

Spatial convergence of real wages in Russian cities

Vera Ivanova¹ 

Received: 27 August 2016 / Accepted: 16 October 2017 / Published online: 30 October 2017
© Springer-Verlag GmbH Germany 2017

Abstract This paper provides an empirical test of spatial wage convergence in Russian cities. Using geo-coded data covering 997 Russian cities and towns from 1996 to 2013, I show that real city wages (i) converge over time and (ii) are significantly affected by the initial levels of real wages in neighboring cities. I also find that cities of the Far North, where a special wage policy is implemented, were converging more slowly than the rest of the country. I find a significant negative impact of regional subsidies on real wages in cities outside the Far North and that the effect of extractive industries on real wage has become weaker. These results are robust to the radius of spatial interaction, and my conclusions hold if remote settlements are not taken into account.

JEL Classification R11 · O18

1 Introduction

There is strong empirical evidence showing that regional economies grow at different rates which (i) tend to converge over time (see [Baumol 1986](#), and the subsequent literature on “convergence”) and (ii) are spatially correlated ([Abreu et al. 2005](#)). Convergence in per capita incomes finds its theoretical basis in models of economic growth ([Barro and Sala-i-Martin 1992](#)), while spatial convergence has been theoretically justified by [Ertur and Koch \(2007\)](#), who derive a spatial autoregressive equation from a growth model which takes into account technological interactions between regions based on proximity and neighborhood effects. Furthermore, new economic geog-

✉ Vera Ivanova
viivanova@hse.ru; verunina@gmail.com

¹ National Research University Higher School of Economics, 123 Griboedov Canal Embankment, Saint-Petersburg, Russian Federation 190068

raphy models predict spatial convergence in *real wages*, since migration between cities/regions leads to real wage equalization in the long run. This idea was proposed by Krugman (1991) in a setting with two regions. I derive a spatial beta-convergence equation for real wages using the simple multi-region model proposed by Tabuchi et al. (2005) in Sect. 3.

This paper tests these spatial convergence predictions for Russian cities. Since Russia is a large country with a very uneven distribution of economic activity, it is natural to conjecture that the “spatial” component of convergence in both real wages and per capita incomes is essential. Until now, the analysis of the spatial determinants of regional economic growth in Russia has been addressed in the literature by using the subjects-of-federation-level data.¹ This approach, however, is problematic, for it leads to a loss of information due to extremely high *intra-regional* heterogeneity (Zubarevich 2015). Moreover, because the regions are few, this limits the use of advanced econometric methods in order to estimate the relevant effects. In this paper, I suggest a way of overcoming these difficulties.

Using disaggregated geo-coded data, I study whether the properties of city location patterns foster convergence in real wages, and quantify the impact of the spatial structure of the economy on β -convergence in real wages in Russia. This has the advantage of exploiting a finer location pattern as a source of variation, and allows me to design a flexible empirical strategy that yields robust inferences. I apply Bayesian spatial econometric models that allow a comparison of the estimation results for different spatial weight matrices. The flexibility of this econometric procedure also comes from the possibility of fine-tuning the *sparsity* of the spatial weight matrix, a property that is of paramount importance in modern spatial econometrics (LeSage and Pace 2009). To the best of my knowledge, no similar setting has ever been used in previous empirical work on Russian regional development.

My main findings can be summarized as follows. First, real city wages (i) converge over time and (ii) are significantly affected by the initial levels of real wages in neighboring cities. Second, the radius of significant spatial interactions between Russian cities is around 300–600 km for cities west of the Urals, and around 1000 km for the Far North (cities in the Arctic Circle and some other isolated cities), where convergence is slower. Third, the effect of regional subsidies on real wage is negative, and the effect of natural resources on real wages has become weaker over time. Finally, although no official statistical data on per capita incomes are available at city level in Russia, I find a strong positive correlation between per capita income and wages at the subject-of-the-federation level, which persists over time. Based on that, I believe that my main results provide indirect evidence for spatial convergence of Russian cities in per capita incomes.

The rest of the paper is organized as follows. Section 2 reviews the literature, mostly based on Russian regional data. Section 3 provides a theoretical foundation for real

¹ The scale of a federal subject unit in Russia corresponds, more or less, to that of a state in the US.

wage convergence in cities. Section 4 describes the dataset. I discuss spatial weight matrices applied to Russian regional data and show a significant positive autocorrelation of real wages in Russia using a series of spatial matrices. Section 5 shows the findings: (i) conditional sigma-convergence of spatially weighted real wages and (ii) spatial beta-convergence of real city wages. I show that my conclusions are robust to the threshold distance, i.e., the estimation results are not sensitive to the exclusion of remote towns from the dataset. Section 6 concludes.

2 Literature review

There is an extensive empirical literature following Baumol's (1986) seminal paper on income convergence, using different datasets, explanatory variables and evaluation methods (including Barro 1991; Barro and Sala-i-Martin 1992; Sala-i-Martin 1996; Williamson 1996; Taylor 1999, and many more). Combining the baseline growth model with the fundamentals of the new economic geography (NEG), the regional science literature has introduced spatial dependence into the growth regression model. The earliest studies on spatial growth models include Armstrong (1995), Bernat (1996), and Fingleton and McCombie (1998). Thorough reviews on spatial growth studies are provided by Rey and Montouri (1999), Arbia (2006), Fingleton and López-Bazo (2006), Rey and Le Gallo (2009) and Le Gallo and Fingleton (2014). Most of them find a significant impact of spatial location patterns on regional income convergence.

Empirical work on Russian regional income growth covers different time periods. Extensive surveys are provided by Glushchenko (2010); Gluschenko (2012) and Guriev and Vakulenko (2012). The convergence hypothesis has been systematically rejected in papers focusing on the earliest post-Soviet period, which may be due to either short time horizons, or missing data, or both. On the contrary, more recent studies provide growing evidence of regional convergence in per capita income across subjects-of-the-federation in Russia. The results also tend to depend on the choice of a regional income measure: while per capita gross regional product (GRP) differentials persist, the cross-regional gaps in per capita incomes and regional wages have shrunk substantially.

Earlier papers on regional income inequalities fully ignored the spatial structure of Russian regions and therefore did not use the tools of spatial econometrics. There is, however, growing empirical support for there being a substantial impact of geography on the dynamics of regional incomes, see Table 5 in "Appendix 1" for a summary. In general, it is fair to say that the evidence provided by studies of regional income convergence in Russia is inconclusive. It is also worth noting that very few papers discuss possible dissimilarities in regional interaction for different parts of the country. Demidova (2015) revealed an asymmetric influence of eastern and western regions on each other using a partitioned spatial matrix. Ivanova (2014) tested whether estimates of a spatial growth model are sensitive to the spatial weight choice. To the best of my knowledge, neither wage convergence nor per capita income convergence in Russia has been studied at lower levels of spatial aggregation than at the subject-of-the-federation level.

3 A theoretical model of spatial convergence in real wages

As pointed out in Introduction, spatial convergence in *per capita incomes* has found its theoretical grounds in the spatial growth literature (Ertur and Koch 2007). What about theoretical foundations of spatial beta-convergence in *real wages*? The purpose of this section is to demonstrate that such convergence can be justified by NEG models. To show this, I use the multi-city setting proposed by Tabuchi et al. (2005) to derive a spatial beta-convergence equation for real wages across cities. For the sake of consistency, I start with a brief description of the model setup.²

There are n cities. Denote the share of total population residing in city i by λ_i , and the distance between cities i and j by $d_{ij} > 0$. Define the distance matrix by $\mathbf{D} \equiv (d_{ij})$ for $i, j = 1, \dots, n$; the vector of population distribution across cities by $\boldsymbol{\lambda} \equiv (\lambda_1, \dots, \lambda_n)^T$, where T denotes the transpose operator; and the vector of *real wages* in cities by $\mathbf{V}(\boldsymbol{\lambda}, \mathbf{D}) \equiv [V_1(\boldsymbol{\lambda}, \mathbf{D}), \dots, V_n(\boldsymbol{\lambda}, \mathbf{D})]^T$.³ Because λ_i is city i 's population share, we have

$$\sum_{j=1}^n \lambda_j = 1. \quad (1)$$

Equation (1) can be equivalently restated as $\boldsymbol{\lambda} \in \Delta_{n-1}$, where Δ_{n-1} is the standard $(n - 1)$ -dimensional simplex. This formulation, although less intuitive than (1), will prove convenient below.

The key assumption of the model is that migration decisions made by workers are *fully driven by pairwise comparisons of real wages between cities*.⁴ More specifically, the net migration flow from city i to city j is proportional to (i) the *real wage differential* $V_i - V_j$ between cities, and (ii) the *speed of adjustment* $f_{ij}(\boldsymbol{\lambda}, \mathbf{D})$ of migration decisions between cities i and j , which depends, in general, on the whole population distribution $\boldsymbol{\lambda}$ and the whole distance matrix \mathbf{D} . For simplicity, the pattern of adjustment speeds is assumed to be symmetric: $f_{ij}(\boldsymbol{\lambda}, \mathbf{D}) = f_{ji}(\boldsymbol{\lambda}, \mathbf{D})$, $i, j = 1, \dots, n$. Finally, it is required that $f_{ij}(\boldsymbol{\lambda}, \mathbf{D})$ are sufficiently differentiable in $\boldsymbol{\lambda}$.

As shown by Tabuchi et al. (2005, Eq. (6) on p. 431), the migration dynamics in the model can be described by the following ordinary differential equation (ODE) system:

$$\frac{d\boldsymbol{\lambda}}{dt} = \mathbf{F}(\boldsymbol{\lambda}, \mathbf{D}) \cdot \mathbf{V}(\boldsymbol{\lambda}, \mathbf{D}), \quad (2)$$

² See Tabuchi et al. (2005) for more details. My notation differs slightly from theirs, as my purpose is not to characterize completely the spatial equilibrium, but just to show convergence in real wages.

³ Formally speaking, $\mathbf{V}(\boldsymbol{\lambda}, \mathbf{D})$ is the vector of the *indirect utility levels*, which can be interpreted as real wages when preferences are homothetic (e.g., CES). In this case, an ideal price index is well defined, and the indirect utility can be represented as a ratio of the income (which coincides with the nominal wage in this type of models) to the ideal price index. This is exactly what economists typically refer to as a "real wage."

⁴ Thus, other factors of labor mobility, such as idiosyncratic tastes about place of residence (Tabuchi and Thisse 2002) and migration costs (Tabuchi et al. 2016) are ruled out here.

where t denotes time, while the matrix $\mathbf{F}(\boldsymbol{\lambda}, \mathbf{D})$ is constructed as follows: its off-diagonal entries are the speeds of adjustments taken with the opposite sign, while its ii th entry is given by $\sum_{j \neq i}^n f_{ij}(\boldsymbol{\lambda}, \mathbf{D})$.

Given the above assumptions, it is readily verified that $\mathbf{F}(\boldsymbol{\lambda}, \mathbf{D})$ is a symmetric positive semi-definite $n \times n$ matrix, which is continuously differentiable in $\boldsymbol{\lambda}$ and satisfies the following identity:

$$\mathbf{F}(\boldsymbol{\lambda}, \mathbf{D}) \cdot \mathbf{1} = \mathbf{0}, \tag{3}$$

where $\mathbf{1} \equiv (1, \dots, 1)^T$.

Let $\boldsymbol{\lambda}^*$ be a stable spatial equilibrium of the system (2), which requires that $d\boldsymbol{\lambda}/dt = 0$. Then, as implied by (2) and (3), it must be that $\mathbf{F}(\boldsymbol{\lambda}, \mathbf{D}) \cdot \mathbf{V}(\boldsymbol{\lambda}, \mathbf{D}) = V^* \cdot \mathbf{F}(\boldsymbol{\lambda}, \mathbf{D}) \cdot \mathbf{1}$, which yields $\mathbf{V}(\boldsymbol{\lambda}^*, \mathbf{D}) = V^* \mathbf{1}$, i.e., the equilibrium real wage must be the same in all cities. In other words, the long-run equilibrium $\boldsymbol{\lambda}^*$ displays *real wage equalization* across cities.

Assume that the initial state of the system is off-equilibrium: $\boldsymbol{\lambda}(0) \neq \boldsymbol{\lambda}^*$. This can be interpreted in two ways: (i) migration frictions, which were strong in the past, i.e., at times $t < 0$, but have been relaxed drastically at $t = 0$; (ii) the existence of compensating wage differentials (Roback 1982). Both these cases are highly relevant for Russia. First, until the end of the Soviet era, the interplay between agglomeration and dispersion forces, which constitute the whole essence of NEG, was not a key factor for city size distribution or city-level wages in Russia. Therefore, one may consider the process of convergence in real wages from their disequilibrium values (generated by the inertia typical for an ex-planned economy with low and impeded mobility) to their equilibrium values as being shaped by the interaction of centripetal and centrifugal forces. Second, there was a special wage policy in the USSR, to attract labor to settlements with bad amenities by setting higher real wages to compensate, see Sect. 4 for details.

It remains to obtain testable convergence equations. To do so, we need to derive a system of ODE describing real wage dynamics. Intuitively, this can be done by simply changing variables in (2). There is, however, a technical difficulty. Due to (1), the mapping $\mathbf{V}(\boldsymbol{\lambda}, \mathbf{D})$ is not invertible, since it maps Δ_{n-1} , which is an $(n - 1)$ -dimensional surface, onto \mathbb{R}_+^n , which is an n -dimensional space. Therefore, (i) not any vector $\mathbf{V}(0, \mathbf{D}) \in \mathbb{R}_+^n$ can serve as a vector of initial real wage levels and (ii) we cannot compute $d\mathbf{V}/dt$ as

$$\frac{d\mathbf{V}}{dt} = \left(\frac{\partial \mathbf{V}}{\partial \boldsymbol{\lambda}} \right)^{-1} \frac{d\boldsymbol{\lambda}}{dt},$$

because the inverse Jacobi matrix $(\partial \mathbf{V} / \partial \boldsymbol{\lambda})^{-1}$ is not well defined.

To tackle these difficulties, we introduce $\mathbf{y}(\boldsymbol{\lambda}, \mathbf{D}) \equiv (y_1(\boldsymbol{\lambda}, \mathbf{D}), \dots, y_n(\boldsymbol{\lambda}, \mathbf{D}))$, where $y_i(\boldsymbol{\lambda}, \mathbf{D})$ is the *normalized real wage* in city i defined by

$$y_i(\boldsymbol{\lambda}, \mathbf{D}) \equiv \frac{V_i(\boldsymbol{\lambda}, \mathbf{D})}{\sum_{j=1}^n V_j(\boldsymbol{\lambda}, \mathbf{D})}.$$

The mapping $\mathbf{y}(\boldsymbol{\lambda}, \mathbf{D})$ maps Δ_{n-1} onto itself, and it is invertible at any point of the interior of $\mathbf{y}(\Delta_{n-1}, \mathbf{D})$. We are now equipped to obtain the following result.

Proposition 1 *Assume the migration dynamics ODE system (2) has a unique steady state $\boldsymbol{\lambda}^*$, which is interior and stable. Then:*

- (i) *in a neighborhood of $\boldsymbol{\lambda}^*$, the system (2) can be equivalently recast into the normalized real wage dynamics ODE system:*

$$\frac{d \ln y_i}{dt} = g_i(y_i, \mathbf{y}_{-i}, \mathbf{D}); \quad (4)$$

- (ii) *the system (4) has a unique steady state given by $\mathbf{y}^* = \mathbf{1}/n$, and it is asymptotically stable;*
 (iii) *for any fixed moment of time $T > 0$, the following “convergence equations” hold (approximately) in the vicinity of the steady state \mathbf{y}^* :*

$$\ln y_i(T) - \ln y_i(0) \approx \alpha + \beta_i(\mathbf{D}) \ln y_i(0) + \sum_{j \neq i} w_{ij}(\mathbf{D}) \ln y_j(0), \quad (5)$$

where $y_i(t)$ is a solution to (4) as a function of t .

Proof See “Appendix 2.”

Equation (5) can be viewed as a spatial beta-convergence equation for the following reasons. First, it relates the growth rate of real wages to the initial levels of real wages in neighboring cities. Second, the coefficients $w_{ij}(\mathbf{D})$ depend on the distance matrix \mathbf{D} , whence they can be interpreted as spatial weights. \square

4 Real wages in Russian cities

4.1 Data description

City wage data I use average monthly wages in Russian cities provided in the Multistat database.⁵ I omit settlements of the Chechen republic and the Ingushetia republic because of missing data. Also, I exclude settlements that have lost the status of a city and have become a part of another city between 1996 and 2013,⁶ resulting in a dataset of 997 cities, and the time span is 1996–2013.

To compute real wages, I deflate nominal wages by a year- and region-specific measure of the cost of living. This measure reflects the value of a fixed set of goods and services, typically referred to as the “market basket of subsistence goods” (Remington 2015). The market basket values are not reported by Russian statistical agencies at the city level, but they are available at the subject-of-the-federation level starting from 2000.⁷ For the years before 2000, I use

⁵ Multistat. Economy of Russian cities, web page: <http://www.multistat.ru/> (accessed 1 February 2016).

⁶ Multistat does not provide data on settlements which do not have city status in the current year.

⁷ Russian Federal State Statistics Service, web page: <http://www.gks.ru/> (accessed 1 September 2016).

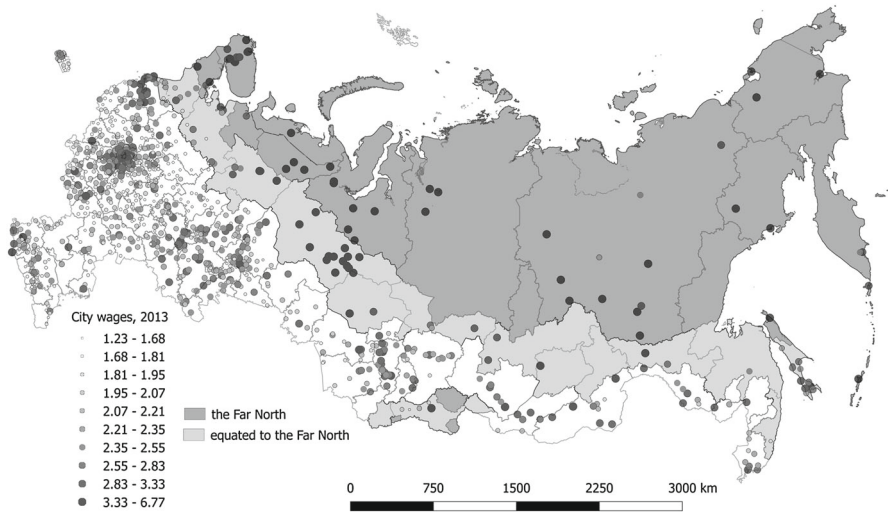


Fig. 1 City wages 2013 (in market baskets) and the Far North territories

the regional consumer price index (CPI) as a proxy for the basket value. The real wage data reveal very large differences between Russian cities, especially at the end of 1990s. In 1996, the richest city in the sample had a real wage 29.1 times larger than the poorest one. By 2013, this ratio had decreased to 5.5.

Next, I have geo-coded city-level real wages. The spatial distribution of real wages in 2013 is shown in Fig. 1, where classes are defined as deciles. Richer circles (i.e., cities with higher real wages) tend to correspond to cities with larger population: Moscow, Saint Petersburg, Yekaterinburg. However, cities located in oil and gas regions (e.g., Tyumenskaya or Sakha Yakutia) also display high wages although they have much smaller populations.

The Far North Recent studies on Russian income data show that results may depend substantially on the part of the country (Ahrend 2005; Demidova 2015; Sardadvar and Vakulenko 2016). The standard east-west split, which separates the European and Asian parts of Russia, runs along the Ural Mountains. Although real wages in eastern cities are higher than in western cities, there are no a priori reasons to believe that they converge faster or slower. Therefore, the east-west partition is hardly relevant in the present context. I choose instead to split the sample into Far North (FN) cities, where a special wage policy is implemented, and non-Far North cities.

The notion of FN territories was established during the Soviet period and comprised the regions of the USSR with severe climate conditions and/or poor transport connection to the rest of the country. Most of these territories (see Fig. 1) are in the Arctic Circle, whence the label “Far North.” In order to attract skilled workers, e.g., doctors and teachers, to Far North territories, the Soviet government offered higher wages and other benefits (longer vacation periods, lower retirement ages, etc.) to those who chose to work in the Northern regions. This practice is referred to as the “Northern” benefits.

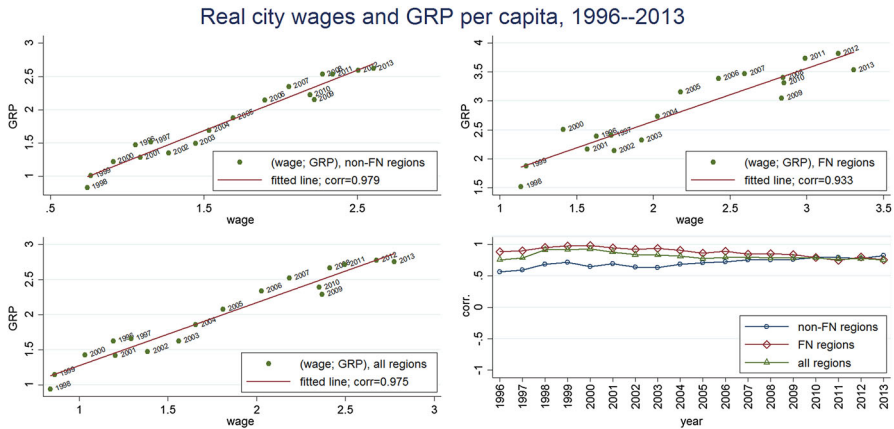


Fig. 2 Real city wages and GRP per capita in market baskets, 1996–2013

These benefits were continued in Russia after 1991 and are now regulated by the [Federal Law \(2004\)](#). In 2005, most of the benefits provided by the Law were converted into cash allowances, while the burden of payments was redistributed among the Federal government and local governments. The availability of “Northern” benefits now depends on the ability of enterprises to pay bonuses, which varies greatly both across sectors and cities ([Wengle and Rasell 2008](#)).

The whole FN area comprises about 70% of Russia’s total area, the so-called territories equated to the Far North in Russia, which have either zero road connections with the rest of the country, or connections which suffer from seasonal isolation. Loosely speaking, these regions are “island regions.” Settlements located there may hence be considered as remote areas with peculiar conditions of spatial interactions.

I split the dataset into two subsets: the FN cities located in the Far North regions or regions equated to the Far North (119 cities in 19 regions), and the non-FN cities located mainly in the western part of the country (878 cities in 59 regions). The real wages are higher on average in the FN cities than in the rest of the country (see [Figs. 1 and 6](#) in “Appendix 3”).

Real wages as a proxy for per capita GRP Using labor force as weights, I compute the monthly weighted average real wages and the weighted average GRP per capita measured in market baskets. I average real wages across cities and GRP across subjects-of-the-federation, since there are no data available on per capita GRP at the city level. [Figure 2](#) shows that these two indicators are highly correlated. The bottom-right panel of [Fig. 2](#) shows the dynamics of the cross-sectional correlation coefficients between real city wages and GRP per capita at the subject-of-the-federation level (see also [Table 6](#) in “Appendix 3”). The correlation coefficient is always above 0.5. The other three panels of [Fig. 2](#) show the scatterplots of real wages versus GRP per capita in market baskets computed, respectively, for the whole country, for the FN part and for the non-FN part. In all three cases, the fitted lines are very close to 45° line, and the corresponding correlation coefficients are very close to 1. This strong positive correlation makes me conclude that real city wages in Russia may serve as a reasonable

proxy for gross regional income at a more disaggregate administrative level, at least within the time period of my study.

4.2 Spatial autocorrelation of real wages

Figure 1 suggests that there are spatial patterns in the distribution of real wages across Russian cities and towns. The choice of spatial weights describing the intensity of the spatial effect on a given location is one of the key issues in spatial analysis. [Harris et al. \(2011\)](#) provided an extensive review of the standard approaches to constructing a spatial weights matrix. It is common practice to use either contiguity or distance-based spatial weight matrices. Distances between regions are measured as distances between the regional central (or largest) cities, and this type of spatial weights leads to rather crude measures of the spatial distribution at the regional level. However, it suits better for cities. Furthermore, there are papers constructing weights based on geodesic distances along highways or railways. In Russia, this way of measuring distances may lead to mixing up the effects of overall spatial interactions between regions with those of particular infrastructural improvements. While for most of the western part, land transportation networks are well developed, there are regions that are poorly connected with other territories via railroads (e.g., Tomskaya, Tyva, Altay), but have much better highway connections. There are also remote territories (e.g., Magadanskaya, Chukotka) for which the common way to access other regions is air transport combined with land transport. These considerations reveal the difficulties of constructing sophisticated measures of distance based on detailed information on the Russian transportation system.

An alternative strategy for constructing spatial weights relies on a fixed number of neighbors for each spatial unit. However, this method is also problematic, as the spatial distribution of population in Russia is highly asymmetric, as is the size distribution of the regional units. Therefore, assuming that each region has the same number of neighbors to interact with would mean that the radius of spatial interaction for the FN regions would be several times longer than that for the rest of the country. For example, 10 neighbors, as used by [Sardadvar and Vakulenko \(2016\)](#) in the case of Sakha Yakutia (a region in the Far North), comprise about 37% of the area of Russia, while the same number of nearest regional neighbors for the Moscow city, which is treated as a separate subject-of-federation in the Russian official statistics, covers less than 3% of the country's area.

I construct the spatial weights as follows:

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

where d_{ij} is the great circle distance between cities i and j , while $C > 0$ is a cutoff distance. When C is greater than the maximum distance between spatial units in my data, I get standard inverse distances as a limiting case.

Given the cutoff C , the spatial weights $w_{ij}^{(C)}$ as defined above are functions of the great circle distances only. The great circle distance may serve as a proxy for

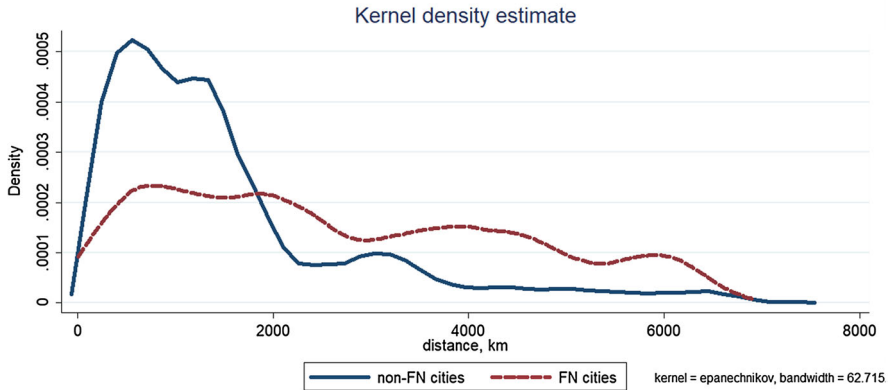


Fig. 3 Kernel density estimation of pairwise great circle distances between Russian cities

Table 1 Number of cities in connected inverse great circle matrices, 100–2000 km

C (km)	100	200	300	400	500–600	700–800	900–1000	2000
Total number of cities	635	855	892	960	987	988	997	997
Number of non-FN cities	623	816	816	857	878	878	878	878
Number of FN cities	12	39	76	103	109	110	119	119
Neighbors	15	39	73	109	147–192	214–280	320–361	693

distance covered by land transport (for the densely populated part of the country) and by air transport (for remote regions). Meanwhile, the cutoff distance C is not *a priori* specified, which provides an additional dimension of flexibility.

The distribution of pairwise distances between FN cities and between non-FN cities is shown in Fig. 3. The median distance for FN cities is 2305 km, and the mean is 2698 km; for non-FN cities the median distance is 1159 km and the mean is 1545 km. There are no grounds to take median or mean values of pairwise distances as a reasonable cutoff distance for the spatial interactions. As a robustness check, I construct weights with different cutoff distances starting from 100 km with 50-km steps. For small values of C , I get matrices with “islands,” i.e., at least two blocks not connected with each other. In such cases, I cut my sample by focusing on observations which generate a connected weight matrix of the highest dimensionality. The number of cities in the largest connected matrix with corresponding cutoff distances C are reported in Table 1. The smallest cutoff distance for which all the cities form a connected matrix is 900 km. When $C < 250$ km, no more than 40% of FN cities are involved in the spatial interaction matrix, while the percentage of non-FN cities is greater than 90 within the same cutoff distance.

As shown in Fig. 3 and Table 1, the radius of interaction up to 300–400 km corresponds to the biggest changes in number of city pairs. The average number of neighbors of each city within the cutoff distance is also shown in Table 1. Within a radius of 1000 km, each city interacts with one-third of all other cities, on average. When the cutoff

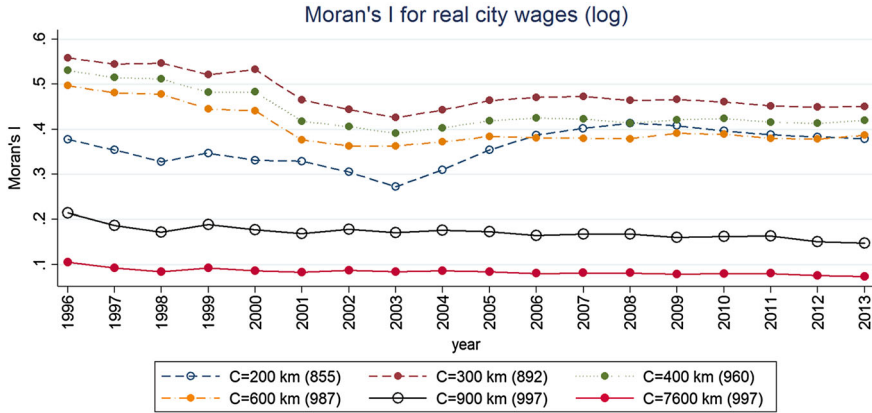


Fig. 4 Moran's *I* for log of city wages (in market baskets), 1996–2013. Spatial weights—row-standardized inverse great circle distances with cutoffs. Number of cities in parentheses

distance is 2000 km, this share is about 70%. I obtain a weight matrix corresponding to the complete graph when $C \geq 7600$ km.

In order to measure the overall tightness of wage co-movements in spatially close cities, I compute Moran's global spatial autocorrelation index *I* using row-standardized spatial weights w_{ij} . Figure 4 shows annual changes in the global Moran's *I* for log real wages calculated using inverse great circle distances with different cutoffs. All indices are positive and significant at the 1% level, whence city-level wages are spatially autocorrelated. Starting from a cutoff distance of 300 km, the pattern Moran's *I* follows barely changes as *C* increases further. Hence, a radius of 300 km captures almost all spatial autocorrelation in real city wages.

To sum up, there is significant spatial autocorrelation in real wages in Russian cities. Ignoring possible inter-city spatial links may therefore lead to biased estimates of wage convergence.

5 Spatial convergence

5.1 Spatial sigma-convergence

Figure 5 shows that the variation of real city wages (measured by the standard deviation of log real city wages) decreases over time. Changes in the standard deviation of the spatial lags of log real wages are presented in Fig. 5, where the lags are constructed as $v_{i,t} = \sum_{j=1}^n w_{ij} y_{i,t}$, with $y_{i,t}$ the log real wage in city *i* and year *t*. These two figures provide evidence for the decreasing volatility of wages in city neighborhoods as defined by the cutoff distance.

To test the hypothesis of spatial sigma-convergence, I use a panel unit root test for spatially lagged real wages and consider a first-order autoregressive component

$$v_{i,t} = bv_{i,t-1} + \gamma_i + \delta_i t + \varepsilon_{i,t}, \tag{6}$$

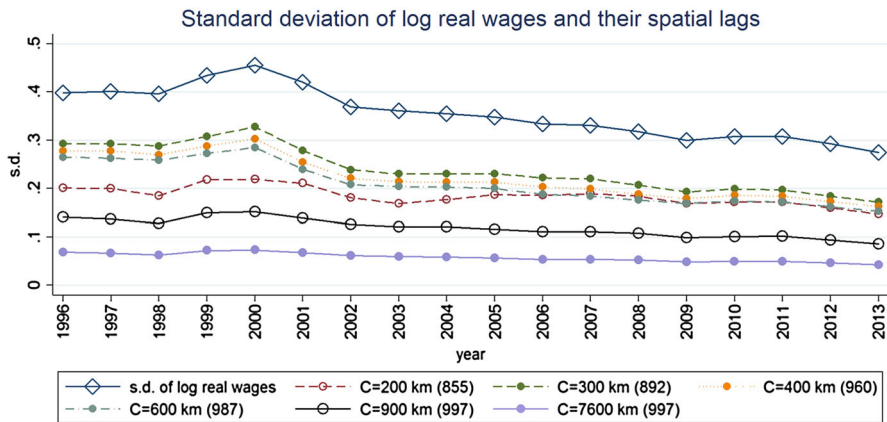


Fig. 5 Standard deviation of real wages (log) and their spatial lags, 1996–2013. Spatial weights—row-standardized inverse great circle distances with cutoffs. Number of cities in parentheses

where γ_i is an individual fixed effect; where δ_{it} is a time trend that captures the fact that average city-level wages increase over time (see Fig. 6); and where $\varepsilon_{i,t} \sim iid(0, \sigma^2)$ is the error term.

I split the time period into two sub-periods: 1996–2005 and 2006–2013. This is done to capture the possible impacts of: (i) changes in the fixed set of goods and services used to define the market basket, which was revised in late 2005 and (ii) changes in the Federal Law of 2004, when the practice of converting “Northern” benefits into cash allowances started. Note that both time periods are relatively short and the number of observations in each panel is large. This has been taken into account when running panel unit root tests. I test the null hypothesis $H_0: b_i = 1$ for $i = 1, \dots, n$ (the panels contain unit roots) using two panel unit root tests.

The first test by Harris and Tzavalis (1999; henceforth HT) tests the above-mentioned null hypothesis against the “homogeneous” alternative $H_a^{HT}: -1 < b_i = b < 1$ for $i = 1, \dots, n$ (the panels are stationary). Estimates for the HT tests using spatial lags of real wages with different cutoff distances are provided in Table 7 in “Appendix 4.” Spatially determined panels of real city wages are stationary during 1996–2005 when the time trend is not included, and this does not depend on C , except for the FN dataset with a large radius ($C \geq 4000$ km). When panel-specific time trends are included, only large distances ($C \geq 900$ km) yield spatially weighted real wages that converge during the first time period. For the second period (2006–2013), neighboring real city wages diverge for distances of less than 900 km and converge for larger distances, when there are no time trends. If I include panel-specific time trends, spatially lagged real city wages diverge between 2006 and 2013.

The alternative H_a^{HT} imposes the restriction that $b_i = b$ for all i , which implies that the rate of convergence would be the same for all cities (regions). Maddala and Wu (1999) showed that this implication is too restrictive in practice.

I hence run a second unit root test that relaxes this restrictive assumption. This test by Im, Pesaran, and Shin (2003; henceforth, IPS) tests H_0 against the “heterogeneous”

alternative H_a^{IPS} : $-1 < b_i < 1$ for $i = 1, \dots, n_1$ and $b_i = 1$ for $i = n_1 + 1, n_1 + 2, \dots, n$ (the fraction of panels that follow stationary processes is nonzero).

The estimation results for the IPS test for spatially determined real wages are given in Table 8 in ‘‘Appendix 4’’. Relaxing the strong alternative H_a^{HT} and taking H_a^{IPS} instead suggest that there is a nonzero share of stationary panels of weighted city real wages, i.e., there is spatial sigma-convergence of real wages during both sub-periods, when city-specific time trends are included.

I thus reject the null hypothesis of a unit root and find support for conditional (in terms of the IPS alternative) sigma-convergence in wages for neighboring cities.

5.2 Spatial beta-convergence

As shown by Quah (1993), beta-convergence is a weaker property than sigma-convergence. I have shown in the previous subsection that the strong alternative of spatial sigma-convergence does not hold for real wages in Russian cities. I thus now relax the assumption and use a test of beta-convergence which is based on the cross-sectional Barro regression:

$$\ln y_{i,T} = \alpha + (1 + T\beta) \ln y_{i,0} + \gamma^T \mathbf{X}_i + \varepsilon_i, \quad (7)$$

where $y_{i,0}$ and $y_{i,T}$ are the real wages in the initial period and the final period; where \mathbf{X}_i is a vector of controls; and where $\varepsilon_i \sim iid(0, \sigma^2)$ is the error term.

I consider two groups of factors as controls: (i) public sector-related variables and (ii) natural resource-related variables. I use these factors to construct proxies that capture their possible effects on real city wages.

Public sector A substantial part of the labor force in Russia is employed in the public sector. According to the Federal State Statistic Service, 42% of the total labor force in Russia were employed in state- and municipality-owned companies in 1995. That share had decreased to 28% in 2013. The consolidated budgets of Russian regions get revenues from the following sources: corporate income taxes, taxes on personal income, property taxes, and gratuitous receipts. The later source is, mainly, a federal non-refundable transfer, i.e., subsidies. The Russian fiscal system redistributes taxes from the federal budget to regional budgets; regions with poor economic performance receive more subsidies than prosperous ones. Ponomareva and Zhuravskaya (2004) considered the reasons for subsidizing firms and noted that all levels of government may have incentives to subsidize firms to secure inefficiently high employment or bribes. The incentives of regional governments to redistribute subsidies efficiently may be weak, because when a region increases in its own budget revenues, federal subsidies to the region are reduced.

There are no available city-level data on the labor force and wages in the public/private sectors. In order to control for the public sector in the Barro regression, I use the percentage of non-refundable subsidies from the federal budget in the local regional budget. Annual subject-of-the-federation-level data on subsidies has been available since 2000. I take the average for 2000–2005 and the average for 2006–2013 for the

two time periods, and I include a variable *Subs* defined as $\ln(1 + \text{average subsidies})$ into the Barro regression.

Natural resources The share of natural resource rents in Russian gross domestic product is large. According to the World Bank Open Data, it was 9.83% in 1996 and 19.35% in 2005.⁸ This increased during 1996–2005, while the 2006–2013 period was characterized by a decline in the percentage of total natural resource rents in GDP: it has decreased from 19.42% in 2006 to 13.73% in 2013. In order to capture the possible impact of natural resource rents on real wages, I include a variable *Emp* computed as $\ln(1 + \text{the share of labor employed in the mining sector in total city labor})$ into the Barro regression. City-level data on labor in the mining sector have been available since 2005. I take the average over 2006–2013. In order to alleviate endogeneity concerns, I instrument the share of city labor in mining in 2005 using the share of mining in GRP for the previous year, 2004.

Spatial model specification The standard approach to studying regional growth involves constructing control variables describing, among others, economic diversity and specialization, measures of physical and human capital, market potential and the geographical characteristics of a spatial unit. Most of such data are not available for Russian cities. Hence, I face a problem of omitted variables in the regression. This problem is intrinsic to models run on Russian city-level data. **LeSage and Fischer (2008)** argued that **spatial Durbin models** are 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. I also include a spatial lag of the initial wage values. My baseline spatial regression of beta-convergence of real city wages in Russia is a spatial Durbin model (SDM) with a “Far North–non-Far North” interaction term:

$$\begin{aligned} \mathbf{y}_T = & \alpha \mathbf{1} + \left(1 + T\beta^{nonFN}\right) \mathbf{y}_0^{nonFN} + \left(1 + T\beta^{FN}\right) \mathbf{y}_0^{FN} \\ & + \gamma^{nonFN} \mathbf{X}^{nonFN} + \gamma^{FN} \mathbf{X}^{FN} + c \cdot \mathbf{dnorth} \\ & + \rho \mathbf{W} \mathbf{y}_T + \delta^{nonFN} \mathbf{W} \mathbf{y}_0^{nonFN} + \delta^{FN} \mathbf{W} \mathbf{y}_0^{FN} \\ & + \mathbf{W} \mathbf{X}^{nonFN} \boldsymbol{\theta}^{nonFN} + \mathbf{W} \mathbf{X}^{FN} \boldsymbol{\theta}^{FN} + \boldsymbol{\varepsilon}_T, \end{aligned} \quad (8)$$

where \mathbf{y} is the $n \times 1$ vector of log real wages; T and 0 denote the final year and the initial year, respectively; $\mathbf{1}$ is the $n \times 1$ vector of ones; β^{nonFN} and β^{FN} are the convergence parameters; \mathbf{dnorth} is a $n \times 1$ vector of dummy variables indicating whether city i is located in the Far North or not; \mathbf{X} is a matrix of control variables; \mathbf{W} is a $n \times n$ weight matrix built from inverse great circle distances with cutoffs C (the number of cities n depends on C ; see Table 1 and Fig. 3); ρ is the spatial autoregressive parameter; $\boldsymbol{\theta}$ is the vector of the parameters of the spatially lagged control variables; and $\boldsymbol{\varepsilon}_T \sim N(0, \sigma^2 \mathbf{I}_n)$ is the error term. In what follows, superscripts FN and $nonFN$ refer to the FN and non-FN datasets, respectively.

⁸ World Bank Open Data, web page: <http://data.worldbank.org/> (accessed 14 July 2017).

Equation (8) captures spatial general-equilibrium effects in a simple way: changes in one observation drive a series of changes in all regions in the sample until a new long-run steady-state equilibrium arises. For example, high wages in city i may attract migrants from neighboring cities j and k , so the employment rate in i will change. Employers may offer lower wages in i , while the employment rate will decrease in cities j and k , which may lead to higher salaries there. Neighbors of j and k may react to new wages in a similar way, which eventually generates a global spatial feedback effect.

5.3 Estimates and interpretations

Because of the presence of the spatial lag of the dependent variable, the coefficients in the SDM (8) cannot be directly interpreted as marginal effects which describe the magnitude of changes in the dependent variable that arise from changes in the explanatory variables. A change in an explanatory variable in city i will result in a direct impact on the wages of city i in the final year, and an indirect impact arising from spatial relationships with cities $j \neq i$. The averaged sum of such impacts for each explanatory variable is called **the average direct effect** of the explanatory variable (LeSage and Pace 2009).

When the wages in city i increase, i.e., when the population of city i becomes richer, then, according to the beta-convergence hypothesis, the wages in city i will grow more slowly. The direct impact of spatially close cities is positive and less than 1.

Estimates of the SDM with different spatial matrices are summarized in Tables 2 and 3. I use Bayesian estimation methods (see “Appendix 5” for a detailed description of my estimation strategy), and the spatial weights are row-standardized inverse great circle distances with cutoffs. I estimate Eq. (8) with different distance cutoffs: 200–1000 km with 100-km steps, and 1000–8000 km with 1000-km steps. In order to interpret my estimation results, I do not report coefficients for explanatory variables but provide their average direct effects.

As shown in Table 2, the spatial parameter ρ for the SDM with different spatial matrices introduced above is positive and highly significant. Therefore, the real city wage is related to the real wages of neighboring cities after conditioning for the effect of initial real wages y_0 and various controls. The values of ρ increase with the cutoff distance, C .

The direct effects of the initial wages represent the impact of this variable on wages of the final year, both immediate (own effect) and mediated (feedback effects via neighbors). For the first time period, the direct effects of initial wages vary from 0.577 to 0.609 for non-FN cities and from 0.596 to 0.838 for cities of the Far North (see Table 2). For the second time period, they vary from 0.602 to 0.661 for non-FN cities, and from 0.691 to 0.838 for the FN cities (see Table 3).

One of the main results of the SDM is that the direct effect of initial wages on final wages is stronger for FN cities than for non-FN cities. This holds for both time periods, as shown in Tables 2 and 3. Furthermore, the values of the direct effects of initial wages on real wages in the final year in non-FN cities are fairly robust to the

Table 2 Estimates of SDM for the period 1996–2005

C (km)	Dependent: wage (log) in market baskets, 2005				
	(1) 200	(2) 300	(3) 400	(4) 500	(5) 900
ρ	0.617***	0.692***	0.642***	0.703***	0.788***
Direct effects					
y_0^{nonFN}	0.609***	0.603***	0.589***	0.585***	0.588***
y_0^{FN}	0.838***	0.596***	0.716***	0.763***	0.739***
Emp^{nonFN}	0.078***	0.080***	0.091***	0.087***	0.080***
Emp^{FN}	0.010	−0.003	−0.012	0.003	0.037
$Subs^{nonFN}$	−0.041***	−0.045***	−0.054***	−0.059***	−0.057***
$Subs^{FN}$	0.007	0.103	0.101	0.093	0.042
τ^{nonFN}	15.6	15.4	14.8	14.7	14.8
τ^{FN}	38.2	15.1	21.6	26.0	23.6
n^{nonFN}	816	816	857	878	878
n^{FN}	39	76	103	109	119
n	855	892	960	987	997
C (km)	(6) 1000	(7) 2000	(8) 4000	(9) 6000	(10) 8000
ρ	0.781***	0.782***	0.934***	0.948***	0.941***
Direct effects					
y_0^{nonFN}	0.588***	0.588***	0.583***	0.577***	0.578***
y_0^{FN}	0.739***	0.727***	0.709***	0.685***	0.692***
Emp^{nonFN}	0.080***	0.077***	0.077***	0.081***	0.087***
Emp^{FN}	0.037	0.042	0.065*	0.067	0.063
$Subs^{nonFN}$	−0.057***	−0.059***	−0.037***	−0.039***	−0.035**
$Subs^{FN}$	0.042	0.044	0.034	0.034	0.033
τ^{nonFN}	14.8	14.8	14.6	14.4	14.4
τ^{FN}	23.6	22.5	21.1	19.5	19.9

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. In (6)–(10): $n = 997$, $n^{nonFN} = 878$, $n^{FN} = 119$

choice of C , which is consistent with [LeSage and Pace \(2009\)](#). A tendency for the stabilization of direct effect estimates starts from about $C = 400$ km, when $n < 997$, i.e., when some remote settlements are excluded. The differences between the direct effects of the initial real wages in the Far North decrease in the case of long distances.

The direct effects of regional subsidies on real wages in non-FN cities are significantly negative during both time periods. This captures the effect of federal non-refundable transfers to regional budgets: regions that are able to grow faster do not need federal support. My estimates of the SDM do not provide evidence of a significant impact of federal subsidies on wages in FN cities. A possible reason is that

Table 3 Estimates of SDM for the period 2006–2013

C (km)	Dependent: wage (log) in market baskets, 2013				
	(1) 200	(2) 300	(3) 400	(4) 500	(5) 900
ρ	0.589***	0.693***	0.636***	0.682***	0.735***
Direct effects					
y_0^{nonFN}	0.658***	0.661***	0.61***	0.609***	0.605***
y_0^{FN}	0.838***	0.813***	0.725***	0.706***	0.691***
Emp^{nonFN}	-0.005	-0.006	-0.009	-0.009	-0.008
Emp^{FN}	-0.003	-0.001	-0.008	-0.005	-0.001
$Subs^{nonFN}$	-0.053***	-0.054***	-0.06***	-0.057***	-0.049***
$Subs^{FN}$	0.245	0.021	-0.083	-0.071	-0.057
τ^{nonFN}	13.8	14	12.1	12.1	11.9
τ^{FN}	29.6	25.6	17.3	16.2	15.4
n^{nonFN}	816	816	857	878	878
n^{FN}	39	76	103	109	119
n	855	892	960	987	997
C (km)	(6) 1000	(7) 2000	(8) 4000	(9) 6000	(10) 8000
ρ	0.74***	0.776***	0.927***	0.931***	0.944***
Direct effects					
y_0^{nonFN}	0.606***	0.605***	0.602***	0.604***	0.608***
y_0^{FN}	0.692***	0.702***	0.756***	0.776***	0.774***
Emp^{nonFN}	-0.008	-0.007	-0.005	-0.005	-0.006
Emp^{FN}	-0.001	-0.002	-0.001	0.001	0.005
$Subs^{nonFN}$	-0.048***	-0.049***	-0.027**	-0.024*	-0.029
$Subs^{FN}$	-0.057	-0.055	0.011	0.05	0.041
τ^{nonFN}	12	11.9	11.8	11.9	12
τ^{FN}	15.4	15.9	19.5	21.3	21.1

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. In (6)–(10): $n = 997$, $n^{nonFN} = 878$, $n^{FN} = 119$

the policy of the “Northern” benefits creates inertia and slows down convergence in wages. This result echoes [Sardadvar and Vakulenko \(2017\)](#) who find that “Eastern regions suffer from circumstances which are difficult or even impossible to change by regional policies”.

Finally, the direct effects of employment in mining are significantly positive only for the non-FN cities during 1996–2005, when the share of natural resource rents in GDP was growing. This finding is in line with [Solanko \(2008\)](#), who studied regional income growth and used the share of extractive industries in GRP as a proxy for natural resources. The impact of employment in mining on real wages is no longer significant for 2006–2013, when the share of the mining sector in GDP was decreasing.

Table 4 Posterior model probabilities for SDM

C (km)	900	1000	2000	3000	4000	5000	6000	7000	8000
1996–2005	0.000	0.00	0.00	0.00	0.008	0.010	0.095	0.349	0.539
2006–2013	0.420	0.577	0.001	0.001	0.000	0.000	0.000	0.000	0.000

To sum up, estimates of the spatial Durbin model of the Barro regression are consistent with the expected spatial impact of initial wages on final real wages. The parameters for spatial lags of both initial and final wages are significant and quantitatively large.

5.4 Model comparison

My estimates of the SDM provided in Tables 2 and 3 depend on cutoff distances. Starting from $C = 900$ km, the estimates are based on the same number of cities. Bayesian posterior model probabilities are shown in Table 4. These probabilities have been calculated using information stored during the sampling of the log-marginal density vectors for each model (i.e., for each cutoff distance). They can be used for the purpose of model comparison.

The highest posterior probability for the regression using the first time period corresponds to the case of $C = 8000$ km, i.e., it captures spatial effects using all cities. For the second time period, the radius of spatial interaction is much smaller at $C = 1000$ km. It should be stressed that this does not mean restricting the spatial impacts only to cities located within 1000 km or geographically closer. The spatial matrix is connected, and in the case of the SDM, the effect can arise from neighbors of neighbors, with corresponding reduced spatial weights.

Speed of spatial convergence The dependence of the dependent variable ($y_{t+\tau}$) on its initial level (y_0) disappears for large τ . The time required for this is usually analyzed based on the so-called half-life to convergence: $HL = \tau \cdot T$, where the coefficient τ stands for a number of T -year-long periods it takes to reach a half-way point on the path to the new steady-state equilibrium, $\tau = -\ln 2 / \ln(1 + \beta)$. As noted, the coefficients of the SDM cannot be interpreted as a measure of a regressor's impact on the dependent variable. For this purpose, I use estimates of the direct effects of initial wages instead of $1 + T\beta$, compute a spatial analogue of β and substitute it into the formula for τ and HL for both parts of the dataset. See Tables 2 and 3 for values of τ .

Choosing the best model in terms of posterior probabilities (column 10 of Table 2), I find that the half-life to convergence is 14.4 rounds of 9 years for the non-FN cities, and 19.9 rounds of 9 years for FN cities. The best model for the second time period corresponds to $C = 1000$ km (column 6 of Table 3), which implies 12 rounds of 7 years for the non-FN cities and 15.4 rounds for FN cities. Hence, my estimation results reveal that the Far North real wages converge more slowly than in the rest of the country during both periods.

During the first period, commodity prices were increasing rapidly. The commodity boom reached its peak in 2007, after that commodity prices fell sharply (Åslund 2013).

During boom periods, commodity bases seem to drive divergence (we observe a slower convergence speed during the first time span), but that effect disappears in times of commodity turndowns.

In the presence of amenity differentials, one can never see full convergence. Therefore, it is hard to state when convergence will be achieved, since there is no information on which equilibrium real wage differences are compatible with spatial equilibrium.

6 Conclusion

This paper tests the convergence in real wages across Russian cities and studies the impact of spatial proximity on the speed of convergence. I find conditional sigma-convergence of real wages in spatially close cities and show that this result is robust to different cutoff distances and the exclusion of remote towns. I also show that there is spatial beta-convergence of real city wages. The impact of the public sector, the share of which is still large in Russia, on real wage growth is negative. The impact of natural resources on real wage growth was significantly positive during 1996–2005, and this finding is consistent with earlier studies. However, this effect disappeared after 2005, when the share of extractive industries in GDP was decreasing. I also find that settlements of the Far North with a special state wage policy were converging more slowly than the non-Far North cities, including western cities.

A final comment is in order. Little work has been done on city-level income convergence in the Russian economy until now. This is arguably due to poor data quality, which precludes the construction of many required control variables. Using location patterns as a source of additional information, and applying spatial econometric techniques, yields a way to alleviate some of these difficulties.

Acknowledgements I owe special thanks to Kristian Behrens, Philip Ushchev, an anonymous referee of the HSE Working Paper Series in Economics, the editor, and two anonymous referees for their detailed comments and suggestions. Earlier versions of this paper were presented at 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; comments by participants are greatly appreciated. I am also grateful to Giuseppe Arbia, Olga Demidova, Evgenia Kolomak, Leonid Limonov, Florian Mayneris, Tatiana Mikhailova, Eirini Tatsi, Jacques Thisse, Elena Vakulenko, Natalia Volchkova, and Jeffrey Wooldridge for their advice on some of the issues considered in this paper. The study has been funded by the Russian Academic Excellence Project ‘5-100’.

Compliance with ethical standards

Conflict of interest No potential conflict of interest was reported by the author.

Appendix 1: empirical papers on regional income convergence using spatial Russian data

See Table 5.

Table 5 Literature on Russian regional income convergence using spatial data

Paper	Time period	Indicator	Space	Results
Berkowitz and De Jong (2003)	1993–1997	Personal income	Control variable “distance from Moscow”	β -divergence
Ahrend (2005)	1990–1998	GRP, personal income, industrial production	Regressions for Russia and for its European part	β -convergence and β -divergence, depending on the indicator
Carluer (2005)	1985–1999	Personal income	Geographical dummies and distance from Moscow	β -convergence and β -divergence, depending on the time period. Convergence clubs
Buccellato (2007)	1999–2004	Personal income	SAR, SEM, W : contiguity	β -convergence
Lugovoy et al. (2007)	1996–2004	GRP	Moran's <i>I</i> . SEM, SDM, W : traveling time, inverse distance	σ -convergence and σ -divergence, depending on the time period and methodology. Conditional β -convergence
Solanko (2008)	1992–2005	Personal income	Distance from Moscow	σ -divergence, β -convergence
Zverev and Kolomak (2010)	1995–2006	GRP, budget revenue, budget expenses	Moran's <i>I</i> . SEM, SAR, W : inverse distances, contiguity	σ -divergence. β -convergence and β -divergence, depending on the indicator
Kholodilin et al. (2012)	1998–2006	GRP	Moran's <i>I</i> . SEM, SAR, W : inverse squared distances	Weak β -convergence. Strong β -convergence among high income regions and among low income regions
Ivanova (2014)	1996–2012	Personal income	Moran's <i>I</i> . SAR, W : contiguity, inverse great circle distances with cutoffs, inverse railway distances	σ -convergence, β -convergence

GRP, gross regional product; SAR, spatial autoregressive model; SEM, spatial error model; SDM, spatial Durbin model; **W**, spatial weight matrix

Appendix 2: Proof of Proposition 1

To prove part (i), we use the chain rule to obtain

$$\frac{d\mathbf{y}}{dt} = \frac{\partial \mathbf{y}}{\partial \boldsymbol{\lambda}} \cdot \frac{d\boldsymbol{\lambda}}{dt}.$$

Combining this with (2) yields

$$\frac{d\mathbf{y}}{dt} = \mathbf{H}(\boldsymbol{\lambda}, \mathbf{D}), \tag{9}$$

where

$$\mathbf{H}(\boldsymbol{\lambda}, \mathbf{D}) \equiv \frac{\partial \mathbf{y}}{\partial \boldsymbol{\lambda}} \cdot \mathbf{F}(\boldsymbol{\lambda}, \mathbf{D}) \cdot \mathbf{V}(\boldsymbol{\lambda}, \mathbf{D}).$$

As discussed above, the mapping $\mathbf{y}(\boldsymbol{\lambda}, \mathbf{D})$ is invertible, so we can define

$$\mathbb{G}(\mathbf{y}, \mathbf{D}) \equiv (\mathbf{H} \circ \mathbf{y}^{-1})(\mathbf{y}, \mathbf{D}),$$

where $\mathbf{y}^{-1}(\cdot)$ is the mapping inverse to $\mathbf{y}(\boldsymbol{\lambda}, \mathbf{D})$ for each given distance matrix \mathbf{D} , while \circ denotes composition of mappings. Doing so and setting

$$g_i(y_i, \mathbf{y}_{-i}, \mathbf{D}) \equiv \ln G_i(y_i, \mathbf{y}_{-i}, \mathbf{D}),$$

we find that (9) becomes (4). This completes the proof of part (i).

To prove part (ii), observe first that uniqueness of the steady state follows immediately from the necessity and sufficiency of real wage equalization for a spatial configuration to be a long-run equilibrium (see the discussion after equation (3)). To show stability, let $\mathcal{U} \in \Delta_{n-1}$ be an open neighborhood of $\boldsymbol{\lambda}^*$, such that any solution $\boldsymbol{\lambda}(t)$ to (4) starting in \mathcal{U} converges asymptotically to $\boldsymbol{\lambda}^*$. Set $\mathcal{V} \equiv \mathbf{y}(\mathcal{U})$. Clearly, \mathcal{V} is an open neighborhood of \mathbf{y}^* . Furthermore, by definition of \mathcal{V} for any solution $\mathbf{y}(t)$ to (4) such that $\mathbf{y}(0) \in \mathcal{V}$ there exists a solution $\boldsymbol{\lambda}(t)$ to (2) such that $\boldsymbol{\lambda}(0) \in \mathcal{U}$. Because $\boldsymbol{\lambda}(t)$ asymptotically converges to $\boldsymbol{\lambda}^*$, and because $\mathbf{y}(\boldsymbol{\lambda}, \mathbf{D})$ is continuous, it must be that $\mathbf{y}(t)$ asymptotically converges to \mathbf{y}^* . This means \mathbf{y}^* is a stable steady state. This completes the proof of part (ii).

To prove part (iii), we log-linearize the system (4) in the vicinity of \mathbf{y}^* . This yields

$$\frac{d \ln y_i}{dt} = \sum_{j=1}^n b_{ij}(\mathbf{D})(\ln y_j - \ln y_j^*), \tag{10}$$

where b_{ij} are defined by

$$b_{ij}(\mathbf{D}) = g_i(\mathbf{y}^*) \cdot \left. \frac{\partial \ln g_i}{\partial \ln y_j} \right|_{\mathbf{y}=\mathbf{y}^*}.$$

Solving (10), we obtain

$$\ln \mathbf{y}(t) - \ln \mathbf{y}(0) = [\mathbf{I} - \exp(t \cdot \mathbf{B}(\mathbf{D}))] \cdot [\ln \mathbf{y}^* - \ln \mathbf{y}(0)], \tag{11}$$

where

$$\mathbf{B}(\mathbf{D}) \equiv (b_{ij}(\mathbf{D}))_{i,j=1,\dots,n},$$

while $\exp[t \cdot \mathbf{B}(\mathbf{D})]$ is the matrix exponential defined as follows:

$$\exp[t \cdot \mathbf{B}(\mathbf{D})] \equiv \mathbf{I} + t \cdot \mathbf{B}(\mathbf{D}) + \frac{t^2}{2!} \cdot [\mathbf{B}(\mathbf{D})]^2 + \dots$$

Fix some moment of time $T > 0$ and set:

$$\mathbf{W}(\mathbf{D}) \equiv \mathbf{I} - \exp[T \cdot \mathbf{B}(\mathbf{D})], \quad \boldsymbol{\alpha} \equiv \mathbf{W}(\mathbf{D}) \ln \mathbf{y}^*.$$

Denote the diagonal entries $w_{ii}(\mathbf{D})$ of the matrix $\mathbf{W}(\mathbf{D})$ by β_i . Using this notation and stating (11) in coordinate form yields (5). This completes the proof of part (iii) and proves the whole proposition.

Appendix 3: descriptive statistics

See Fig. 6 and Table 6.

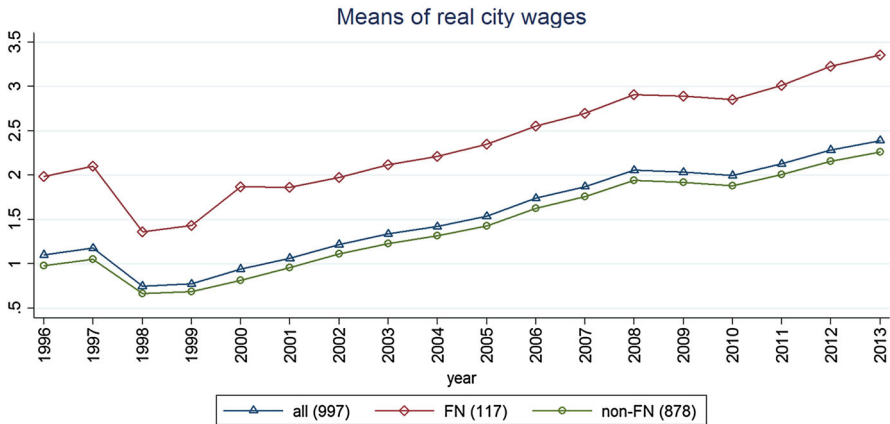


Fig. 6 Means of real city wages. Number of cities in parentheses

Table 6 Correlations between real city wages by regions and real GRP per capita

Year	Pearson's			Spearman's			
	All	Non-FN	FN	All	Non-FN	FN	
1996	0.760	0.563	0.886	0.561	0.573	0.421	
1997	0.788	0.595	0.902	0.594	0.615	0.339	
1998	0.915	0.688	0.958	0.806	0.734	0.926	
1999	0.920	0.718	0.976	0.846	0.801	0.760	
2000	0.930	0.650	0.987	0.838	0.785	0.856	
2001	0.882	0.694	0.946	0.867	0.830	0.809	
2002	0.841	0.645	0.925	0.824	0.809	0.733	
2003	0.835	0.634	0.938	0.818	0.773	0.812	
2004	0.817	0.688	0.907	0.825	0.796	0.833	
2005	0.779	0.710	0.865	0.821	0.831	0.714	
2006	0.797	0.728	0.894	0.830	0.849	0.728	
2007	0.801	0.759	0.849	0.819	0.836	0.705	
2008	0.784	0.760	0.857	0.787	0.803	0.788	
2009	0.785	0.761	0.840	0.783	0.802	0.851	
2010	0.789	0.796	0.794	0.795	0.840	0.739	
2011	0.767	0.794	0.750	0.790	0.829	0.756	
2012	0.779	0.783	0.806	0.775	0.814	0.800	
2013	0.753	0.829	0.757	0.749	0.842	0.725	
All coefficients are significant at $p < 0.001$	<i>n</i>	78	59	19	78	59	19

Appendix 4: panel unit root test results

See Tables 7 and 8.

Table 7 HT panel unit root test results

<i>C</i> (km)	All (1) 300	Non-FN (2) 300	FN (3) 300	All (4) 400	Non-FN (5) 400	FN (6) 400	All (7) 500	Non-FN (8) 500	FN (9) 500	All (10) 900	Non-FN (11) 900	FN (12) 900
Time trend: NO												
1996–2005												
<i>b</i>	0.67	0.62	0.68	0.66	0.59	0.64	0.65	0.57	0.66	0.66	0.52	0.82
<i>z</i>	-3.02	-7.90	-0.65	-4.62	-10.3	-1.86	-5.36	-12.8	-1.27	-4.57	-17.6	4.16
<i>p</i>	0.00	0.00	0.26	0.00	0.00	0.03	0.00	0.00	0.10	0.00	0.00	1.00
2006–2013												
<i>b</i>	0.68	0.68	0.69	0.68	0.67	0.72	0.68	0.65	0.75	0.55	0.55	0.56
<i>z</i>	4.80	3.97	1.61	4.94	3.49	2.68	4.90	1.79	3.62	-6.43	-6.22	-1.83
<i>p</i>	1.00	1.00	0.95	1.00	1.00	0.99	1.00	0.96	1.00	0.00	0.00	0.03
Time trend: YES												
1996–2005												
<i>b</i>	0.39	0.37	0.30	0.36	0.33	0.35	0.34	0.29	0.34	0.23	0.18	0.29
<i>z</i>	5.50	3.55	-0.32	3.25	0.98	0.84	1.98	-2.00	0.66	-7.12	-10.7	-0.60
<i>p</i>	1.00	1.00	0.37	1.00	0.84	0.80	0.98	0.02	0.75	0.00	0.00	0.27
2006–2013												
<i>b</i>	0.96	0.95	0.97	0.96	0.95	0.96	0.96	0.95	0.96	0.96	0.97	0.97
<i>z</i>	52.61	49.9	15.5	54.8	51.0	17.8	55.7	51.6	18.5	56.29	52.7	19.4
<i>p</i>	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
<i>n</i>	892	816	76	960	857	103	987	878	109	997	878	119

Table 7 continued

C (km)	(13) 2000	(14) 2000	(15) 2000	(16) 4000	(17) 4000	(18) 4000	(19) 6000	(20) 6000	(21) 6000	(22) 8000	(23) 8000	(24) 8000
Time trend: NO												
1996–2005												
<i>b</i>	0.61	0.61	0.56	0.61	0.6	0.68	0.61	0.6	0.68	0.61	0.59	0.68
<i>z</i>	-9.53	-8.47	-4.99	-9.21	-9.96	-0.86	-9.67	-10.3	-0.58	-9.83	-10.4	-0.82
<i>p</i>	0.00	0.00	0.00	0.00	0.00	0.19	0.00	0.00	0.28	0.00	0.00	0.21
2006–2013												
<i>b</i>	0.46	0.46	0.43	0.48	0.48	0.47	0.49	0.49	0.5	0.49	0.49	0.51
<i>z</i>	-14.2	-13.4	-5.81	-12.2	-11.7	-4.52	-11.1	-10.7	-3.7	-11.2	-10.9	-3.39
<i>p</i>	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Time trend: YES												
1996–2005												
<i>b</i>	0.21	0.23	0.09	0.23	0.22	0.30	0.22	0.21	0.28	0.21	0.2	0.28
<i>z</i>	-9.07	-6.8	-6.6	-7.38	-7.81	-0.47	-8.49	-8.65	-0.97	-8.61	-8.79	-1.03
<i>p</i>	0.00	0.00	0.00	0.00	0.00	0.32	0.00	0.00	0.16	0.00	0.00	0.15
2006–2013												
<i>b</i>	0.98	0.97	1.00	0.98	0.97	1.00	0.97	0.97	1.00	0.97	0.97	1.00
<i>z</i>	57.1	53.1	20.3	57.0	53.2	20.2	56.7	52.9	20.1	56.5	52.7	20.1
<i>p</i>	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
<i>n</i>	997	878	119	997	878	119	997	878	119	997	878	119

Table 8 IPS panel unit root test results

<i>C</i> (km)	All (1)	Non-FN (2)	FN (3)	All (4)	Non-FN (5)	FN (6)	All (7)	Non-FN (8)	FN (9)	All (10)	Non-FN (11)	FN (12)
Time trend: NO												
1996–2005												
<i>W</i>	-20.47	-26.7	1.06	-19.61	-26.5	-1.76	-19.2	-27.3	-1.19	-17.9	-14.7	1.77
<i>p</i>	0.00	0.00	0.85	0.00	0.00	0.04	0.00	0.00	0.12	0.00	0.00	0.96
2006–2013												
<i>W</i>	-2.60	-0.10	-9.96	-4.77	-1.76	-4.95	-4.35	-1.66	-2.63	-18.45	-14.9	-10.5
<i>p</i>	0.00	0.46	0.00	0.00	0.04	0.00	0.00	0.05	0.00	0.00	0.00	0.00
Time trend: YES												
1996–2005												
<i>W</i>	-14.12	-13.1	-3.06	-11.7	-11.8	-2.27	-16.15	-16.4	-2.78	-17.5	-16.6	-7.66
<i>p</i>	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00
2006–2013												
<i>W</i>	-26.8	-29.7	-17.3	-33.47	-30.95	-16.8	-43.3	-34.95	-12.0	-30.3	-29.7	-10.23
<i>p</i>	0.00	0.00	0.08	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
<i>n</i>	892	816	76	960	857	103	987	878	109	997	878	119
<i>C</i> (km)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)	(21)	(22)	(23)	(24)
	2000	2000	2000	4000	4000	4000	6000	6000	6000	8000	8000	8000
Time trend: NO												
1996–2005												
<i>W</i>	-14.3	-14.0	-2.87	-15.5	-15.6	-2.26	-16.7	-16.2	-2.28	-17.2	-16.9	-2.67

Table 8 continued

C (km)	(13) 2000	(14) 2000	(15) 2000	(16) 4000	(17) 4000	(18) 4000	(19) 6000	(20) 6000	(21) 6000	(22) 8000	(23) 8000	(24) 8000
2006–2013												
W	-17.8	-16.0	-4.66	-15.8	-15.0	-5.35	-16.3	-14.6	-6.14	-15.8	-14.2	-5.33
Time trend: YES												
1996–2005												
W	-12.8	-13.2	-2.96	-14.5	-15.5	-4.60	-15.0	-15.6	-2.57	-15.6	-16.1	-3.55
2006–2013												
W	-20.2	-19.0	-3.11	-16.0	-15.4	-2.99	-20.7	-19.2	-6.02	-19.7	-19.4	-7.46

In columns (13)–(24): all $p < 0.01$; $n = 997$, $n^{nonFN} = 878$, $n^{FN} = 119$

Appendix 5: Bayesian model estimation

Consider an SDM

$$\mathbf{y} = \mathbf{X}\beta + \rho\mathbf{W}\mathbf{y} + \theta\mathbf{W}\mathbf{X} + \varepsilon \quad (12)$$

here \mathbf{X} is an $n \times k$ matrix of explanatory variables, β is a $k \times 1$ parameter vector, \mathbf{W} is a $n \times n$ spatial matrix, ρ and θ are spatial scalar parameters, ε is an $n \times 1$ disturbance vector, $\varepsilon \sim N(0, \sigma^2\mathbf{I}_n)$, here \mathbf{I}_n is the identity matrix of size n . An intercept parameter in (12) is omitted for simplicity.

Following LeSage (1997), LeSage and Parent (2007), I 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).

I carry out 100,000 draws of the Metropolis within Gibbs sampling procedure, excluding the first 10,000 draws, to produce posterior estimates for each model in Tables 2 and 3. Estimates are mainly based on a spatial toolbox developed by LeSage (2001) for MATLAB software. 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), and s is a probability associated with r (e.g., 0.90). The above-mentioned number of draws of the Metropolis within Gibbs sampling procedure is sufficient for $q = 0.001$, $r = 0.0005$, and $s = 0.999$ for all the estimated models. Significance levels provided in Tables 2 and 3 are derived from simple descriptive statistics of the Markov chain iterations for each sample.

References

- Abreu M, De Groot HL, Florax RJ (2005) Space and growth: a survey of empirical evidence and methods. *Rég Dev* 21:13–44
- Ahrend R (2005) Speed of reform, initial conditions or political orientation? Explaining Russian regions' economic performance. *Post Communist Econ* 17(3):289–317
- Arbia G (2006) *Spatial econometrics: statistical foundations and applications to regional convergence*. Springer, Berlin
- Armstrong HW (1995) Convergence among regions of the European Union, 1950–1990. *Pap Reg Sci* 74(2):143–152
- Åslund A (2013) *How capitalism was built: the transformation of Central and Eastern Europe, Russia, the Caucasus, and Central Asia*. Cambridge University Press, Cambridge
- Barro RJ (1991) Economic growth in a cross section of countries. *Q J Econ* 106(2):407–443
- Barro RJ, Sala-i-Martin X (1992) Regional growth and migration: a Japan–United States comparison. *J Jpn Int Econ* 6(4):312–346
- Baumol WJ (1986) Productivity growth, convergence, and welfare: what the long-run data show. *Am Econ Rev* 76(5):107–285
- Berkowitz D, De Jong DN (2003) Policy reform and growth in post-Soviet Russia. *Eur Econ Rev* 47(2):337–352
- Bernat GA (1996) Does manufacturing matter? A spatial econometric view of Kaldor's laws. *J Reg Sci* 36(3):463–477
- Buccellato T (2007) Convergence across Russian regions: a spatial econometrics approach. *Economics Working Paper 72*, Centre for the Study of Economic and Social Change in Europe. UCL, London
- Carlier F (2005) Dynamics of Russian regional clubs: the time of divergence. *Reg Stud* 39(6):713–726

- Demidova O (2015) Spatial effects for the eastern and western regions of Russia: a comparative analysis. *Int J Econ Policy Emerg Econ* 8(2):153–168
- Ertur C, Koch W (2007) Growth, technological interdependence and spatial externalities: theory and evidence. *J Appl Econ* 22(6):1033–1062
- Federal Law (2004) Federal Law 122-FZ of August 22, 2004. Moscow. <http://base.garant.ru/12136676/>. Accessed 15 Jan 2017
- Fingleton B, López-Bazo E (2006) Empirical growth models with spatial effects. *Pap Reg Sci* 85(2):177–198
- Fingleton B, McCombie JSL (1998) Increasing returns and economic growth: some evidence for manufacturing from the European Union regions. *Oxf Econ Pap* 50(1):89–105
- Glushchenko K (2010) Studies on income inequality between Russian regions. *Reg Econ Sociol* 4:88–119 (in Russian). http://www.sibran.ru/journals/issue.php?ID=120718&ARTICLE_ID=130871. Accessed 20 July 2016
- Gluschenko K (2012) Myths about beta-convergence. *J New Econ Assoc* 16(4):26–44 (in Russian). <http://journal.econorus.org/pdf/NEA-16.pdf>. Accessed 1 Oct 2015
- Guriev S, Vakulenko E (2012) Convergence between Russian regions. WP 180. Center for Economic and Financial Research (CEFIR), Moscow
- Harris RDF, Tzavalis E (1999) Inference for unit roots in dynamic panels where the time dimension is fixed. *J Econom* 91(2):201–226
- Harris R, Moffat J, Kravtsova V (2011) In search of ‘W’. *Spat Econ Anal* 6(3):249–270
- Im KS, Pesaran MH, Shin Y (2003) Testing for unit roots in heterogeneous panels. *J Econom* 115(1):53–74
- Ivanova VI (2014) Regional convergence of income: spatial analysis. *Spat Econ* 4:100–119. <https://doi.org/10.14530/se.2014.4.100-119> (in Russian)
- Kholodilin KA, Oshchepkov A, Siliverstovs B (2012) The Russian regional convergence process: Where is it leading? *East Eur Econ* 50(3):5–26
- Krugman P (1991) Increasing returns and economic geography. *J Polit Econ* 99(3):483–499
- Le Gallo J, Fingleton B (2014) Regional growth and convergence empirics. In: Fischer MM, Nijkamp P (eds) *Handbook of regional science*. Springer, Berlin, pp 291–315
- LeSage JP (1997) Bayesian estimation of spatial autoregressive models. *Int Reg Sci Rev* 20(1–2):113–129
- LeSage JP (2001) *Econometrics toolbox. MATLAB software*. <http://www.spatial-econometrics.com/>. Accessed 1 October 2015
- LeSage JP, Fischer MM (2008) Spatial growth regressions: model specification, estimation and interpretation. *Spat Econ Anal* 3(3):275–304
- LeSage JP, Pace RK (2009) *Introduction to spatial econometrics*. Chapman & Hall/CRC, Boca Raton, FL
- LeSage JP, Parent O (2007) Bayesian model averaging for spatial econometric models. *Geogr Anal* 39(3):241–267
- 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
- Maddala GS, Wu S (1999) A comparative study of unit root tests with panel data and a new simple test. *Oxf Bull Econ Stat* 61(S1):631–652
- Ponomareva M, Zhuravskaya E (2004) Federal tax arrears in Russia: Liquidity problems, federal redistribution or regional resistance? *Econ Transit* 12(3):373–398
- Quah D (1993) Galton’s fallacy and tests of the convergence hypothesis. *Scand J Econ* 95(4):427–443
- Remington TF (2015) Why is interregional inequality in Russia and China not falling? *Communist Post Commun* 48(1):1–13
- Rey S, Montouri B (1999) US regional income convergence: a spatial econometric perspective. *Reg Stud* 33(2):143–156
- Rey SJ, Le Gallo J (2009) Spatial analysis of economic convergence. In: Mills TC, Patterson K (eds) *Palgrave handbook of econometrics, vol 2, applied econometrics*. Palgrave Macmillan, Basingstoke, pp 1251–1290
- Roback J (1982) Wages, rents, and the quality of life. *J Polit Econ* 90(6):1257–1278
- Sala-i-Martin XX (1996) Regional cohesion: evidence and theories of regional growth and convergence. *Eur Econ Rev* 40(6):1325–1352
- Sardadvar S, Vakulenko E (2016) Interregional migration within Russia and its east-west divide: evidence from spatial panel regressions. *Rev Urb Reg Dev Stud* 28(2):123–141
- Sardadvar S, Vakulenko E (2017) A model of interregional migration under the presence of natural resources: theory and evidence from Russia. *Ann Reg Sci* 59(2):535–569

- Solanko L (2008) Unequal fortunes: a note on income convergence across Russian regions. *Post Communist Econ* 20(3):287–301
- Tabuchi T, Thisse JF (2002) Taste heterogeneity, labor mobility and economic geography. *J Dev Econ* 69(1):155–177
- Tabuchi T, Thisse JF, Zeng DZ (2005) On the number and size of cities. *J Econ Geogr* 5(4):423–448
- Tabuchi T, Thisse JF, Zhu X (2016) Does technological progress magnify regional disparities? DP 599. Institute of Developing Economies, Japan External Trade Organization (JETRO), Tokyo
- Taylor AM (1999) Sources of convergence in the late nineteenth century. *Eur Econ Rev* 43(9):1621–1645
- Wengle S, Rasell M (2008) The monetisation of l'goty: changing patterns of welfare politics and provision in Russia. *Eur Asia Stud* 60(5):739–756
- Williamson JG (1996) Globalization, convergence, and history. *J Econ Hist* 56(2):277–306
- Zubarevich N (2015) Regional inequality and potential for modernization. In: Oxenstierna S (ed) *The challenges for Russia's politicized economic system*. Routledge, London, pp 182–201
- Zverev DV, Kolomak EA (2010) 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, Moscow (in Russian)