect can be expressed as $\alpha_{i,j,t} = \alpha_{i,j,0} + \delta t$. If there is no dynamics in the coefficients then δ, η and λ are equal to zero, and that is one of the hypotheses we test.

The model with linear dynamics in the coefficients can be easily estimated as it is, in fact, an ordinary linear regression model. Using the assumption of linear trends we can rewrite equation (1) as follows:

$$\ln M_{i,j,t} = (\alpha_{i,j,t} + \delta t) + (\beta_0 + \eta t)' Y_{i,t-1} + (\gamma_t + \lambda t)' Y_{j,t-1} + \varepsilon_{i,j,t} \qquad (2)$$

If we open the brackets we obtain:

$$\ln M_{i,j,t} = \alpha_{i,j,t} + \delta t + \beta_0 ' Y_{i,t-1} + \eta ' t Y_{i,t-1} + \gamma_t ' Y_{j,t-1} + \lambda ' t Y_{j,t-1} + \varepsilon_{i,j,t} \qquad (3)$$

Equation (3) is an ordinary regression equation that can be estimated by techniques common in panel data analysis. We estimate it using ordinary least squares with the so-called "within" transformation to handle fixed effect. Standard errors are calculated with respect to possible heteroscedasticity and autocorrelation. We use clustered standard errors technique (Stock, Watson, 2006) to allow for correlation between the random errors in observations related to the same region of origin.

There are well-known alternatives to the fixed effect approach, which are random effect and pooled regression model. We use F-test to choose between fixed effect and pooled regression and Hausman test to choose between random and fixed effects. Both tests reject their null hypotheses in favor of the fixed effect model at 1% level of significance.

About 2% of the observed migration flows are equal to zero and cannot be taken into account in logarithmic model. Such observations are excluded from the analyzed sample. We have also tried to handle them by substituting zeros for a small positive number (0.1 or 0.5) – a method considered in (Flowerden, Aitkin, 1982). That has not lead to substantial changes in estimates, so we present the results obtained without such a trick.

Apart from estimation of our model using the whole sample, we also analyze groups of regional pairs, grouped by the distance between the regional centers. We use the *k*-means method to divide all pairs into nine groups which are presented in table 2.

The use of distance as a grouping variable is based on the following consideration. We assume that migrants usually prefer to move to nearby regions. So, those who travel at long distances seem to have some uncommon motives, different from those that drive migration at short distances.

5. Results

Our model fits the data very poorly when estimated using the whole sample of regions. We use R^2-within coefficient as a measure of goodness-of-fit and it equals 0.07 for the model with constant coefficients and 0.08 for the model with linear dynamics in coefficients. But when we estimate the gravity equation using separate groups of regional pairs we find that R^2-within substantially varies between those groups (see table 2).

The strongest correlation between migration flows and social and economic characteristics of regions is observed in the first group which includes pairs of nearby regions (<500 km between regional centers) and accounts for the highest number of migrants. R^2-within is also relatively high in the group of the most distant (>10000 km) regions but the migration flows in that group are scanty. Those are mainly the flows between the European part of Russia and the Far East region and they include less than 2% of all migrants. R^2-between and R^2-overall coefficients are partly presented in Appendix B, their variation among the groups seems to be rather chaotic. However, interpretation of these coefficients in the framework of the fixed effect model is vague so we do not discuss them here.

Table 2. Goodness of fit in the groups of regional pairs

Distance between regional centers	R^2-within, constant coefficients	R^2-within, linear dynamics in coefficients	F-statistics for significance of linear trend	% of migrants, 2001-2008.
≤ 500 km	0.26	0.32	F(41, 1875)=4.34	32.66%
500-1000 km	0.11	0.14	F(41, 4778)=3.78	20.59%
1087-2022 km	0.08	0.11	F(41, 7128)=6.10	19.04%
2026-3186 km	0.06	0.10	F(41, 4885)=4.62	10.12%
3204-4585 km	0.05	0.07	F(41, 3528)=2.43	6.60%
4593-6094 km	0.07	0.10	F(41, 2523)=2.24	3.82%
6111-7891 km	0.08	0.12	F(41, 1906)=2.36	2.53%
7941-9985 km	0.11	0.16	F(41, 2325)=3.47	2.69%
> 10000 km	0.18	0.24	F(41, 1224)=3.20	1.97%

In all groups the model with the linear trend in coefficients outperforms the model with constant coefficients. P-values for relevant F-statistics never exceed 1%. Thus we can conclude that the influence of different factors on migration flows has changed over time.

Despite the rather low R^2-within, our model fits the data quite well. Correlation between the observed migration flows and their predicted values is over 0.95 in every group. Such a discrepancy means that migration flows are determined more by the individual effect of regional pairs than by the influence of social and economic factors included in the model.

Estimates of the gravity model with linear dynamics in coefficients are presented in Appendix B. It includes results only for the "<500 km" and the ">10000 km" groups where correlation between explanatory variables and migration flows is relatively high. Instead of presenting estimates for the slopes of linear trend in coefficients, we present coefficients and their standard errors for years 2002 and 2008 separately.

Appendix C contains results of a visual test for model specification. For the group of nearby regions the test does not reveal any specification error. In the group of distant regions the visual test reveals heteroscedasticity which could occur due to different forms of unobserved heterogeneity that cannot be taken into account using ordinary fixed effect model.

Now let us turn to description and interpretation of obtained estimates.

Population and demographical factors. It is natural that the intensity of a migration flow is determined, first of all, by the population in the region of origin. In the distant regions' group the corresponding coefficient appears to be insignificant but that seems to be due to the large standard error.

Migrants are attracted to regions with lower percentage of urban population. That effect grows stronger with time but it is significant only in the group of nearby regional pairs.

As for the group of distant regions, we find that the role of age composition in the region of destination has become more important during the analyzed period. Regions with a large share of young (younger than 16) population has become less attractive to migrants.

Labor market. In the beginning of the analyzed period, the only indicator of labor market which had significantly affected migration flows at short distances was the share of unprofitable enterprises. People tried to leave regions where that share was high. By the year 2008 the role of that factor in the region of origin has diminished but it has become significant in the region of destination. Ratio of wages in the regions of origin and destination and tension coefficient (number of unemployed per one vacancy) in the region of destination have also appeared to be influential. Migrants have become more motivated by the possibility of finding a job and by the expected change in their wages.

Another factor that started to substantially affect migration by the end of the analyzed period is the wage growth rate. Both in regions of origin and destination it is positively correlated with the intensity of migration flow. So, migration is more active between the rapidly evolving regions.

Effects of the wage ratio and the share of unprofitable enterprises are significant in the group of distant regional pairs. Our estimates also demonstrate that the elasticity of migration flow with respect to the tension coefficient is significantly negative during the analyzed period, but that result is hard to interpret. It could appear randomly, as the consequence of statistical error which is likely to occur when estimating models with many variables.

Real estate. Here the key determinant is housing availability. It influences migration in both groups of regional pairs but in a different way. Migration flows at short distances are attracted by housing availability in the region of destination while distant migration is negatively correlated with the similar indicator of the region of origin. In both cases, affordability is the attracting factor and remains significant during the whole analyzed period.

By the year 2008 the role of housing price has grown. In the group of nearby regions negative correlation between the coefficient of affordability (i.e. the ratio of the average housing prices to the average monthly income) in the region of destination and intensity of migration flow has become significant.

The amount of newly-built dwellings in the region of destination was an attractive factor in 2002 but its effect has diminished. Perhaps, the real estate development was caused by the growing demand from "stayers", not migrants. That could cause an increase in prices and make the new dwellings unaffordable for immigrants.

Quality of life. The most noticeable factor is the air pollution. In the group of distant regional pairs, migration flows originate from polluted regions. On the contrary, flows at short distances are attracted to polluted areas. Probably, that is because those areas are more industrially developed.

In the group of nearby regions the number of health care employees in the region of destination is significant attractor. Hospital bed complement in the region of destination positively affects migration flows. In 2002 this effect is observed in both groups of regional pairs but it becomes insignificant by 2008 in the group of nearby regions.

We have to say that the assumption of linear dynamics in coefficients is rather restricting and can lead to incorrect conclusions. Consider some explanatory variable that affected migration in 2002 but its influence has diminished over time and the corresponding coefficient has become zero. If we apply a model with linear trend then our estimate of that coefficient in 2008 can change the sign and become negative and, probably, significant. But, on the other hand, using more flexible specification means including more variables into already overburdened model.

6. Conclusion

Our model has a substantial drawback: it's overburdened with explanatory variables. A large amount of estimated coefficients makes the obtained results hard to interpret. Significance of some of the coefficients can be the consequence of type I statistical errors which are highly probable to occur in the multiple hypothesis testing. Maybe, that problem can be solved by using integral indicators of labour market, real estate and so on, instead of a great number of variables. However, we keep that for further research, and now we'll point at the results which seem to be both most interesting and most reliable.

First of all, *the determinants of migration flows have changed substantially during the 2002-2008 period.* It follows from the highly significant advantage that the model with dynamic coefficients has over the one where coefficients are assumed to be constant (see table 2). That advantage takes place in all the groups of regional pairs.

The second finding is that *social and economic factors affect migration mainly between nearby regions.* Our attempts to model the flows between distant (from 500 to 8000 km) regions lead to very poor goodness of fit measured by R^2-within coefficient. After 8000 km R^2-within increases again but the migration flows at such distances are scanty and do not contribute much to the overall interregional migration.

References

Andrienko, Y. and Guriev, S., 2004. Determinants of Interregional Mobility in Russia.

Evidence from Panel Data. *Economics of Transition*, Vol., 12, (1), pp. 1-27.

Andrienko, Y. and Guriev, S., 2006b. Razrabotka pricladnoy modeli vnutrennih i vneshnih migracionnih potokov naseleniya dlya regionov Rossiyskoy Federacii. Otchet po proektu v ramkah programmi podderjki nezavisimih economicheskih analiticheskih centrov MONF, CEFIR.

Arellano, M. and Bond, S., 1999. Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations. *The Review of Economic Studies*, Vol. 58, No. 2, pp. 277-297.

Brown A., 1997. The Economic Determinants of Internal Migration Flows in Russia during Transition. *WDI WP №. 89*.

Denisenko, M.B., (1994). Determinanti mejregionalnoy migracii v Rossii i ih economicheskoe znachenie. M., S-Pb.

Flowerden, R. and Aitkin, M., 1982. A method of fitting the gravity model based on the Poisson distribution. *Journal of regional science*, vol. 22, No. 2.

Gerber, T., 2005. Individual and Contextual Determinants if Internal Migration in Russia, 1985-2001. University of Wisconsin, Mimeo.

Gerber, T., 2006. Regional Economic Performance and Net Migration Rates in Russia, 1993-2002. *International Migration Review*, Vol. 40, №3, pp. 661-697.

Korel, I. and Korel, L., 1999. Migracionnii i makroeconomicheskie processi v postsocialisticheskoy Rossii: regionalniy aspect. RPAI Fond Evrasiya.

Kumo, K., 2006a. Migraciya naseleniya v postsovetskoy Rossii. *DP Series A, №480*, The Institute of Economic Research, Hitotsubashi University, Tokyo, 2006a.

Kumo K. Interregional Population Migration in Russia: Using an Origin to Destination Matrix. *DP Series A, №483*, The Institute of Economic Research, Hitotsubashi University, Tokyo, 2006b.

Oschepkov, A.Y., 2008. Mejregionalnaya migraciya: ravnovesniy podhod. SU-HSE. (unpublished).

Paas, T., 2000. Gravity Approach for Modeling Trade Flows between Estonia and the Main Trading Partners. Tartu University Press.

Regiony Rossii, 2009. [The Regions of Russia, 2009.] Moskva: Rosstat. CD-ROM version.

Regiony Rossii, 2001. [The Regions of Russia, 2001.] Moskva: Goskomstat. CD-ROM version.

Stock, J.H. and Watson, M.W., 2006. Heteroskedasticity-Robust Standard Errors for Fixed Effects Panel Data Regression. *NBER Technical Working Paper 323*.

Tatevosov, R.V., 1973. Issledovanie prostranstvennih zakonomernostey migracii naseleniya. Statistika migracii naseleniya. Moskva: Statistika.

Verbeek, M., 2000. A Guide to Modern Econometrics. John Wiley & Sons, Ltd.

Appendix A. Variables and their descriptive statistics in the analyzed sample.

Variable	Definition	Number of observation	Mean	Std. Dev.	Min	Max
Population						
Migration ij	Number of people migrated from region i to region j in a given year	90090	171.6384	550.2828	0	31291
Population	Average number of population per year	114114	1854319	1577652	49892	1.05e+07
Urban	Urban population, %	114114	69.30	12.49	23.6	100
Young	Share of 0-16 age group	114114	20.95	4.62	12.3	37.1
Labor market						
Unemployment	Unemployment rate (ILO definition), %	89859	9.97	4.79	0.77	31.97
Tension	Number of unemployed per one	102102	10.74	24.23	0.3	467.5

Variable	Definition	Number of observation	Mean	Std. Dev.	Min	Max
coefficient	vacancy					
Wage	Average monthly wage (regionally adjusted)	76596	1.74	1.18	0.0007	7.27
Income	Average monthly income (regionally adjusted)	82602	1.41	1.10	0.0007	6.06
Grwage	Growth rate of real wage, percent	85176	104.41	14.83	52	152.1
Grincome	Growth rate of real income, percent	91260	104.64	12.95	62.1	172
Unprofitable enterprises	Share of unprofitable firms, percent	103350	39.54	14.30	1.9	85.5
Real estate						
New flats	New flats constructed	114037	7231.47	8979.07	8	83026
Housing affordability	The ratio of average housing prices of 1 square meter to the average monthly income per capita	74459	3.32	1.26	0.36	11.16
Housing availability	Availability of dwellings per capita in square meters at the end of the year	90090	19.78	2.76	12.1	30.8
Characteristics of quality of life						
Infant mortality	Infant mortality rate	114114	15.38	4.87	4.28	42.1
Life expectancy	Life expectancy at birth, years	114114	65.54	2.88	53.76	74.37
Health care workers	Number of health care workers per 10000 residence	114114	44.36	10.18	26.8	95.3
Hospital bed complement	Hospital beds per 10 000 residence	114114	126.64	22.46	68.1	252.4
Crime	Number of registered crimes per 10 000 residents	113421	1965.49	678.15	430	4941
Pollution	Air pollution emitted by stationary structures (tons per square kilometer)	112651	285.01	498.89	1	4179
Student	Number of students, thousands	107800	61.21	112.34	0	1312.7
Mobile phones	Number of registered mobile phones, thousands	52822	1422.49	3492.52	0.1	33961.8
Geographic characteristics						
Distance	Distance between two regional centers, km	114114	3628.08	3103.93	50	13621

Appendix B. Estimated coefficients of the gravity model.

Variable[1]	Distance (≤500 km), 2002 year		Distance (≤500 км), 2008 year		Distance (≥10000 км), 2002 year		Distance (≥10000 км), 2008 year	
	Coeff.	**St. Error**	**Coeff.**	**St. Error**	**Coeff.**	**St. Error**	**Coeff.**	**St. Error**
Population								
Population i	3.910***	1.167	4.374***	1.191	3.755	2.554	3.801	2.666
Population j	-1.541	1.378	-1.495	1.400	3.598	2.294	3.839	2.571
Urban i	0.649	0.500	0.692	0.535	1.012	0.894	-0.530	1.136
Urban j	-1.003**	0.487	-1.434***	0.462	-0.035	0.857	-0.763	1.037
Young i	0.233	0.669	0.610	0.710	0.307	2.626	0.622	1.883
Young j	0.632	0.586	0.502	0.619	-0.068	1.562	-3.503**	1.487
Labor market								
Unemployment i	-0.008	0.030	-0.015	0.033	-0.037	0.146	0.121	0.128
Unemployment j	0.049	0.035	-0.000	0.032	-0.045	0.134	0.019	0.099
Tension coefficient i	-0.024	0.024	0.015	0.027	-0.172**	0.085	-0.104*	0.059
Tension coefficient j	-0.026	0.024	-0.106***	0.024	0.110	0.076	0.051	0.093
Δwage	-0.091	0.094	-0.364**	0.158	0.349	0.333	-0.776**	0.392
Δincome	-0.044	0.081	0.145	0.112	-0.180	0.165	0.868**	0.365
Grwage i	0.190	0.230	0.139	0.271	0.018	0.695	-0.179	0.692
Grwage j	0.268	0.208	-0.246	0.287	1.111***	0.357	-0.982	0.854
Grincome i	-0.091	0.197	0.428***	0.154	1.001	0.973	0.144	0.665
Grincome j	-0.212	0.196	0.335*	0.172	-0.096	0.309	0.246	0.482
Unprofitable enterprises i	0.177**	0.070	0.060	0.074	0.493*	0.270	0.312	0.200
Unprofitable enterprises j	-0.037	0.084	-0.156**	0.076	-0.022	0.263	-0.879***	0.128
Housing market								
New flats i	0.026	0.034	-0.035	0.035	0.111	0.081	0.200***	0.059
New flats j	0.091**	0.041	-0.075**	0.033	0.128**	0.062	-0.041	0.067
Housing affordability i	-0.002	0.049	0.051	0.059	-0.028	0.107	0.104	0.130
Housing affordability j	0.014	0.054	-0.231***	0.048	-0.102	0.132	-0.240	0.197
Housing availability i	-0.284	0.696	-0.105	0.609	-4.459**	1.820	-2.824*	1.510
Housing availability j	2.693***	0.617	2.880***	0.509	-0.172	2.651	-1.390	2.374
Quality of life								
Life expectancy i	0.232	0.978	-0.987	0.862	4.858*	2.932	1.200	2.325
Life expectancy j	1.068	0.987	0.518	0.865	2.147	2.262	3.615	3.008
Hospital bed complement i	0.020	0.134	0.239*	0.128	0.619	0.541	0.462	0.342

[1]All variables are in logarithms. All independent variables are lagged one year. Significance levels: *** −1%, ** −5%, * −10%. Index 'i' denotes source region and 'j' denotes destination.

Variable[1]	Distance (≤500 km), 2002 year		Distance (≤500 км), 2008 year		Distance (≥10000 км), 2002 year		Distance (≥10000 км), 2008 year	
	Coeff.	**St. Error**	**Coeff.**	**St. Error**	**Coeff.**	**St. Error**	**Coeff.**	**St. Error**
Hospital bed complement j	0.472***	0.131	0.210	0.142	1.105*	0.586	1.120**	0.482
Health care workers i	0.344	0.266	0.209	0.273	-0.971	0.696	-0.369	0.639
Health care workers j	0.560**	0.279	0.644**	0.295	-0.141	0.803	0.038	0.711
Infant mortality i	0.005	0.071	0.092*	0.051	-0.115	0.243	-0.069	0.140
Infant mortality j	0.138*	0.078	-0.208***	0.049	-0.180	0.135	-0.085	0.123
Students i	-0.056	0.060	-0.172**	0.070	0.079	0.190	-0.042	0.216
Students j	-0.004	0.048	0.135*	0.079	0.159	0.164	-0.020	0.268
Crime i	-0.050	0.051	-0.033	0.047	-0.112	0.173	-0.278	0.173
Crime j	0.127**	0.058	0.056	0.050	-0.092	0.190	0.208	0.155
Pollution i	0.038	0.038	0.026	0.037	-0.515***	0.124	-0.360***	0.137
Pollution j	0.105***	0.032	0.131***	0.034	0.196	0.129	0.169	0.130
Mobile phones i	0.049***	0.014	-0.062*	0.034	0.055	0.034	0.087	0.077
Mobile phones j	-0.024**	0.012	-0.026	0.029	0.099***	0.034	0.171*	0.094
Trend	0.433	1.304	0.433	1.304	5.838**	2.622	5.838**	2.622
Constant	-51.367*	27.685	-48.767*	26.351	-130.094***	35.379	-95.068**	37.719
R2-within	0.32				0.25			
R2-between	0.0021				0.38			
R2-overall	0.0021				0.35			
Number of observations (NT)	2282				1525			

Notes:
*** – p-value<0.01, ** – p-value<0.05, * – p-value<0.1.

Appendix C. Visual test for specification errors.

Notes. Scatter plots do not reveal substantial departures from linear relationship between predicted and observed values, which is an argument in favor of the chosen specification of regression equation. Decreasing dispersion in Figure 3 may appear due to heteroscedasticity or unobserved heterogeneity between regional pairs.

Figure 2. Specification test for nearby regional pairs (<500 km between regional centers)

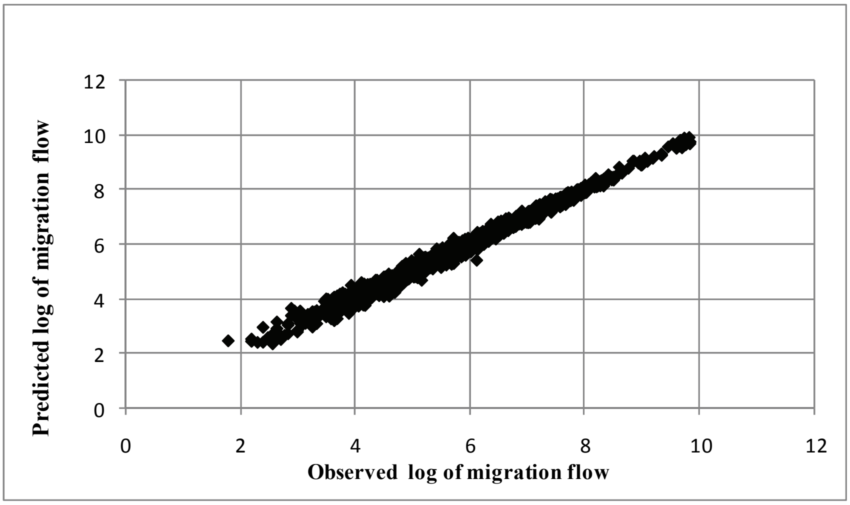
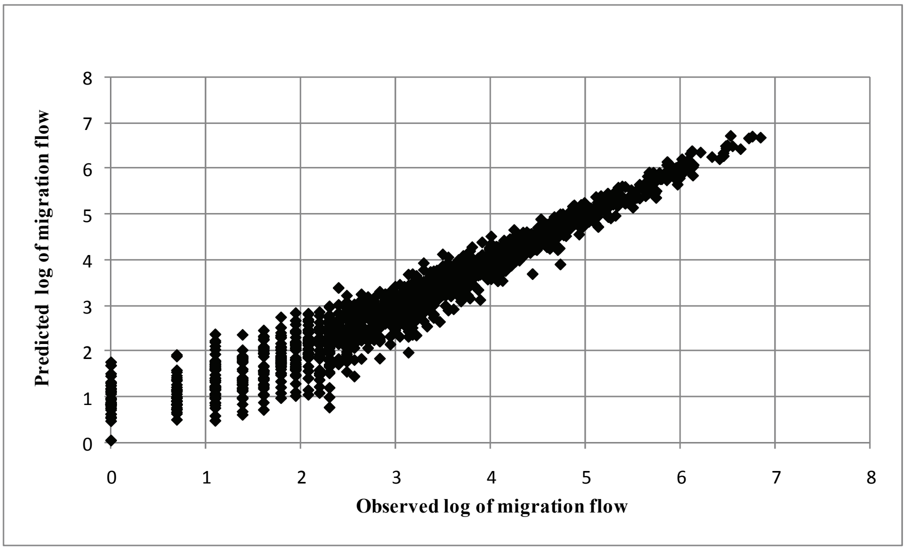


Figure 3. Specification test, distant regional pairs (>10000 km between regional centers).



Zaključak : *Naš model ima jedan bitan nedostatak: preopterećen je objašnjenjima i varijablama. Veliki dio procijenjenih koeficijenata otežava interpretiranje dobijenih rezultata. Značaj nekih od koeficijenata može da bude posledica statističkih grešaka tipa I, što je čest slučaj kod ispitivanja višestrukih hipoteza. Možda bi se ovaj problem mogao riješiti korišćenjem intergralnih indikatora tržišta rada, nekretnina itd., umjesto velikog broja varijabli. Međutim, to će biti predmet nekog daljeg istraživanja, a sada ukazujemo na rezultate koji su se pokazali kao najinteresantniji i najpouzdaniji. Prije svega, sve determinante migracionih tokova su se značajno izmijenile tokom perioda 2002.-2008. Ovo je u skladu s ogromnom prednošću koju model sa dinamičkim koeficijentima ima u odnosu na onaj kod kojega se pretpostavlja da su koeficijentni konstantni (vidi tabelu 2). Data prednost se zapaža u svim grupama regionalnih parova. Drugo saznanje je da socijalni i ekonomski faktori utiču na migracije uglavnom između susjednih regiona. Naši pokušaji da utvrdimo model tokova između udaljenih regiona (od 500 do 8000 km) rezultirali su gotovo neupotrebljivim vrijednostima izmjerenim R^2-unutrašnjim koeficijentom. Poslije 8000 km R^2-unutrašnji koeficijent se ponovo povećava, ali su migracioni tokovi pri takvim distancama oskudni i ne doprinose puno ukupnim međuregionalnim migracijama.*