NATIONAL RESEARCH UNIVERSITY
HIGHER SCHOOL OF ECONOMICS

Vera Ivanova

How space channels wage convergence: the case of Russian cities

BASIC RESEARCH PROGRAM

WORKING PAPERS

SERIES: ECONOMICS

WP BRP 120/EC/2015

Vera Ivanova[1]

How space channels wage convergence: the case of Russian cities[2]

Abstract

Existing empirical work on the growth of Russian regions mostly covers a short time period and considers only regional-level data, while city-level spatial data of the post-reform era remain largely ignored. Using city-level geo-coded data covering 997 cities and towns from 1996 until 2013, I find sigma- and beta-convergence across Russian cities in wages. City wages during the period under consideration display significant and positive spatial autocorrelation. Spatial Durbin models of the Barro regression are estimated using Markov chain Monte Carlo methodology. Estimates of the spatial models for different weight matrices indicate that the city wage growth is significantly affected by wage growth rates in neighboring cities, after conditioning on initial wages.

Keywords: Russian cities, wages, convergence, spatial autocorrelation, spatial econometrics, Markov chain Monte Carlo.

JEL codes: R12; O18.

---

[1]National Research University Higher School of Economics, Center for Market Studies and Spatial Economics, research fellow. E-mail: verunina@gmail.com, viivanova@hse.ru

[2]I owe special thanks to Kristian Behrens, Philip Ushchev, and an anonymous referee for detailed comments and suggestions.

I am grateful to Tatiana Mikhailova, Eirini Tatsi, Jeffrey Wooldridge, Giuseppe Arbia, Jacques Thisse, Florian Mayneris, Natalia Volchkova, Leonid Limonov, Elena Vakulenko, Olga Demidova, Evgenia Kolomak, participants of the IX World Conference of the Spatial Econometrics Association, 2nd Annual Conference of the IAAE, 55th Congress of the ERSA, 13th International Workshop Spatial Econometrics and Statistics, 2014 Asian Meeting of the Econometric Society, 15th International Meeting of the APET, 2nd International Conference on Applied Research in Economics, International conference "Modern Econometric Tools and Applications", and seminars of HSE CMSSE.

I acknowledge the financial support from the Government of the Russian Federation under the grant 11.G34.31.0059.

1 Introduction

Empirical evidence shows that the economies of different regions grow at different speed, and their growth rates are probably spatially correlated (Abreu et al., 2004). Russia is a large and heterogeneous country, so it is natural to conjecture that the role of its geography is essential. Are the economic differences between Russian spatial units due to a more favorable geographic location? Is location a source of local economic advantage? Do spatial spillovers across Russian regions impact substantially on regional growth rates? To answer these questions, the literature on spatial determinants of regional economic growth in Russian regions uses the subjects-of-Federation level data. This approach, however, is problematic, for regions are highly heterogeneous. Moreover, because the regions are few, this creates serious limitations in using advanced econometric methods in order to estimate the relevant effects. In this paper, I suggest a way of overcoming these difficulties.

Using disaggregate geo-coded data, we study whether spatial spillovers across Russian cities foster convergence in wages, and quantify the impact of cities' location patterns on the wage β-convergence process. Thus, although the problem of identifying spatial determinants of regional economic growth in Russia has already been addressed in the literature, we deviate from previous work based on subject-of-Federation-level data by working with city-level data. This entails having the advantage of exploiting a finer location pattern, and creating more room for applying modern spatial econometric tools, which allows us to design an empirical strategy that yields robust inference and displays enough flexibility. First, we estimate spatial econometric models of city wage convergence under three different types of spatial weight matrices. Second, we apply Bayesian spatial econometric models that allow comparison of the estimation results for different spatial weight matrices. The flexibility of our econometric procedure is also due to the possibility of fine-tuning the degree of the spatial weight matrix *sparsity*, a property that has been argued by LeSage and Pace (2009) to be of paramount importance in modern spatial econometrics. To the best of our knowledge, no similar setting has ever been used in previous empirical work on Russian regional development.

Before proceeding, it is worth discussing why we should expect at all that propagation of spillovers over space should affect convergence of urban economies over time. The impact of spatial spillovers on convergence or divergence may be intuitively explained as follows.

Think of a city which is in a very bad economic condition (e.g. it has a very low per capita income). Hence, the current level of city economic performance is below the steady-state level. Consequently, given that positive spatial spillovers are at work, they will foster income growth in the unsuccessful city, which will result in a higher speed of convergence. On the contrary, under negative spatial spillovers, the relatively high income levels in the neighboring cities will lead to slower convergence of the poor city's income to its long-run level. The converse argument works for prosperous cities. To sum up, positive (or negative) spatial spillovers generate a convergence- (or divergence-) enhancing force.

My main findings can be summarized as follows. First, I find that Russian cities beta-converge in wages. Taking the inter-cities spatial spillovers into account does not lead to impressive changes in the convergence speed estimates. However, a dramatic consequence of considering the spatial dimension of convergence is that it results in a striking reduction of the half-life period. Using the Markov chain Monte Carlo methodology, I find that the half-life period varies from 30 to 100 years (depending on the choice of spatial weight matrix) versus around four hundred years in the standard non-spatial convergence estimation setting. Second, I study the dynamics of the spatial autocorrelation index during the period 1996–2013 and explain its non-monotone behavior over time. I also show that the scope of the spatial externalities is almost fully captured by assuming the radius of the spatial interaction being around 1,500 km. Third and lastly, I provide evidence for sigma converge of Russian cities in wages, i.e. the decay of wage dispersion across cities over time.

The rest of the paper is organized as follows. Section 2 reviews the literature, which is mostly empirical studies based on Russian regional data. In Section 3, I describe my dataset and test whether Russian cities converge in wages in terms of sigma- and beta-convergence. In Section 4, I construct different spatial matrices and calculate spatial autocorrelation indices. I show that spatial autocorrelation is significant for city wages. Section 5 tests whether the role of space is significant for convergence. Estimates of spatial Durbin models of the city wage growth equation are provided. Based on Bayesian methodology, I compare models with different spatial weights. Section 6 concludes.

2 Literature review

In a seminal paper on empirical convergence studies, Baumol (1986) found that poorer countries like Japan and Italy have substantially reduced the per capita income gap with richer countries like the United States and Canada in the years from 1870 to 1979. Growth rates are positively correlated with the initial gap between per capita income of a region and the steady-state per capita income level, which is the same for all regions. Regions on the steady growth trajectory are characterized by constant growth rates of per capita income. According to the model, poor regions should grow at a higher pace than wealthy regions, therefore the long-run perspective should tend to smooth regional differences in economic development.

There is extensive empirical literature on convergence using different datasets, explanatory variables and methods (including Barro 1991, Barro and Sala-i-Martin 1992, Sala-i-Martin 1996, Williamson 1996, Taylor 1999, etc). Combining the baseline growth model with the fundamentals of the new economic geography, regional science literature has introduced spatial dependence in the growth regression model. The earliest studies on spatial growth models include Armstrong (1995), Bernat (1996), Fingleton and McCombie (1998). Thorough reviews on spatial growth studies have been provided by Rey and Montouri (1999), Arbia (2006), Fingleton and López-Bazo (2006), Ertur et al. (2006).

Empirical work on Russian regional income growth cover vary in the length of time horizon. Extensive reviews on Russian regional income inequalities are provided by Glushchenko (2010, 2012), Guriev and Vakulenko (2012). Convergence hypothesis is rejected in papers based on the earliest post-Soviet time period data, while in more recent studies there is growing evidence of Russian regional convergence. Results also depend on the key indicator chosen for the convergence studies. While interregional GDP per capita gaps persist, the differentials in incomes and wages have decreased substantially. The earliest papers on regional income inequalities ignore any spatial localization of Russian regions and do not exploit spatial analysis. There is, however, growing evidence that space is crucial in regional income studies.

In general, it is fair to say that evidence provided by studies of regional income convergence in Russia is inconclusive (see Table 7 in Appendix 1 for a summary). It is also worth noting that very few papers consider two or more types of spatial matrices. Demidova (2015) tests the hypothesis of a possible difference in the spatial effects in western and eastern regions. She develops a special class of empirical growth models with a partitioned spatial matrix (west-west, east-east, west-east, and east-west) for Russian regions over the period 2000–2010. Her estimates for spatial externalities depend on the referent variable and reveal the asymmetric influence of eastern and western regions on each other. Ivanova (2014) estimates the Barro regression for five different spatial matrices and tests whether a spatial autoregressive parameter is sensitive to the spatial weights choice.

The choice of spatial weights is crucial in spatial econometrics. Harris et al. (2011) provide an extensive review of the standard approaches in constructing the spatial weights matrix. There is an overview of spatial models comparison procedures in LeSage and Pace (2009, Chapter 6). They consider spatial versus non-spatial models, models with different spatial weights, and spatial models with different sets of explanatory variables. In this study, I exploit methodology for comparing models with different spatial weight matrices based on a Markov chain Monte Carlo model composition (MC^3) approach proposed for spatial regression models by LeSage and Parent (2007).

3 Do Russian cities converge in wages?

3.1 Data description

Convergence studies are based on various regional per capita income variables; like gross domestic product, gross value added, or total income. The Russian Statistical Agency does not provide city level GDP per capita, and there is no available information on city-level income per capita. The only available proxy for per capita income is the average monthly wage in Russian cities provided by the Multistat database[3]. The database contains 1098 Russian cities and towns. I omit settlements of the Chechen republic, the Ingushetia republic and observations with missing values on wages. As a consequence, the number of cities of the main dataset for my study is 997, and they cover 78 regions of Russia. Summary statistics for nominal wages are presented in Table 1[4].

Table 1: Summary statistics for nominal wages

Year	Mean	Standard deviation	Min	Max
1996	748.68	452.29	164	4049
1997	890.45	549.07	105	5074
1998	979.63	577.86	300	5247
1999	1390.52	846.28	350	7628
2000	2055.08	1543.64	624	20284
2001	2870.15	1951.44	844	17012
2002	3872.18	2325.14	1230	19801
2003	4788.29	2831.43	1791	24347
2004	5827.33	3294.95	2200	29116
2005	7237.62	4113.06	895	48965
2006	9230.72	7000.58	1326	173403
2007	11340.14	5615.04	1544	44958
2008	14424.36	7086.00	4915	80299
2009	15751.20	7216.92	5830	57746
2010	17426.67	8000.40	6767	66024
2011	19653.85	8973.56	7957	72551
2012	22732.45	9977.81	9640	78819
2013	25736.91	10763.89	12557	84410

[3] www.multistat.ru

[4] Wages in 1996 and 1997 are recalculated (divided by 1000) into rubles redenominated in 1998.

I convert nominal wages into real (spatially comparable) ones. I use the regional prices of a fixed set of goods and services for adjusting wages by PPP as follows. The real wages measured in 2010 rubles are obtained by deflating the nominal values of wages by the cost of a fixed basket of goods and services in the regions (subjects of Federation). Data for the cost of a fixed basket of goods and services in the regions has been directly available since 2002[5], so I use the regional consumer price index (CPI) as a proxy for the basket cost in earlier years.

Next, I geo-coded the dataset with geographical latitudes and longitudes for each observation. Spatially comparable wages in cities are shown in Fig.1. One can see that some dark points (i.e. cities with high wages) correspond to cities with large population: Moscow, Saint-Petersburg, Ekaterinburg. These cities seem to generate clots of cities with big wages. It is worth mentioning, however, that not only big cities are rich. Cities in oil and gas regions (Tyumen oblast, Saha (Yakutia) republic etc.) also display high wages.

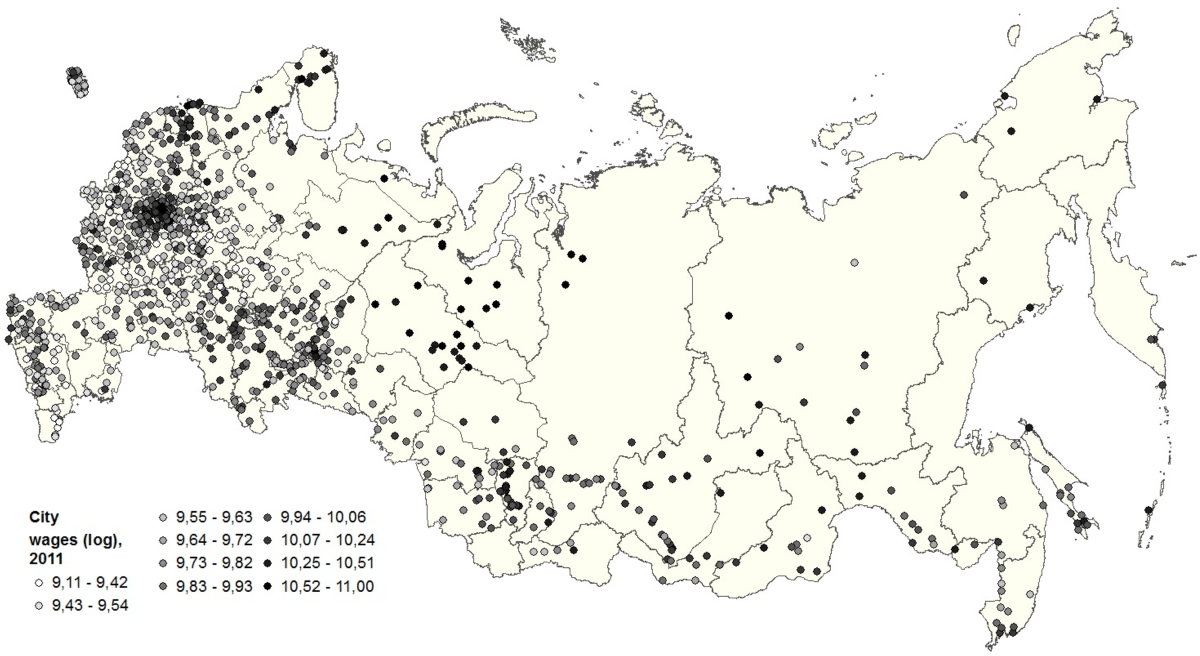

Figure 1: Real wages in Russian cities, 2011 (in 2010 rubles, log).

3.2 Convergence methodology

I study whether city level sigma- and beta-convergence of wages holds. The concept of sigma-convergence refers to a reduction of wage dispersion, while beta-convergence means that poorer cities grow faster than the richer.

I test the hypothesis of sigma-convergence using a coefficient of variation for real wages and performing a unit-root test for a time-detrended panel of relative city wages. The hypothesis of beta-convergence is tested based on estimates of the Barro regression.

[5] www.gks.ru

3.3 Sigma-convergence

Consider variation of real (spatially comparable) city wages. Real wages increase over time, so I compute the coefficient variation (a ratio of the standard deviation to the mean) in order to exclude the scale effect. The dynamics of the coefficient variation of log real wages is depicted by Fig. 2. Unified social tax with regressive taxation is introduced in 2001 in Russia, and this may cause the change in the dynamics. Variation of log real city wages is decreasing, which suggests that sigma-convergence of cities holds.

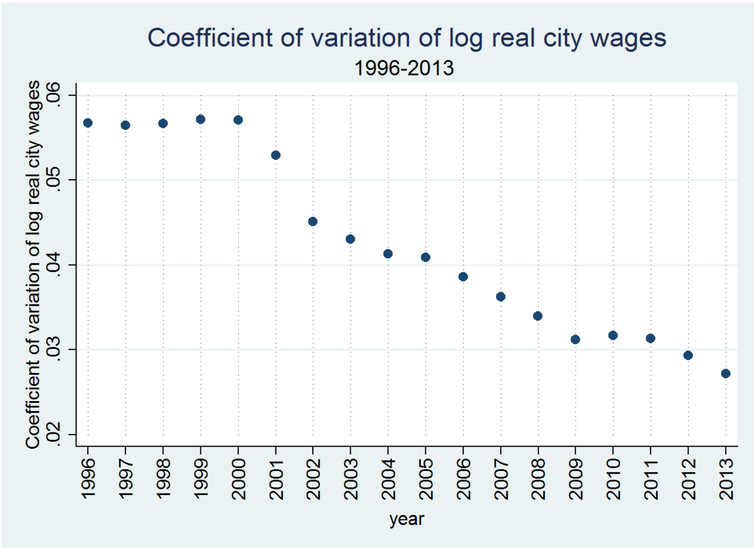

Figure 2: Coefficient of variation for log real wages, 1996–2013.

In order to exclude the time trend from the real wages, I compute relative real city wages as a ratio of a city wage to a base, where the base is the national average wage.

Consider the model (1) with a first-order autoregressive component

$$z_{i,t} = bz_{i,t-1} + \gamma_i + \varepsilon_{i,t}, \tag{1}$$

where $z_{i,t}$ - relative real wage (log) in city $i = 1, ..., 997$ in year $t = 1996, ..., 2013$, γ_i – individual fixed effects, $\varepsilon_{i,t} \sim iid$. I exploit Harris-Tsavalis panel unit root test (1999) which is based on the OLS estimator. The null hypothesis $H_0 : b = 1$ (panels contain unit roots) versus the alternative $H_a : |b| < 1$ (panels are stationary). Estimates are as follows: $\hat{b} = 0.561$, z statistics is -54.439, p-value< 0.0001, so I strongly reject the null hypothesis of a unit root, finding support for sigma-convergence in wages across cities.

3.4 Beta-convergence

The test of beta-convergence hypothesis is based on the cross-sectional Barro regression:

$$y_i = \alpha + \beta \ln x_{i,0} + \varepsilon_i, \tag{2}$$

$y_i = \frac{\ln x_{i,t} - \ln x_{i,0}}{T}$ is the average annual growth of real wages between 1996 and 2013, $x_{i,0}$ – initial real wages (for the year 1996), $x_{i,t}$ – real wages in 2013, T =17, ε_i is the error term, $\varepsilon_i \sim iid$ with 0 mean and finite 2nd moment.

When β is significantly negative, beta-convergence holds. The dependence of $y_{t+\tau}$ on the initial levels of y_0 disappears for large τ. The time required for this usually is analyzed based on the so-called "half life time to convergence":

$$HL = \tau T = -\frac{\ln 2}{\ln(1+\beta)} T \tag{3}$$

Fig. 3 depicts a scatterplot for real wages in 1996 and average annual growth of wages between 1996 and 2013 (in logs). The slope of the line is negative. The fitted values correspond to OLS estimates of equation (2) provided in Table 2, column 1. The coefficient β is negative and significant, so there is beta-convergence by city wages. Because initial wages in 1996 are measured in 2010 rubles, the result means that an increase in initial monthly wages by 1000 rubles lowers the wage growth by 28 percent per year over the period 1996–2013.

Figure 3: Scatterplot for real wages in 1996 and average annual growth of wages between 1996 and 2013.

Cities are not randomly located in the country's territory. In order to take their location into account, I estimate Barro regression (2) with additional explanatory geographical variables: longitude and latitude of each city (in logs), and the squares of these variables. The estimation results are given by Table 2, column 2.

Table 2: **OLS estimates of the Barro regression**

Variables	Dependent: Average annual growth of wages (log), 1996–2013 (1)	(2)
Constant	0.397***	3.081***
	(0.007)	(0.428)
Wages in 1996 (log)	-0.028***	-0.031***
	(0.001)	(0.001)
Longitude (log)		-0.063***
		(0.013)
Latitude (log)		-1.290***
		(0.216)
Squared longitude (log)		0.008***
		(0.002)
Squared latitude (log)		0.163***
		(0.027)
Adjusted R^2	0.444	0.480
Number of cities	997	997
τ	24.14	21.94

standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The effect of the geographical variables on city wage growth is significant. The estimates of τ yield very long half life time to convergence: 24.14 rounds of 17 years without locational variables (HL=410.38 years) and 21.94 rounds of 17 years with locational variables (HL=372.98 years). In other words, convergence is *tremendously slow*. In the next two sections, I struggle to show that accounting for spatial spillovers renders the estimates for the speed of convergence to become more realistic.

Hence, hypotheses on both sigma- and beta-convergence of Russian cities in wages have been confirmed. Furthermore, locational variables in the Barro regression turn out to be highly significant, i.e. location matters for convergence in wages. This raises a conjecture that city-level wages may be spatially autocorrelated, which we test in the next section.

4 Is there spatial autocorrelation of wages?

I now come to estimating the impact of a city's location pattern on the convergence in wages across cities.

Most researchers studying the growth of Russian regions use contiguity and distance-based matrices. Distances between regions are measured as distances between regional central cities, i.e. such a type of spatial matrix in the case of regions is a rather rough measure of the spatial distribution, and this is more suitable for cities. Furthermore, in such papers distance matrices are based on highway or railway distances, so these spatial matrices cannot be considered as exogenous. In addition, because of the huge and non-uniformly populated territory, regional contiguity matrices in estimating spatial models may lead to significantly different effects in the western and eastern parts of Russia (Demidova, 2015).

To obtain robust inference, I construct 4 types of spatial weights for cities:

1) inverse great circle distances:

$$w_{ij}^{(d)} = \begin{cases} 1/d_{ij} & \text{if} \quad i \neq j, \\ 0 & \text{if} \quad i = j, \end{cases}$$

where d_{ij} – a great circle distance between cities i and j;

2) inverse great circle distances and a cutoff distance C of spatial interaction:

$$w_{ij}^{(d_C)} = \begin{cases} 1/d_{ij} & \text{if} \quad d_{ij} < C, \, i \neq j, \\ 0 & \text{if} \quad i = j, \end{cases}$$

3) contiguity based on a cutoff great circle distance: cities i and j are considered as neighbors if the distance between them is less than a cutoff C:

$$w_{ij}^{(nei_C)} = \begin{cases} 1 & \text{if} \quad 0 < d_{ij} < C, \, i \neq j, \\ 0 & \text{if} \quad i = j, \end{cases}$$

4) contiguity based on a given number N of neighbors: the N nearest (in terms of the great circle distance) cities are considered as neighbors: $w_{ij}^{(N)} = 1$ if $j \neq i$ are N nearest neighbors, and $w_{ij}^{(N)} = 0$ otherwise.

In order to measure the overall degree of wage similarity between spatially close cities, I calculate Moran's global spatial autocorrelation index I :

$$I = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}} \tag{4}$$

where w_{ij} are spatial weights, the (i, j)th element of the $n \times n$ spatial weights matrix W, spatial weights w_{ij} measure the locational similarity of units i and j; x is a variable of interest; and $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$.

Following the literature, I use row-standardized spatial weights, so that $\sum_{j=1}^{n} w_{ij} = 1$ for each i.

Fig. 4 shows changes in the global Moran's I for log wages calculated using row-standardized inverse great circle distance matrices with different cutoff distances. All the indices are positive and significant at the 5% significance level, hence, city wages are spatially autocorrelated.

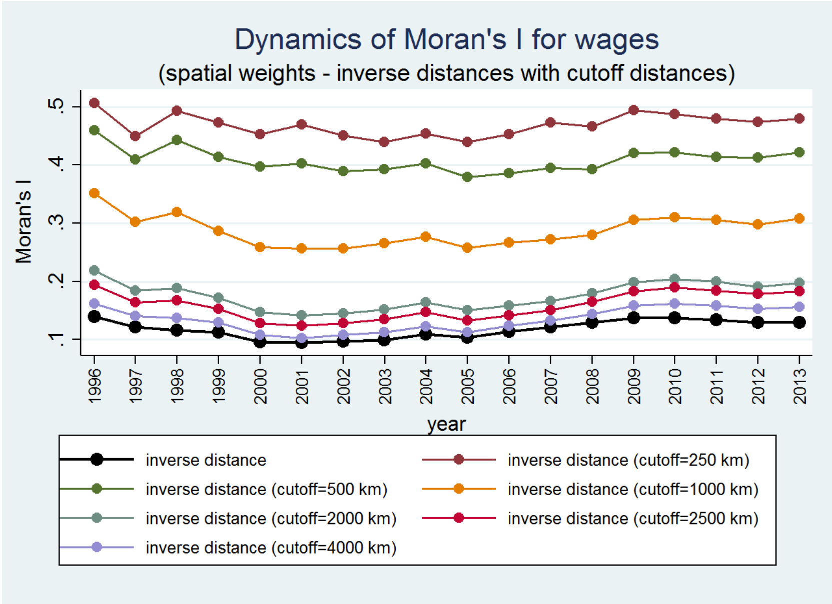

Figure 4: Moran's I for real wages (log), inverse distance matrices with cutoffs.

Also I compute Moran's index using row-standardized contiguity distance matrices with different cutoff distances (Fig. 5). As in Fig. 4, all the spatial indexes are 5% significant and positive. The territory of Russia is huge, so it should not come as a surprise that the measure of spatial autocorrelation becomes higher when geographically remote cities' pairwise spatial weights are considered as zero. Changes in Moran's I for inverse distances over the period 1996–2013 are highlighted in Fig. 5. As we can see, most similar dynamics corresponds to neighbors within a radius of 1500 km. So, the radius of 1500 km can be considered as the neighbors' radius capturing almost all possible interactions in terms of spatial autocorrelation of real city wages.

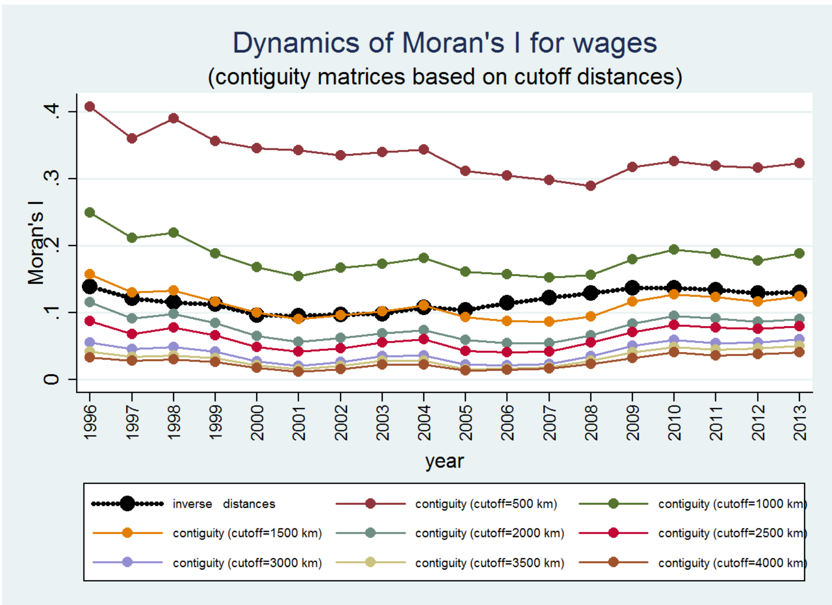

Figure 5: Moran's I for real wages (log), contiguity weight matrices with cutoffs.

The fourth type of spatial matrix depending on N nearest neighbors is usually applied in studies of spatial autocorelation for US states and European regions. There seem to be no empirical papers on Russian regional data that analyze regional spatial interactions of regions based on a given number of neighbors. Moran's I for wages with spatial weights defined by the number of neighbors decreases while number of neighbors becomes greater (Fig. 6).

There are several local minimal and maximal values of the global spatial autocorrelation index, and these correspond to different values of the year variable. We can find local maximums in financial crisis years (1998 and 2008) or a year later in all three plots with Moran's I dynamics.

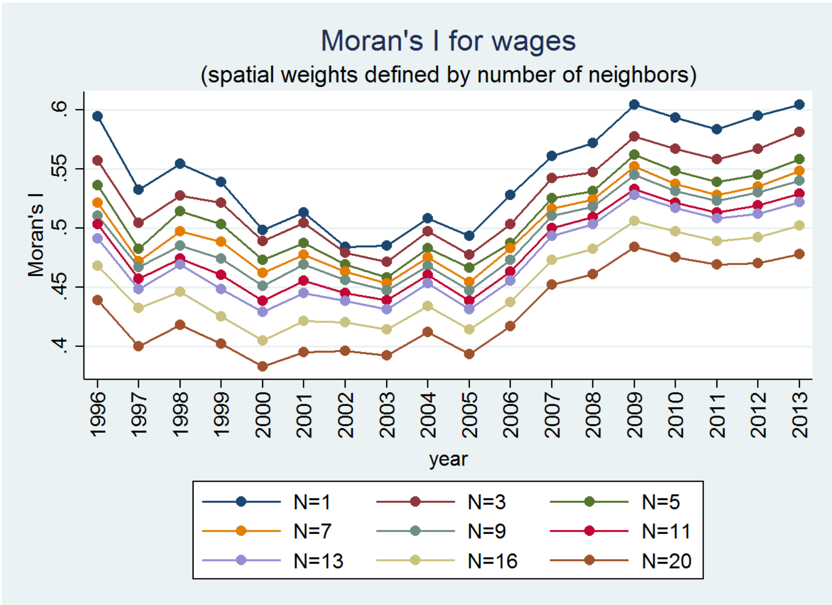

Figure 6: Moran's I for real wages (log), weights based on number of neighbors.

All the three plots display significant and positive spatial autocorrelation of real city wages. Hence, ignoring possible spatial links of observations may lead to biased estimates of the Barro regression. Therefore, in the next section, I introduce a spatial component into the Barro regression.

5 Spatial convergence

5.1 Spatial model specification of the Barro regression

Before choosing between various spatial regression models, I run spatial diagnostic tests (Moran's test, Lagrange multiplier error test and Lagrange multiplier lag test) for spatial dependence in OLS residuals. These tests show significant spatial autocorrelation of residuals for most of the spatial matrices constructed in Section 4. There is a criticism of choosing a specification of the spatial model based only on formal diagnostic tests (LeSage and Fischer, 2008). The Barro regression (2) includes the only explanatory variable – the initial level of income – while wage dispersion cannot be explained solely by the initial levels of wages. Hence, we face the problem of omitted explanatory variables. Moreover, statistics on Russian cities are insufficient for constructing the standard bunch of variables tested in conditional beta-convergence models. The aim of the paper is to find spatial matrices that could explain spatial interactions in the Barro regression in the case of omitted variables. LeSage and Fischer (2008) argue that the SDM is a natural choice over competing alternatives of spatial models in the case of missing explanatory variables. Following LeSage and Fischer (2008), I include a spatial lag of the dependent variable on the right hand side of the Barro regression in order to account for missing explanatory variables. Also I include a spatial lag of initial values of wages. So, the regression to be estimated is **a spatial Durbin model (SDM)**:

$$y_i = \alpha + \beta \ln x_{i,0} + \rho \sum_{j=1}^{n} w_{ij} y_j + \theta \sum_{j=1}^{n} w_{ij} \ln x_{i,0} + \varepsilon_{i,t}, \tag{5}$$

where $y_i = \frac{\ln x_{i,t} - \ln x_{i,0}}{T}$,

w_{ij} are spatial weights, $i, j = 1, ..., n$, number of cities $n = 997$,

ρ is a spatial autoregressive parameter, $|\rho| < 1$ for conventional row-stochastic matrices,

θ is a spatial parameter of lagged initial year wages,

ε is a vector of homoscedastic and uncorrelated errors.

One more reason for considering the SDM is that the model (5) is a global spatial spillover specification. The spatial spillover arises when the value of variable x of the observation i in space has a significant impact on the outcomes y_j of an observation located at position j. Global spatial spillovers mean that there is an endogenous interaction and a spatial feedback effect (LeSage, 2014). Endogenous interaction implies that changes in one observation drive a series of changes in all regions in the sample such that a new long-run steady state equilibrium arises. In other words, equation (5) yields a simple way of capturing the general-equilibrium effects. For example, high wages in city i may attract migrants from neighboring cities j and k, so the employment rate in i will change, and employers may offer lower wages in i, while in cities j and k the employment rate will decrease, ceteris paribus, which may lead to higher salaries. Neighbors of j and k may react to new wages in the same scenario, and this will lead to a global spatial feedback effect.

Because of the presence of the spatial lag of the dependent variable in (5), coefficients of explanatory variables in the SDM cannot be interpreted as partial derivatives which describe the magnitude of changes in the dependent variable that arise from changes in explanatory variables. LeSage and Pace (2009) introduce direct, indirect and total effects for the correct interpretation of coefficients in spatial models (see Appendix 2).

What about the expected signs of impact? A change in the initial level of the wage of city i will result in: 1) a direct impact on the growth rate of city i, 2) an indirect impact arising from spatial relationships with cities $j \neq i$, because of the presence of the spatial lag. When the wage in city i increases, i.e. the population of city i becomes richer, then, according to beta-convergence hypothesis, it must follow that the wage in city i grows slower.

Hence, the expected sign of the direct impact is negative. The predicted sign of the indirect impact that comes from changes in wages in neighboring cities is negative, because when neighboring cities become richer, then city i gets poorer than the neighboring cities, and beta-convergence suggests that it grows faster.

Half life time to convergence in a spatial version of the Barro regression depend on not only the coefficient of the initial income. LeSage and Fischer (2008) provide a formula for half life time to convergence in the case of the SDM

$$\tau = \frac{\ln 2}{\ln \lambda} \tag{6}$$

where $\lambda = (1 + \beta - \rho)/(1 - \rho)$. Coefficient τ stands for a number of rounds of T years to reach a half-way point of the path to the new equilibrium steady-state, so half-life time in years is $HL = \tau T$.

5.2 Estimates and interpretations

Estimates of the SDM with different spatial matrices are in Tables 3, 4, and 5. The estimation method that I use is Markov chain Monte Carlo model composition methodology (MC^3) introduced by Madigan and York (1995) and extended by LeSage and Parent (2007) to the case of SDMs (see Appendix 3 for details of the estimation procedure). Posterior probabilities of each model are provided in Tables 3, 4, 5, and can be considered as weights in a simple linear combination of estimates presented in each table for constructing model averaged estimates (LeSage and Fischer, 2008).

Table 3: **Estimates of SDM, weights - inverse distances with cutoffs**

	Dependent: Annual growth of wages, log			
Variables	(1)	(2)	(3)	(4)
Constant	0.026**	0.016	0.010	0.008
	(0.016)	(0.016)	(0.015)	(0.015)
Wages in 1996 (log)	-0.0334***	-0.0334***	-0.0333***	-0.0332***
	(0.0012)	(0.0012)	(0.0012)	(0.0012)
ρ	0.879***	0.896***	0.906***	0.909***
	(0.034)	(0.033)	(0.032)	(0.032)
θ	0.033***	0.034***	0.035***	0.035***
	(0.002)	(0.002)	(0.002)	(0.002)
Direct effect	-0.033***	-0.033***	-0.033***	-0.033***
Indirect effect	0.032***	0.041**	0.050**	0.055**
Total effect	-0.001	0.008	0.017	0.022*
τ	2.15	1.79	1.58	1.53
Spatial weights	$w^{(d_{650})}$	$w^{(d_{700})}$	$w^{(d_{750})}$	$w^{(d_{800})}$
Model probabilities	0.0255	0.6573	0.3097	0.0070

standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Estimates of the SDM are presented in Table 3. Spatial weights are inverse great circle distances with cutoffs. I estimate (5) with distance cutoffs 650, 700, 750, ..., 2000 km[6], and inverse distances without cutoffs. In Table 3, estimates are provided for models with posterior probabilities greater than 0.0001.

The convergence parameter β is negative and significant, as well as in the non-spatial case. Estimates of β are robust to the spatial cutoff distance, and this is consistent with LeSage and Pace (2009). I get positive and significant estimates of the spatial parameter ρ for the SDM with different spatial matrices introduced above, therefore the city wage growth is related with the growth rates of neighboring (in terms of the weights matrix) cities, after conditioning on the effect of initial wages $\ln y_{i,0}$. Values of ρ depend on the type of the spatial matrix.

Table 4: **Estimates of SDM, weights - neighbors defined by cutoff distances**

	Dependent: Annual growth of wages, log			
Variables	(1)	(2)	(3)	(4)
Constant	0.116***	0.099***	0.086***	0.088***
	(0.028)	(0.027)	(0.028)	(0.030)
Wages in 1996 (log)	-0.0319***	-0.0317***	-0.0315***	-0.0314***
	(0.0012)	(0.0012)	(0.0012)	(0.0012)
ρ	0.662***	0.705***	0.736***	0.724***
	(0.072)	(0.072)	(0.070)	(0.077)
θ	0.025***	0.027***	0.027***	0.027***
	(0.003)	(0.003)	(0.003)	(0.003)
Direct effect	-0.032***	-0.032***	-0.031***	-0.031***
Indirect effect	0.011*	0.012*	0.013*	0.013*
Total effect	-0.021***	-0.020**	-0.018*	-0.018*
τ	6.99	6.10	5.46	5.74
Spatial weights	$w^{(nei_{650})}$	$w^{(nei_{700})}$	$w^{(nei_{750})}$	$w^{(nei_{800})}$
Model probabilities	0.0335	0.5287	0.4362	0.0017

standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

The direct effects of the initial wages represent the impact of this variable on the growth rates directly as well as the feedback effect of going through neighbors. Magnitudes of direct effects of initial wages are close to the estimates of β, this can be explained by low values of θ.

One of the main results from estimating the spatial Durbin model is that the indirect effect of the initial wages is positive in all the cases (Tables 3, 4, 5), which suggests the existence of a global spatial spillover effect for Russian city wages. Changes in the explanatory variables in the SDM leads to a series of space-time changes that result in a new equilibrium. The time to reach the new equilibrium takes τ rounds of 17 years, which in the case of the

[6] Cutoff values less than 650 km are not considered because of the presence of some remote towns in Sakha (Yakutia) Republic and Krasnoyarsk kray. The nearest observations to those towns are located at a distance more than 600 km. Therefore, spatial matrices with a cutoff distance less than 600 km are unconnected, and they cannot be used in the SDM.

model with the highest posterior probability in Table 3 (cutoff distance 700 km) is equal to $HL = 1.79 \times 17{=}30.43$ years.

Estimates of the SDM with spatial contiguity weights based on cutoff distances are presented in Table 4. In this case, the highest posterior probability model is also defined by a cutoff distance of 700 km, and while also using inverse distances. The time to reach the new equilibrium takes 6.1 rounds of 17 years, which is much greater than for spatial inverse distance weights.

Estimates of the SDM (5) for a given number of neighbors are presented in Table 5. I consider N=1, 2, ..., 20, while estimates of models with posterior probabilities less than 0.01 are omitted. Posterior probabilities in Table 5 indicate that the span of local externalities varies from 11 to 15 neighbors. In terms of distances this number depends largely on the location of the city, because the European part of the country is densely populated, and the 10-15 neighbors may envelop a radius of 100 km, while in the Asian part such distance may be around 500 km or even longer.

Table 5: **Estimates of SDM, weights defined by N nearest neighbors**

	Dependent: Annual growth of wages, log					
Variables	(1)	(2)	(3)	(4)	(5)	(6)
Constant	0.089***	0.086***	0.081***	0.078***	0.071***	0.069***
	(0.013)	(0.013)	(0.013)	(0.014)	(0.014)	(0.014)
Wages in 1996 (log)	-0.0337***	-0.0337***	-0.0338***	-0.0341***	-0.0341***	-0.0341***
	(0.0011)	(0.0011)	(0.0011)	(0.0011)	(0.0012)	(0.0011)
ρ	0.733***	0.742***	0.749***	0.759***	0.770***	0.774***
	(0.029)	(0.030)	(0.030)	(0.031)	(0.031)	(0.031)
θ	0.029***	0.029***	0.029***	0.030***	0.031***	0.031***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Direct effect	-0.033***	-0.033***	-0.033***	-0.034***	-0.034***	-0.034 ***
Indirect effect	0.014**	0.015**	0.016**	0.017**	0.118**	0.019**
Total effect	-0.019***	-0.019***	-0.018***	-0.017***	-0.015 **	-0.015**
τ	5.14	4.95	4.79	4.54	4.32	4.24
Spatial weights	$w^{(11)}$	$w^{(12)}$	$w^{(13)}$	$w^{(14)}$	$w^{(15)}$	$w^{(16)}$
Model probabilities	0.4078	0.0913	0.0776	0.0909	0.2687	0.0421

standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

To sum up, estimates of the spatial Durbin model of the Barro regression are consistent with the expected spatial impact of initial wages on the wage growth. Spatial parameters for spatial lags of both initial wages and growth rates are significant and rather high.

5.3 Spatial sigma-convergence

In the non-spatial case, I test the hypothesis of sigma-convergence and find a reduction of wage dispersion. In order to test whether the neighbors' wage dispersion also decreases, I consider spatial lags of log wages: $v_{i,t} = \sum_{j=1}^{n} w_{ij} \ln x_{i,t}$, where t =1996,...,2013, w_{ij} are row-normalized spatial weights. Wages $\ln x_{i,t}$ increase over time, i.e. involve a time trend. Therefore, I consider a first-order autoregressive component with a time trend:

$$v_{i,t} = bv_{i,t-1} + \gamma_i + \delta_i t + \varepsilon_{i,t}, \quad (7)$$

As in subsection 3.3, I exploit the Harris-Tsavalis panel unit test based on the OLS estimator. The null hypothesis $H_0 : b = 1$ (panels contain unit roots) versus the alternative $H_a : |b| < 1$ (panels are stationary). Estimates for different spatial matrices are provided in Table 6, with p-values<0.0001.

Table 6: **Panel unit root test results**

Spatial weights	$W^{d_{1500}}$	$W^{d_{1000}}$	$W^{d_{2000}}$	$W^{nei_{500}}$	$W^{nei_{1000}}$	$W^{nei_{2000}}$
Estimates of b	0.510	0.478	0.546	0.529	0.526	0.560
z-statistics	-15.831	-20.15	-10.81	-13.22	-13.53	-8.88

Thus, I reject the null hypothesis of a unit root and find support for spatial sigma-convergence in wages.

6 Conclusion

The paper studies sigma- and beta-convergence based on recent Russian city-level geocoded data. I find significant convergence of both types. The effect of the initial wage on the wage growth is rather weak, the coefficient β for the initial wage in the Barro regression is about -0.03 for both non-spatial and spatial cases. Russian city wages display significant and positive spatial autocorrelation, and a radius of 1500 km can be considered as an area of almost all possible spatial interactions of city wages in terms of the Moran's I volatility.

Estimates of the spatial Durbin model are rather robust to the specification of the spatial weights matrix. There is a positive and significant spatial spillover effect of initial city wage level on city wage growth, which is consistent with the choice of spatial regression. Finally, using Bayesian spatial regression, I argue that the most intensive spatial interactions between cities that result in changes of wage growth rates are within the distance of 700 km.

Estimates of half life period of convergence are highly dependent of the form of spatial weight matrix. Because of heterogeneous pairwise distances between cities, wage convergence patterns can be studied, for example, subdividing regions into several groups in the spirit of Kholodilin et al. (2009), introducing spatial weights depending on the part of the country (Demidova, 2015), or other techniques, which may shed light on the nature of spatial interactions between Russian cities.

Appendix 1

Table 7: **Empirical papers on regional income convergence using spatial Russian data**

Paper	Time period	Indicator	Results	Space
Solanko (2003)	1992–2001	Personal income	σ-divergence, β-convergence	Control variable "distance from Moscow"
Berkowitz and DeJong (2003)	1993-1997	Personal income	β-divergence	Control variable "distance from Moscow"
Ahrend (2005)	1990–1998	GRP, personal income, industrial production	β-convergence and β-divergence, depending on the indicator	Regressions for Russia and for its European part
Carluer (2005)	1985–1999	Personal income	β-convergence and β-divergence, depending on the time period. Convergence clubs	Geographical dummies and distance from Moscow
Buccellato (2007)	1999–2004	Personal income	β-convergence	SAR, SEM. Contiguity weight matrix
Lugovoy et al. (2007)	1996–2004	GRP	σ-convergence and σ-divergence, depending on the time period and methodology. Conditional β-convergence	Moran's I. SEM, SDM. Spatial weights: travelling time, inverse distance
Zverev and Kolomak (2010)	1995–2006	GRP, budget revenue, budget expenses	σ-divergence. β-convergence and β-divergence, depending on the indicator	Moran's I. SEM, SAR. Spatial weights: inverse distances, contiguity
Kholodilin et al. (2012)	1998–2006	GRP	Weak β-convergence. Strong β-convergence among high income regions and among low income regions	Moran's I. SEM, SAR. Spatial weights: inverse squared distances
Ivanova (2014)	1996–2012	Personal income	σ-convergence, β-convergence	Moran's I. SAR. Spatial weights: contiguity, inverse great circle distances with cutoffs, inverse railway distances

Notes. GRP – gross regional product. SEM – spatial error model, SAR – spatial autoregressive model, SDM – spatial Durbin model.

Appendix 2

Interpreting estimates of SDM

Consider an SDM

$$y = X\beta + \rho Wy + \theta WX + \varepsilon \tag{8}$$

here X is an $n \times k$ matrix of explanatory variables, β is a $k \times 1$ parameter vector, W is a $n \times n$ spatial matrix, ρ and θ are spatial scalar parameters, ε is an $n \times 1$ disturbance vector. We assume that $\varepsilon \sim N(0, \sigma^2 I_n)$, here I_n is the identity matrix of size n. An intercept parameter in (8) is omitted for simplicity.

The data generating process for this model is given by

$$y = \sum_{r=1}^{p} S_r(W)x_r + (I_n - \rho W)^{-1}\varepsilon$$

$$S_r = (I_n - \rho W)^{-1}(I_n \beta_r + W\theta_r)$$

where x_r is the rth explanatory variable.

Impacts (effects) arise in changes in the rth explanatory variable in the SDM (8).

Consider the impact of changes in x_{ir} (each ith observation of x_r) on the dependent variable y_i of the ith observation. These impacts mean effect of changes in x_{ir} directly to the y_i as well as a feedback effect which passing through neighbors and back to the observation itself. The averaged sum of such impacts is equal to $\frac{1}{n}\mathrm{tr}(S_r)$ and called **the average direct effect of** x_r.

The indirect effect means impacts from changes in all observations $j = 1, 2, ..., n$ of an explanatory variable x_{jr}, $j \neq i$. They are calculated as the sum of the off-diagonal elements of the rows i in S_r for each i. These effects can be considered as a numerical measure of spatial spillovers. **The average indirect effect of** x_r is equal to the average sum of the off-diagonal elements of the matrix S_r.

The average total effect of x_r is calculated as the sum of the average direct effect and the average indirect effect.

Appendix 3

Bayesian model comparison

Consider SDMs (8) with different spatial matrices $W_1, W_2, ..., W_m$ and equal set of explanatory variables, denote the set of SDMs: $M = M_1, M_2, ..., M_m$. Denote prior probabilities of each model $\pi(M_i)$, $i = 1, ..., m$ and prior distributions for the parameters of each model: $\pi(\eta)$, $\eta = (\beta, \rho, \theta, \sigma^2)$, where σ^2 is a constant, scalar noise variance parameter.

Each model is equally likely a priori, i.e. $\pi(M_i) = 1/m$, $i = 1, ..., m$. Let D be the sample data. Likelihood for y is conditional on distributions for the parameters of the model and the set of models M: $p(D|\eta, M)$. The joint probability for the set of models, parameters and data:

$$p(M, \eta, D) = \pi(M)\pi(\eta|M)p(D|\eta, M)$$

Apply the Bayes rule and get the joint posterior probability for both models and parameters:

$$p(M, \eta|D) = \frac{\pi(M)\pi(\eta|M)p(D|\eta, M)}{p(D)}$$

The posterior probabilities regarding the models take the form

$$p(M|y) = \int p(M, \eta|y)d\eta \qquad (9)$$

LeSage and Parent (2007) provided formulas for the marginal posterior in (9) for spatial autoregressive and spatial error models that differ in terms of the explanatory variables matrix X and the fixed type of the spatial matrix. LeSage and Fischer (2009) extended the study for the case of parameters for the range of nearest neighbors in the spatial matrix and type of distance.

The likelihood function for the parameters $\eta = (\beta, \rho, \theta, \sigma^2)$, based on the data $D = \{y, X, W\}$ takes the form

$$L(\eta, y, X, W) \propto (\sigma^2)^{-n/2}|I_n - \rho W|^{1/2} \exp\left\{-\frac{1}{2\sigma^2}\varepsilon'\varepsilon\right\},$$

where

$$\varepsilon = (I_n - \rho W)y - X\beta - \theta WX.$$

There is a number of different approach for assigning prior distributions for the parameters in η. Following LeSage (1997), LeSage and Parent (2007), use Metropolis within Gibbs sampling procedure (Metropolis-Hastings sampling for the parameter ρ and Gibbs sampling from the normal distribution for the parameters β, θ and normal inverse gamma distributions for σ^2). The MCMC algorithm for the SDM is provided by LeSage and Pace (2009), see Chapter 5. In the case of homoscedastic disturbances in SDM marginal likelihoods depend only on the parameter ρ.

The log-marginal posterior likelihoods $p(M|y)$ are the key magnitudes for model comparison in the Bayesian approach. Log-marginal density vectors for each unique model found during the MCMC are stored to calculate posterior model probabilities over the set of all unique models covered by the sampler. The posterior model probabilities sum to unity, and they can be used as weights to form a linear combination of estimates based on different explanatory variables.

For each model in Tables 3, 4, 5 1,000,000 draws of the Metropolis within Gibbs sampling procedure were carried out, excluding the first 200,000 draws to produce posterior estimates.[7] I run Raftery-Lewis diagnostic tests for each parameter of the chain, in order to detect convergence to the stationary distribution and to provide a way of bounding the variance of estimates of quantiles of functions of parameters. In the Raftery-Lewis test, q is a quantile of the quantity of interest (e.g. 0.05), r is a level of precision desired (e.g. ± 0.025), s is a probability associated with r (e.g. 0.90), and δ is a convergence tolerance (e.g. 0.01). The above-mentioned number of draws of the Metropolis within Gibbs sampling procedure is sufficient for q=0.001, r=0.0005, s=0.999, and δ =0.001.

Standard errors and significance levels provided in Tables 3, 4, 5, are derived from simple descriptive statistics of the Markov chain iterations for each sample.

References

[1] Abreu, M., De Groot, H. L., Florax, R. J. (2004). Space and growth: a survey of empirical evidence and methods. Tinbergen Institute Discussion Papers, No 04-129/3.

[2] Ahrend, R. (2005). Speed of Reform, Initial Conditions or Political Orientation? Explaining Russian Regions' Economic Performance. *Post-Communist Economies, 17* (3), 289–317.

[3] Arbia, G. (2006). Spatial econometrics: statistical foundations and applications to regional convergence. Springer Science & Business Media.

[4] Armstrong, H. W. (1995). Convergence among regions of the European Union, 1950–1990. *Papers in Regional Science*, 74(2), 143–152.

[5] Berkowitz, D., DeJong, D. N. (2003). Policy reform and growth in post-Soviet Russia. *European economic review*, 47(2), 337–352.

[6] Barro, R. J. (1991). Economic Growth in a Cross Section of Countries. *The Quarterly Journal of Economics*, 106(2), 407–443.

[7] Barro R. J. and X. Sala-i-Martin (1992). Regional growth and migration: a Japan – United States Comparison. *Journal of Japanese and International Economies,* 6, 312–346.

[8] Baumol, W. J. (1986). Productivity Growth, Convergence, and Welfare: What the Long-run Data Show. *American Economic Review,* 76 (5), 107–285.

[9] Bernat, G. A. (1996). Does manufacturing matter? A spatial econometric view of Kaldor's laws. *Journal of Regional Science*, 36(3), 463–477.

[7]Estimates are mainly based on a spatial toolbox developed by LeSage for Matlab, see http://www.spatial-econometrics.com

[10] Buccellato, T. (2007). Convergence across Russian regions: a spatial econometrics approach. *Economics Working Papers 72. Centre for the Study of Economic and Social Change in Europe, SSEES, UCL: London, UK.*

[11] Carluer, F. (2005). Dynamics of Russian regional clubs: The time of divergence. *Regional Studies, 39 (6)*, 713–726.

[12] Demidova, O. (2015). Spatial effects for the eastern and western regions of Russia: a comparative analysis. *International Journal of Economic Policy in Emerging Economies.* 8(2). 153–168.

[13] Ertur, C., Le Gallo, J., Baumont, C. (2006). The European regional convergence process, 1980-1995: Do spatial regimes and spatial dependence matter? *International Regional Science Review*, 29(1), 3–34.

[14] Fingleton, B., McCombie, J.S.L. (1998). Increasing returns and economic growth: Some evidence for manufacturing from the European Union regions. Oxford Economic Papers 50, 89–105.

[15] Fingleton, B., López-Bazo, E. (2006). Empirical growth models with spatial effects. *Papers in Regional Science*, 85(2), 177–198.

[16] Glushchenko, K. (2010). Studies on income inequality between Russian regions. *Region: Economics and Sociology*, 4, 88–119. In Russian: Глущенко К.П. (2010). Исследования неравенства по доходам между российскими регионами. Регион: экономика и социология, 4, 88–119.

[17] Gluschenko, K. (2012). Myths about Beta-Convergence. *Journal of the New Economic Association*, 16(4), 26–44. In Russian: Глущенко К.П. (2012). Мифы о бета-конвергенции. Журнал Новой экономической ассоциации, 4(16), 26–44.

[18] Guriev, S., Vakulenko, E. (2012). Convergence between Russian regions. *Working Papers w0180, Center for Economic and Financial Research (CEFIR).*

[19] Harris, R.D.F., Tzavalis E. (1999). Inference for unit roots in dynamic panels where the time dimension is fixed. *Journal of Econometrics* 91: 201–226.

[20] Harris, R., Moffat, J., Kravtsova, V. (2011). In search of 'W'. *Spatial Economic Analysis*, 6(3), 249–270.

[21] Ivanova, V. I. (2014). Regional convergence of income: spatial analysis. *Spatial Economics*, 4, 100–119. *In Russian:* Иванова В. И. (2014). Региональная конвергенция доходов населения: пространственный анализ. Пространственная экономика. No 4. С. 100–119.

[22] Kholodilin, K. A., Oshchepkov, A., Siliverstovs, B. (2012). The Russian regional convergence process: Where is it leading? *Eastern European Economics*, 50(3), 5–26.

[23] LeSage, J. P. (1997). Bayesian estimation of spatial autoregressive models. *International Regional Science Review*, 20(1-2), 113–129.

[24] LeSage, J. P. (2014). What Regional Scientists Need to Know about Spatial Econometrics. *The Review of Regional Studies*, 44(1), 13-32.

[25] LeSage, J. P., Fischer, M. M. (2008). Spatial growth regressions: model specification, estimation and interpretation. *Spatial Economic Analysis*, 3(3), 275–304.

[26] LeSage, J. P., Pace, R. K. (2009). Introduction to spatial econometrics. Boca Raton, FL: Chapman &Hall/CRC.

[27] LeSage, J. P., Parent, O. (2007). Bayesian model averaging for spatial econometric models. *Geographical Analysis*, 39(3), 241–267.

[28] Lugovoy, O., Dashkeev, V., Mazaev, I. (2007). Analysis of economic growth in regions: geographical and institutional Aspect. Consortium for Economic Policy Research and Advice. Moscow. 132 p.

[29] Madigan, D., York, J. (1995). Bayesian graphical models for discrete data. *International Statistical Review/Revue Internationale de Statistique*, 63(2), 215–232.

[30] Rey, S., Montouri, B. (1999). U.S. regional income convergence: a spatial econometric perspective. *Regional Studies,* 33(2), 143–156.

[31] Sala-i-Martin, X. X. (1996). Regional cohesion: evidence and theories of regional growth and convergence. *European Economic Review*, 40(6), 1325–1352.

[32] Solanko, L. (2003). An empirical note on growth and convergence accross Russian regions. *BOFIT Discussion Papers* 9/2003.

[33] Taylor, A. M. (1999). Sources of convergence in the late nineteenth century. *European Economic Review*, 43(9), 1621–1645.

[34] Williamson, J. G. (1996). Globalization, convergence, and history. *The Journal of Economic History*, 56(02), 277–306.

[35] Zverev, D. V., Kolomak, E. A. Subnational Fiscal Policy in Russia: Regional Differences and Relations. Series "Scientific Reports: Independent Economic Analysis", 214. Moscow Public Science Foundation; Siberian Center for Applied Economic Research, 2010, 160 p. *In Russian*: Зверев Д.В., Коломак Е.А. (2010). Субфедеральная фискальная политика в России: межрегиональные различия и связи. Серия "Научные доклады: независимый экономический анализ", No 214. – М., Московский общественный научный фонд; Сибирский центр прикладных экономических исследований, 2010. – 160 с.

Contact details:

Vera Ivanova
National Research University Higher School of Economics (Moscow, Russia).
Center for Market Studies and Spatial Economics, research fellow.
E-mail: viivanova@hse.ru, verunina@gmail.com