NATIONAL RESEARCH UNIVERSITY
HIGHER SCHOOL OF ECONOMICS

Alexander S. Skorobogatov

AN ONGOING REVERSAL OF FORTUNE AMONG RUSSIAN CITIES: CITY AGE, NATURAL RESOURCES, AND CHANGING SPATIAL INCOME DISTRIBUTION

BASIC RESEARCH PROGRAM

WORKING PAPERS

SERIES: ECONOMICS
WP BRP 60/EC/2014

Alexander S. Skorobogatov[*]

AN ONGOING REVERSAL OF FORTUNE AMONG RUSSIAN CITIES: CITY AGE, NATURAL RESOURCES, AND CHANGING SPATIAL INCOME DISTRIBUTION[**]

This paper documents the negative link between the age of Russian cities and their average wage. This link is robust to various definitions of city age and sample censuring, the inclusion of regional and time fixed effects, dependent variable spatial lag and many urban characteristics. This link is revealed especially for cities founded after the Soviet industrialization and for upper quintiles of cities by their average wage. To determine a mechanism behind the established fact, hypotheses as to spatial patterns of economic performance are discussed, including the increasing return hypothesis, the institutions hypothesis and the geography hypothesis. Following the sophisticated version of the geography hypothesis, a model of growth in n-city and two-sector economy is developed. The model replicates the link between age and per capita income and contains testable hypotheses that enable one to check whether a mechanism outlined in the model is behind the link between city age and wage. Our empirical strategy is based on a quasi-experiment, in which the treatment effect is made by time and various age groups of the cities are broken up into treatment and control groups. The results are strongly in favor of the sophisticated geography hypothesis. The revealed mechanism suggests that the changing spatial patterns of wage differentials are explained by the changing remaining stocks of natural resources. Older cities are getting relatively poorer due to the shrinking of their remaining resource stocks, while new cities are emerging in resource-rich territories with the respective income advantages.

JEL Classification: R12, Q32, O13

Keywords: quasi-experiment, exhaustible resources, internal colonization, differences-in-differences, urban development

[*] National Research University Higher School of Economics (Saint-Petersburg, Russia)
e-mail: skorobogatov@mail.ru

[**] Earlier drafts of this paper were presented in the Second International Conference «Industrial Organization and Spatial Economics», Saint-Petersburg, 27-28 June 2013, and the International Workshop «Natural Resources, Environment, Urban Economics, International Trade and Industrial Organization», Saint-Petersburg, 02-04 October 2013. For valuable comments and suggestions the author is grateful to conferences participants, and especially to Sergey Izmalkov, Grigory Kosenok, Takatoshi Tabucci, Jacques François Thisse, Dao-Zhi Zeng, Ives Zenou. All the views and errors are of the author alone.

History of Russia... [is] the history of a country for ever undergoing colonisation.

Kluchevsky (1911)

1. Introduction

Russian territories display a kind of reversal of fortune. While ancient and famous Russian cities are relatively poor, those founded just a few decades ago are relatively rich and frequently the most flourishing cities. The youngest Russian cities in 2011 were only seven years old, while the age of the most ancient ones was well over a thousand years. The latter include such historical cities of Ancient Rus as Veliky Novgorod, Smolensk, Pskov, Rostov, Vladimir, Bryansk, Belozersk, and Murom. What is important here is that all these cities are not in ruins; quite opposite, they are living cities and relatively densely populated ones. Nevertheless, they are much poorer than many adolescent cities.

The average wage across the five youngest cities in Russia is more than twice that across the five oldest.[1] Examples of the young and rich city and the old and poor one are Nadym being 39 years old and having an average monthly wage of 72,551 rubles and Yelnya with 800 years and 10,671 rubles, respectively. Although among the more than 1,000 cities there are old and rich ones, such as Moscow and other megalopolises, and young and poor ones, the general tendency is strongly toward a negative correlation between city age and wage.

This tendency can be illustrated in maps of modern Russia. To compare spatial allocation of these two urban features – age and wage, – we constructed two regional maps in which color denotes averages of city age and wage for respective regions.[2] Map A of Figure 1 visualizes the mean age of cities across Russian regions. It is readily seen that the vast majority of the oldest regions from the standpoint of mean age of their cities lie in European Russia. Apart from this most densely populated territory, 'old' regions are occupied also in the extreme western part of modern Russia, in Kaliningrad Oblast, with its ancient Prussian cities; in the extreme southern part of the North Caucasus district, in Dagestan where Derbent – one of the most ancient cities in the modern Russian Federation – is situated; in south-western Siberia – the Omsk and Tyumen Oblasts that were among the first lands where Russians began to settle in the process of opening and developing Siberia in 16th-17th centuries; and in the far east.

At the opposite end of the spectrum there are four regions in western Siberia, two regions in southern Siberia, regions in the north Caucasus and in the far East. Unlike the regions of European Russia these regions began to be developed in the Soviet period. For example, Magadan Oblast began to be settled under Stalin in connection with the creation of camps designed to exploit forced labor for extracting gold in the rich deposits of that region. Another example is the oil and gas regions of the western Siberia that were intensively developed mostly under Brezhnev to realize their rich deposits of the fuel resources, in particular, after the oil prices jumps in the world markets.

In the map B of Figure 2 we can see a spatial pattern of the average wages. What is clear from the map is that poorest regions are just the oldest ones, and vice versa. All regions of European Russia, excluding the two capital cities and their neighboring regions as well as the extreme north regions, are among the poorest. It is true also for Kaliningrad Oblast in the extreme western part and for Dagestan in the south. At the same time, the richest regions are just those in Siberia and the Far East with their cities founded in the late Soviet and even post-Soviet periods.

Finally, this link generally holds not only across regions, but within them too. For example, Kursk Oblast displays similar intra-regional disparities between its cities in terms of their age and wage as illustrated in Figure 2. Here the link of our interest is best exemplified by a pair of cities: one is Kurchatov that was given city status in 1983 and is the richest city of the region; other is Rylsk that is both the oldest city of the region (and one of the most ancient in Russia) and the poorest city of that region. As a whole, controlling for regional dummies and allowing for the respective autocorrelation

[1] This ratio is in excess of two for three age variables among the four at hand, while for the remaining age variable this ratio is 1.99. For the description of the age variables see below.

[2] We use choropleth maps rather than proportional symbols ones as in this case the former make the allocation patterns more visual. The respective proportional symbols maps would contain a great number of visually indistinguishable points.

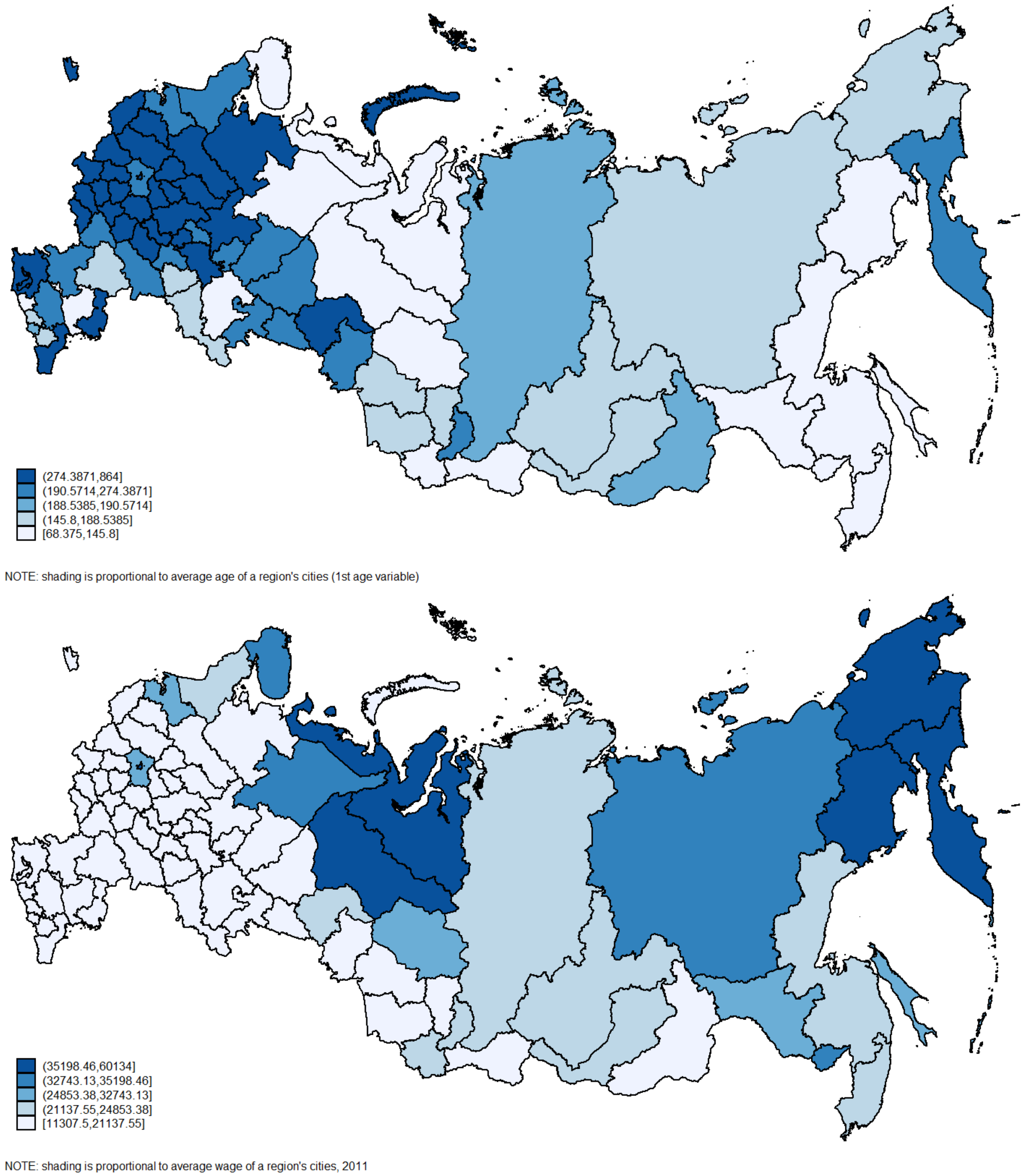

Figure 1. Russian regions shaded by their average city age and wage

in estimating the standard errors we still have a significant negative link between the age and the average wage.[3]

Since pioneering works of Krugman (1991a, b) one of the most influential approaches to the analysis of the spatial patterns of economic activities and their results has been the New Economic Geography. As Davis and Weinstein noted in their discussion of this approach (termed in their paper the increasing returns hypothesis), it suggests, in particular, that "an early start in one location provides that site with advantages at each succeeding stage of locational competition" (2002,

[3]Among 85 Russian subjects, including the sub-regions, there are those featuring a negative link and no significant link and even a positive link. However, the overall association after controlling for regional fixed effects is significantly negative.

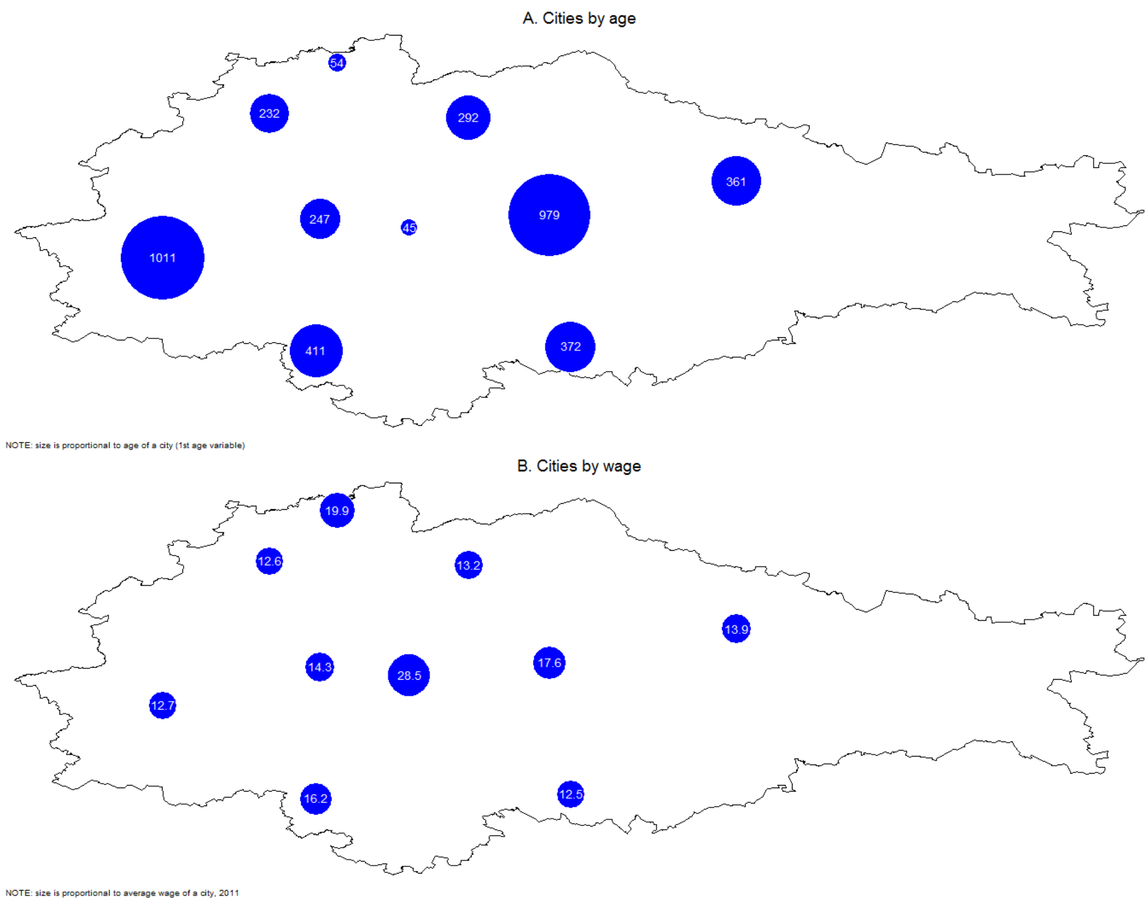

Figure 2. Age and wage across cities of Kursk Oblast

p. 1270). These advantages of older locations are based on the mere fact that the accumulation of population is, on the one hand, a key source of increasing returns and, on the other hand, a process that takes a long time, suggesting that older locations, other things being equal, should be richer than younger ones.

In light of that implication a strongly negative link between age and wage across Russian cities seems puzzling. And examining it could provide some new evidence that would be informative for the ongoing debate in the economics literature concerning the forces behind spatial patterns of economic activities. The agenda here includes questions such as why economic activity is unevenly distributed across space in terms of its intensity and rewards (Combes et al. 2008); or what drives changes through time periods in terms of interregional income inequalities. There are three popular hypotheses to answer these questions. Apart from the just mentioned increasing returns hypothesis, other important hypotheses concerning spatial income distribution are the institutions hypothesis and the geography hypothesis. The latter two were termed and discussed by Acemoglu et al. (2002) when addressing an issue of cross-country differences in economic performance.

Given the negative link between age and wage, and these hypotheses, there are two related questions. The first question is how important age is as a determinant of urban characteristics which are given a crucial role in these hypotheses. In other words, we are interested in a static link between age and the measures of agglomeration effects, the quality of institutions, and time-invariant and time-varying geographical features. The second question is how these features change with time for cities of different ages. This question is important here because, as seen below, the most intense negative link takes place for younger cities: the younger a city, the faster the relative growth of its age year to year; at the same time, current dynamics may reflect the relative age increase, and such a localization of the link implies that the year-to-year age increase may affect younger cities more. An additional year of age may heavily matter for subsample of cities of, say, less than fifty, while being less important for the subsample of older cities or for the whole sample. Thus, we are also concerned with the dynamic link between the age and the urban features.

A key assumption underlying our empirical strategy is that the force generating our data is to replicate the link between age and wage both in statics and dynamics. If some force did replicate the link, it would be evidence in favor of the respective hypothesis. Based on this assumption we can use a method of quasi-experiment. What makes the treatment effect is time measured in years. The treatment and the control groups are age groups of the cities. The youngest cities are to be under the strongest effect of the underlying force, but this effect is to decrease relatively with time. We pick out two age groups of youngest and second youngest cities. We observe that these groups are the richest and second richest, and their relative wage is diminishing with most rapid and second most rapid pace year to year. Hence, the force underlying these regularities is to behave in the same way.

To the best of our knowledge, this paper is the first to document and explore the negative link between age and wage on data about Russia or any other country. As is noted by Giesen and Suedekum, age, while being a fundamental dimension of cities, was neglected in the spatial economics literature for the most part (2012, p. 2). Their own paper is one of the few exceptions, but they address a different issue. In particular, they examined relationships between age and size using data on US cities. One of their results is that older American cities as a rule are larger than younger ones. A similar result concerning US cities was presented in Dobkins and Ioannides (2001). The age of a city turned to be related to the probability of having neighboring cities, i.e., the older a city, the more probable a constellation of cities was in the immediate vicinity. Thus, in the US age is a positive predictor of the size of a city and the larger agglomeration around it. A kind of non-linear relationship between age and size is implied in Michaels et al. (2012). They documented a positive relationship between 1880 population density and its subsequent growth for the middle-size locations. Glaeser and Kahn (2001) examined the age effect on the decentralization of cities with the results that this effect was very weak. As a whole, previous work that explicitly used age as an urban characteristic had a different focus and was devoted to developed economies, mostly the US. However, their results are helpful for our research as a point of departure when choosing controlling variables and interpreting our regression results.

Our own results turned out in favor of the geography hypothesis in its sophisticated version (Acemoglu 2002, pp. 1260-1261). According to our results, the most probable candidate as an underlying force in the data-generating process was relative per capita extractive output. Our interpretation of the results is in terms of the Schumpeterian creative destruction in its spatial manifestation. The latter, according to Hounshell (1984), suggests that new products, technologies and organizational forms may move inputs to new territories where their reward would be even higher than in older territories. In our case this move is driven by new resource deposits. Old areas get poorer because their resources deposits get depleted, while new areas with still abundant resources get settled. This interpretation is in line with the general approach to Russian history proposed by Kluchevsky (1911), according to which, internal colonization is the most persistent pattern of dealing with various political and economic challenges on the part of Russian people and state. In fact, our evidence suggests that this pattern is at work until now.

The structure of this paper is as follows. The next section examines the hypotheses potentially explaining the negative link between age and wage in Russian cities. Section 3 describes the data sources and the most important patterns of the data. We go into detail about the data that contain the link under consideration. Section 4 contains a thorough examination of the spatial reversal of fortune. To confirm the link, we estimate a number of regressions of the wage on the age which differ from each other by the sample used, list of controls, and estimation techniques. In particular, we present the results of estimating regressions using the 2011 cross-section and several panels from 1991 to 2011; without and with regional and time fixed effects and spatial lags and various sets of controls depending on the available data; and finally, without and with censuring the sample to cities with populations over 12,000; finally, to better understand the nature of the link, we determine what part of the sample in terms of the age and wage contains the highest absolute values of the coefficients. Section 5 presents an exhaustive resource-related version of the sophisticated geography hypothesis which, according to our results, is the most relevant for explanation of the link. A model of the economy with several locations and extractive and manufacturing industries is presented with a numerical example to illustrate the main theoretical findings. Section 6 outlines the hypotheses based on the model and the ensuing empirical strategy to test them. Section 7 presents the results of testing the sophisticated geography hypothesis. Section 8 discusses alternative hypotheses in the context of the link. Section 9 concludes.

2. Hypotheses as to spatial income distribution

2.1. The increasing returns hypothesis

Our goal is to determine the mechanism underlying the negative link between age and wage in the context of a number of important hypotheses in the economics literature that explain activity and income distribution across space. One of these hypotheses is referred to here as the increasing returns hypothesis after Davis and Weinstein (2002). This hypothesis, originated in Krugman (1991a), has given a start for the New Economic Geography (for an introduction see, e.g., Combes et al. 2008; and Neary 2001). The main idea of the hypothesis is that the comparative benefits of staying in some location are a function of the economic activity there. The more intensive the economic activity is, the more attractive the location is for firms and labor compared with other locations. Economic activity in turn is intensified as population size grows.

One of the implications of this hypothesis is that a bigger location in terms of population size and/or density, other things being equal, is to be richer due to agglomeration forces at the location level. The latter can result from spatial clustering of varying kinds. There are the Marshallian external economies, and pecuniary economies; demand externalities related to multipurpose and comparative shopping; and longitudinal stability in economic communities, because of which new firms are attracted to places occupied by older firms even if the location is not an optimal one (Mulligan et al. 2012). Larger locations may benefit from labor pooling leading to a better match between workers and employers. For example, Gautier and Teulings (2009) explain the wage differentials between large cities and small ones based on the realization of scarce types of human capital. What differs large and small cities from each other is difference in trade-off between match quality and the cost of extended search among big and smalls cities. Big cities allow a tighter match between skills and jobs making wages higher. There is also a large literature about the centripetal forces related to the home market effect and market potential (Combes et al. 2008). In the urban economics literature, agglomeration effects are examined coupled with other forces behind city growth. Duranton and Puga (2014) mark out factors of city growth. Apart from agglomeration effects, they note transportation and housing, which are important ingredients of such a centrifugal factor as congestion costs, amenities, and technology (see also Duranton 2007). An example of the explanation of growth based on both the increasing returns and amenities is Partridge (2010).

Another implication of this hypothesis is that the spatial distribution of sizes across locations and thereby that of their incomes is historically circumscribed by a snowball effect, so that a location which by accident was initially larger is to grow faster and to become richer than a smaller one. Thus, cumulative causation underlies the spatial allocation of activities and incomes: more populated and well-off locations attract additional inputs, e.g., labor and firms, which makes them even more populated and rich. It implies that initial conditions and therefore history, matters for the allocational patterns in any period (Krugman 1991b).

This in turn suggests a positive link between age and size. As mentioned, the increasing returns hypothesis implies, inter alia, that an early start of a city favors its subsequent development compared with cities originated later (Davis and Weinstein 2002, p. 1270). The point is that the accumulation of inputs, in particular, population is a time-consuming process, owing to which locations with longer history have had more time to accumulate their population. Hence, demographic measures, such as population size and density are to be higher in older cities.

Thus, combining the increasing returns on location level[4] (Krugman 1991a) with cumulative causation gives rise to a kind of path dependency (Krugman 1991b). It suggests that older cities are to be richer. They were able to accumulate more inputs, and the latter as a result of their bringing together make the increasing returns, which, other things equal,

[4] There may be increasing returns on various levels – firm-level, city-level, etc. In our case increasing returns on city-level are of principal interest.

increases wages. Obviously, this logic hardly underlies the mechanism that produces the negative link between age and wage in Russian cities. The increasing returns hypothesis could be reconciled with the link, if one of the two parts of its logic did not work, e.g., if city age turned out to be inversely linked with the inputs measures. One could suppose that for some reasons inputs were accumulated in locations only over a limited period that was subsequently followed by moving of these inputs to other, maybe younger, areas. If so, there is to be observed a negative link between city age and their demographic and economic characteristics. So, one can admit that some element of the increasing returns hypothesis explains the link under consideration, but its overall logic does not fit it. This suggests that there are some forces behind spatial income distribution in modern Russia that differ from forces captured by the increasing returns hypothesis in its straightforward version.

2.2. The institutions hypothesis

Another hypothesis, termed here after Acemoglu et al. (2002) the institutions hypothesis, explains the spatial patterns of economic performance leading to unevenly distributed incomes across space by the varying quality of institutions in various locations. A number of works, in particular that by Acemoglu and his co-authors, empirically support this hypothesis. It is institutional change which according to their conclusions explains the reversal of fortune in the history of ex-colonies, namely, the fact that countries which were relatively rich in 1500 are relatively poor today. Using varying empirical strategies and related instruments for the quality of institutions they identified a positive impact of the quality of institutions on economic performance since the beginning modern era (Acemoglu et al. 2001, 2002). In these studies Europeans were treated as a potential source of better institutions for colonized territories. When it comes to the effect of Atlantic trade on economic performance in Europe itself, it depends on who played the most prominent role in that trade. In countries where this role was played mostly by merchant groups, for example in England, the Atlantic trade through the rise of the merchant class led to an improvement of institutions (Acemoglu et al. 2005). A recent supportive result for the hypothesis was obtained for 175 countries, rather than just ex-colonies, but over a shorter period 1960-2010. Using the advantages of panel data a positive effect of democracy on per capita GDP was established (Acemoglu et al. 2014).

Finally, institutions can be responsible for income inequality not only between countries, but also within them. As shown up in Acemoglu and Dell (2010) using data on the Americas, within-country differences in labor incomes are much more than the cross-country differences, and the within-country differences are ascribed to the differences in the efficiency of production determined by local institutions. Two points deserve attention in our context. First, income differentials across areas within countries are higher and in a sense more important. From the standpoint of an individual it may be immaterial what causes his/her well-being, and if regional differences matter more in this respect than the differences across countries, he/she will care more about which region he/she lives in, than about which country. The second point is that there are significant differences in the quality of institutions not only across countries, but also across regions within countries. Thus, substantial income differences between areas within countries may result from differences in local institutions.

The institutions hypothesis in its local version would explain the current Russian pattern of spatial income distribution if younger cities tended to have better local institutions. In other words, there is to be a negative link between city age and some measures of the quality of institutions. In the literature various measures of institutions quality are proposed. Among them there are property right protection, comparative share of transaction sector (Wallis and North 1986), and the availability of various organizational environments for private sector (Wittfogel 1957; Williamson 1985; North et al. 2009). Acemoglu and Dell (2010) measure the quality of local institutions by the disparities in access to paved roads. This is instructive for choosing urban characteristics to control for and for choosing a dependent variable when explicitly testing the institutions hypothesis. Using the data in our disposal for measuring the quality of institutions we used crime rate constructed as the ratio of total crimes to population size; the share of legal services expenditure in overall expenditures; and the share of firms in all organizations. These serve us as proxies for property right protection, the comparative share of the transaction sector, and the availability of various organizational environments for the private sector, respectively.

2.3. The simple geography hypothesis

Acemoglu et al. (2002) mark out simple and sophisticated versions of the geography hypothesis. The former version, also termed the locational fundamentals hypothesis in Davis and Weinstein (2002), suggests that the key factor in making spatial income distribution is inherent time-invariant geographical features of various locations that circumscribe their comparative economic potential. This hypothesis can be traced into history of economics, e.g., to Ricardo's law of diminishing returns from arable land (1817). The implication of this law is that at first people settle in good sites and they proceed to use inferior ones only as the better sites are occupied. This in turn means that early start of a location indicates its good geographical position compared to locations founded later. A comparatively good position is to remain a major factor of wage gaps because of the ensuing distribution of rents.

There is large body of evidence in favor of this hypothesis. For the most part, this literature indicates a persistent effect of time-constant geographical characteristics on economic prosperity. Famous evidence is presented by Davis and Weinstein (2002) in their study of Japanese urban history. The relative scarcity of good places for living in Japan produces a highly persistent city size distribution. It changed a little when the country underwent deep transformation of its political and economic regimes throughout its long history. And in modern times the city size distribution remains intact in the long run even after such shocks as the destruction due to the massive bombings during World War II. Other evidences about the persistence of city size distributions, including from natural experiments, like that in Davis and Weinstein (2002), are reviewed in Gabaix et al. (2004).

A major shock may have enduring impact on the relative performance if it changes the very geographical characteristics. For example, the American dust bowl was not fully offset by the recovery of the soil and, as a result, induced a relative decline of population sizes (Hornbeck 2012). Or, shocks may matter over the long term when they are an inherent feature of the territory, as with hurricanes in a number of the US coastal counties (Strobl 2011). Drawbacks of coastal territory are offset by some advantages, as is seen in positive link between coastal territory and density controlling for historical conditions. It suggests that this locational fundamental per se plays an important role (Rappaport and Sachs 2003). Similar evidence is presented in Dell et al. (2012) concerning temperature shocks. According to their results, an increase of temperature in poor countries is conducive to agricultural and industrial contractions as well as political instability. Thus, shocks can change the relative performance of various locations if they are a geographical characteristic themselves, or they change other geographical characteristics.

Finally, there exists the literature that explores the impact of time-constant characteristics coupled with other factors. For example, according to Vollrath (2011), agricultural endowments affect labor intensity and thereby the relative price of food. This, via varying fertility decisions, gives rise to various patterns of industrial development. Locational fundamentals also can determine the comparative initial conditions, starting from which the snowball effect leads to a massive accumulation of the population in one location and a relative decline in the other locations (Krugman 1991b). For example, Ayuda et al. (2010) point out that many European regions currently have population densities lower than that in mid-nineteenth century, and, as a whole, regions with lower than average densities continue to shrink, and vice versa. Their explanation is in line with Krugman's reasoning as to the initial conditions, namely, locational fundamentals heavily mattered until the 20th century, i.e., mostly in the preindustrial era, while subsequently the increasing returns became a more important factor. Another attempt to develop a hybrid hypothesis was presented in Wang and Wu (2011). They examined a linkage between natural amenities and increasing returns. The former attract workers which creates a large scale effect. The role of amenities is to encourage moving to some place initially, while subsequently moving is reinforced by the increasing returns.

According to the overall logic of the simple geography hypothesis, the major force behind comparative economic prosperity is the differences in natural conditions; and since these conditions are time-constant, there is an unchangeable spatial structure with respect to inputs, activities, and incomes, whatever temporary shocks. Concerning the link between age and wage, the simple geography hypothesis as such cannot explain it, because its key prediction of the unchangeable spatial distribution is violated. It could be supposed that territories with advantageous local fundamentals were unknown over history, but treating the relative wellbeing in this way means that the value of resources is considered dependent on human activity.

2.4. The sophisticated geography hypothesis

Like the simple geography hypothesis, the sophisticated hypothesis posits that geographical characteristics are a major force in determining spatial income distribution, but with a "focus on time-varying effects of geography" (Acemoglu et al. 2002, p. 1233) meaning that presence and importance of geographical characteristics change with time. So, the key difference from the simple version is treating important geographical features as time-varying. The key implication is that some geographical features may change their relative importance for human activities depending on changing economic conditions such as technological progress or resource constraints. Spatial patterns of economic performance are explained by geographical factors circumscribed by human activity. This pattern admits the possibility that changing spatial allocations of activities and the ensuing wage differentials result from the changing relative importance of various geography characteristics as well as their exposure to human activity.

Among scholars who treated important geographical features in this way were Braudel (1967), Wallerstein (1974), and Wittfogel (1957). More recently, Krugman admitted their existence in his distinction between first and second natures, the latter being time-varying geographical features (1993). One of the examples of using this approach referred to in Acemoglu et al. (2002), which is relevant for our purposes, is related to coal or other resource reserves that enabled industrialization. In this case it is evident that the importance of some resources or other geographical features depends on technological progress. These natural resources might be negligible in one historical period and be of value in other times as was the case with coal as well as oil, gas, and other energy resources. Apart from technology, what may heavily matter over shorter historical periods for changing relative advantages of various sites is an exhaustible nature of many valuable resources. Thus, while some territories may decline because their resources are not as in demand as they were previously, other territories may suffer from depletion of their resources deposits.

In a sense, this pattern can be treated as a special case of the Schumpeterian creative destruction when applied to spatial input allocation and its rewards (Schumpeter 1942; Hounshell 1984; Florida, 1996). As was notoriously exemplified by Detroit, some once rich locations can decline as a result of fall of relative importance of some products, technologies, and organizational forms, or the rise of new alternatives (Glaeser 2011). The decline of some locations due to spatial creative destruction may lead to the rise of other locations as a result of inputs moving there from older locations. If the economic prosperity of some territories result from their resource abundance, the destructive move of inputs can be driven by new resources and resource deposits. A logical sequence would be: new technologies – new resources – new deposits – new rich regions.

Time-varying resource abundance may lead not only to economic prosperity, but also to stagnation or even decline, as stressed by the resource curse theory. This theory addresses a number of political economic issues related to resource rents (e.g., Ross 2001; Egorov et al. 2009). The latter induce mafia-like groups to rule in resource-abundant areas because a key resource one needs to get the resource rents is a brute force. The ensuing social organization in turn induces a further redistribution which even stronger subdues any productive work. Countries and their areas differ from each other in their resource abundance, the share of the extractive output in overall productive activities, and the share that the population has in the rents. Depending on these factors, various countries suffer or benefit differently from their resource stocks.

There is also an explanation of the resource curse related to the crowding effect. This makes the resource abundance result in higher prices on non-traded goods which reduces the profit margins of manufacturing firms and undercuts export possibilities except for extractive industries (Sachs and Warner 2001). Resource abundance hinders the export of manufactured goods, though it induces the export of extractive produce. This fits the "new new" trade theory based on Melitz (2003), according to which, to export its produce, a firm needs to have relatively efficient production, otherwise the firm will not be able to cover the trading cost imposed by moving its goods abroad. Empirical evidence indicates that trade in mining goods is driven by factor endowments, but not by productivity gains as the "new new" trade theory suggests (Davis and Cordano 2013). The resource curse may be seen not only across countries, but also across areas within countries. James and Aadland (2011) ran their study on the US county level, which allowed them to control for many country level features. Their results supported the resource curse on the county level, being an explanation again development at the expense of manufacturing, which is more conducive to growth via increasing returns. Apart from this, resource abundance is positively correlated with inclination to autarky, which is obviously related to the failure to export manufacturing goods, and with social unrest.

To sum up, time-varying geographical features change with time both in their importance and their presence; this may lead to a decline of once prosperous territories and the rise of new settled ones; when it comes to natural resources, this time-varying feature may result in both a blessing and curse. To what extent can the overall economic performance of Russia or the comparative prosperity across its areas be explained by these factors? There were several empirical studies of the impact of Russia's resource abundance. They indicate that Russia benefits more than suffers from resources. Alexeev and Conrad (2009) point out that per capita income in Russia would be lower if Russia lacked extractive industries. At the same time, over the decade of 2000s rapid growth induced by the booming extractive output even was not accompanied by the crowding effect, and, in fact, over the same period manufacturing production did increase (Dobrynskaya and Turkisch 2009).

The sophisticated geography hypothesis would fit the established pattern if there were a negative link between city age and time-varying geographical features leading to relative economic prosperity. In case of Russia, an obvious candidate for such a time-varying feature is valuable resource deposits. Within a long historical horizon, the economic significance of resource endowments depends on the stage of technological development, while over shorter periods the importance of a place may grow sharply owing to newly discovered resource deposits or may fall as the resources are exhausted. If the extractive production is the key factor of comparative economic prosperity in modern Russia, the negative link between age and wage in the short run can be explained by resource exhaustion in some locations and discovering new resource deposits in other locations. Less available resources could lead to an increase in extraction costs or even make the respective extractive firms stop their work, while settling in new sites may be beneficial owing to their not yet exhausted deposits. This process would bring about the decline of older locations and rise of new ones. In the long run the same process could be driven by technological progress changing the relative demand for various resources.

3. Data

3.1. Data sources

There are numerous definitions of a city. For the most part, they refer to either some threshold values of population size or density, or to an administrative status (Rozenfeld et al. 2003, Henderson 2005). The Encyclopedia of Russian cities gives the official modern definition of a city in Russia. A city is defined as a location with population above 12,000 and the share of non-agricultural population in labor force no less than 85%. These are criteria from which the modern Russian state starts when giving city status. However, there are a lot of locations that do not meet these requirements still having city status. In 1989 among 1037 cities, 160 cities had less than 12,000 people, and there were 237 locations with more than 12,000 that did not have city status (Lappo 1998, p. 5). As of 2011 in our dataset (Economica gorodov Rossii 2013) among 1056 cities with data on their population, 215 had population size less than 12,000.

Some cities have become under-populated as a result of the population decrease which a number of Russian regions faced during the later Soviet and early post-Soviet periods. However, most of these cities saved their city status due to their economic and/or administrative functions or other considerations on the part of the government. Some locations even got city status while not meeting the population size requirements. As a rule this is explained by their urban functions they do though not having enough population. The most exciting example of such a new born city is Magas that was given city status immediately after its foundation when only 100 people resided there, because Magas was designed to be the capital city of Ingushetiya, and in minds of the decision makers its administrative functions outweighed its population size when it came to giving it city status.

The main body of our data is from the commercially distributed base “Multistate” where Rosstat publishes data on Russian cities “Ekonomika gorodov Rossii” (2013). The dataset contains data on all Russian locations that have an official city status. The sample size is 1097, and the time span is 1970-2011. The actual sample size varies depending on variables used in the regression analysis. Except for our main interest variables, we use an extensive set of control variables, some of which are used also as dependent and/or interest variables for testing the hypotheses. We also use data on geographical coordinates (in decimal degrees) that we taken from Bariev (2007) for 992 observations; the remaining 97 items were taken from Internet maps. Thus, the subsample of the cities with the available coordinates is 1089. This is almost as many as the subsample of the cities, 1090, with date available on wage for at least one of the 21 years of the period 1991-2011.

3.2. City age

The spatial and urban economics literature lacks a consensus concerning a definition of city age. Glaeser and Kahn refer to definitions such as “years since incorporation of the largest city” and “the first year it hit some population mark” (2001, p. 22). A number of the definitions are given in Lappo (2006) and Giesen and Suedekum (2012). The problem with many of the definitions is that the respective dates are mostly unknown (p. 14). These dates, and those used by Giesen and Suedekum, are not available in our dataset. As a rule, what is saved in recorded sources is information about more important events than just about the founding of the first building or something like that. There is also a logical problem with this definition. All the locations of modern Russia (like any other country) started from something like the first building, but only a tiny part of them became cities. They either disappeared completely as locations or, more often, existed on as villages that never came to such a specialization, or an administrative position, or population size that would make the state give them a city status.

Given the available sources we rely on two alternative definitions of city age. The first one is related to creating something that gave rise to the development of a location as a city; the other is related to getting city status from the state. We have information about age of Russian cities from Economika gorodov Rossii (2013) as well as from three other sources. One of them is Administrativno-territorialnoye delenie (1987). It is the latest official edition with information about all Soviet, including Russian, locations with their foundation and/or city status years. Other source is Lappo (1998) where one can find a brief history of almost all the Russian cities that are in our main dataset (Economika gorodov Rossii 2013). Finally, there are 14 cities for which the three above-mentioned sources lack their foundation or city status years in which cases we used dates from their official Internet sites.

Our main dataset contains a column named “foundation year” for 1053 cities. For 28 cities this column has an additional year in parentheses in which cases the earlier date refers to the actual foundation of the city. This can be established when comparing these dates with the detailed review on each city in Lappo (1998). The later date is either the year of getting city status by the respective location or the year of regaining this status after a temporary loss. With three exceptions, all these double dates are for Leningrad Oblast. Among 31 cities in this region 25 have the two dates.

There are two more regions of note. One is Perm Krai whose 25 cities lack years of their start in Economika gorodov Rossii (2013). The other region is Kaliningrad Oblast – a part of Eastern Prussia that became a part of Russian Federation after World War II. For the 22 Prussian cities of Kaliningrad Oblast, such as Königsberg or Gumbinnen (the Soviet names are Kaliningrad and Gusev, respectively), there are two dates – that of the foundation and that of the Soviet renaming. In these cases we used the dates of the foundation, as their renaming after World War II did not fundamentally change their nature as cities.

For 28 of 1053 observations two possible dates are available when using Economika gorodov Rossii (2013). However, for the remaining 1027 observations they have values that refer to a foundation year in some cases, and to a getting city status, or even later dates in other cases. This fact, again, can be established from other sources about Russian cities, specifically, from Lappo (1998) and Administrativno-territorialnoye delenie (1987). Thus, the dataset mixes incomparable

Table 1. Summary statistics of the age variables for the year 2011

Variables	obs	mean	median	sd	min	max
City age (since a city origin)	1,071	288,2	206	277,9	11	2,561
City age (since a city status)	1,071	181,1	75	220,1	6	1,444
City age (multistat with some earlier dates)	1,053	199,4	79	240	7	1,572
City age (multistat with some later dates)	1,053	194,9	76	238,3	7	1,572

values under the same variable name.

To get a more accurate age variables, we used the alternative sources. Based on them we constructed two other age variables. '1st age variable' refers to the foundation year; '2nd age variable' refers to the year of getting city status. The sample size of cities with available values of these constructed age variables is 1071. Thus, we made more accurate age variables and got 18 more age observations. The latter addition was from cities of Perm Krai. At the same time, for 7 cities from Economika gorodov Rossii (2013) data about their current average wage are absent because of their loss of city status or for other reasons. So, we have four variables that refer to the city beginning – two from the dataset on Russian cities and two constructed from additional sources. Our age variables were generated as $2011 - founded$, where the first term is the latest year for which the data are available, and $founded$ is a date when a city was founded or was given city status.[5]

We used all the age variables to check the robustness of our results. At the same time, we used the constructed age variables as our main age variables. Our considerations here are that these variables are more consistent when dealing with the start and therefore able to give a more accurate picture; and the sample does not exclude the whole region, as it is the case for the start variables in Economika gorodov Rossii (2013), so that it is more representative. Summary statistics for the four age variables are presented in Table 1.

Means, and min and max values for the variables relate to each other in a predictive way. The first age variable, based on foundation year, has the highest mean, min and max values, and the second has the lowest. Three Russian cities on the Black sea were founded as colonies of ancient Greek city-states which makes their age well over two thousand years. The max value 1444 of the second age variable is the former capital city of Dagestan, Derbent, which was a city from the beginning of the Middle Ages. Note also that variation coefficient is markedly lower for the first age variable than for the other variables; it does not exceed unity. It is because this variable is somewhat more homogeneous by its values, not so dispersed among adolescent and very old ages as the other age variables.

As an estimate of the density distribution we present here the quantile plot, depicted in Figure 3, rather than histogram so as to better show the exponential urban growth. The distribution quantiles for the age in conjunction with range of this variable reveal a distinctive pattern of Russian urban history, viz., a rapid speed-up of the number of cities in the modern era. Even without cities founded before the origin of the initial Russian state – Kievan Rus – during about the seven hundred years of pre-Moscow Russian history only ten percent of Russian cities still existing occurred versus the 90 percent that occurred during the modern era.

The median values of all but the first age variables indicate that more than half of existing Russian cities were founded and/or given city status after initiation of the Soviet industrialization under Stalin in 1930. Also, if we take the second age variable we can see that between 1972 and 2011 as many cities were created as in the millennium of the Middle Ages 567-1552; or, for the first variable the same number of cities are for about two millennia versus the period after the war 1949-2011.

[5]Using Lappo (2006), we treated following events as a beginning: the first mention (direct or indirect) in the historical records; the building of a military establishment such as a fortress; the emergence of an economic establishment related to a typically urban specialization, a settlement with urban specialization, a transport center, an administrative center, a custom, a health resort; finally, changing place given a time gap between stopping life in the previous place and its beginning in the new place. As for getting city status, it is straightforward when it comes to the relatively young cities that were officially given their status, but it is not so clear for ancient cities. For the most part, in the dataset their foundation dates are treated as those of giving them city status on the consideration that they were never known as non-urban locations. Following this practice for ancient cities for which the year of being given city status is unknown we treat as those their foundation years. In fact, the share of such cities is 120. Some of these were founded as cities, so that they got city status from the outset. Others have the same dates due to their great ages. Famous examples of both types are Saint-Petersburg and Moscow, respectively. The exclusion was made for the three most ancient cities by their foundation year. They existed under various regimes, and at times they stopped their existence for long, so that their resuming under the Russian state was treated as the new beginning.

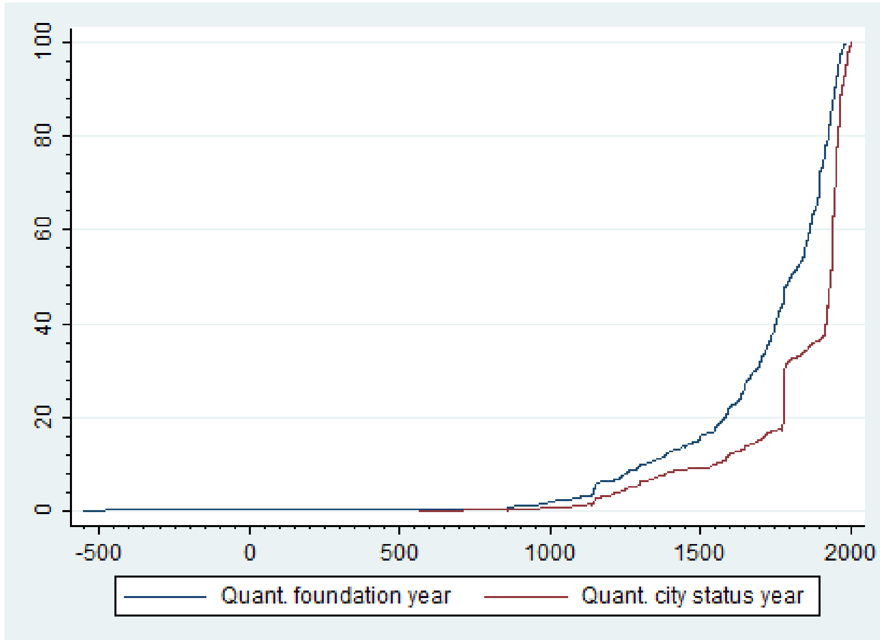

Figure 3. Quantile of a foundation year versus year of a city's foundation

Figure 3 also highlights two periods of extremely rapid urban growth. One occurs under Catherine the Great and was the result of administrative reform. This period roughly coincides with the beginning of the Industrial Revolution in the Great Britain, i.e., approximately 1775-1795. The other period begins after the Russian revolutions in 1917 and speeds up after the beginning of the first Soviet five-year plan in 1930. Such rapid urban growth is still occurring, though in the post-Communist period it has slowed down. Thus, urban growth in Russian history, as a whole, followed a pattern similar to that for many developed countries: slow growth after the rise of the state; its continuous acceleration in the early modern times; and its extreme speed-up after the industrialization, which in Russia actually launched no less than twice.

One of the implications of this growth history is the density of city age in Russia. The late growth in the number of cities results in more than half of the cities being of Soviet origin, and the bulk of the cities are just several decades years old. Though there are a number of very old cities, the vast majority of the cities are of quite recent origin, and it is these young cities that are a decisive factor in making Russian spatial income disparities.

3.3. Wage

Another dataset published by the Rosstat is Regiony Russii (Regions of Russia) (2012). This dataset is in many respects richer than Economika gorodov Rossii (2013), as it contains various data on 83 Russian regions that are absent in dataset on the cities. It contains both nominal wage and the price of the consumer basket as well as per capita income, and it can therefore be helpful when checking the relevance of some variables as proxies and for robustness checks.

Our main dependent variable is a city's average nominal wage. We use this variable as a proxy for per capita income, since the dataset on Russian cities lacks another proxy. Neither per capita income, nor real wage, nor the price of the basket are available in Economika gorodov Rossii (2013). According to Regiony Russii (2012), wages in 2011 comprised only 40.1% of all the incomes (p. 176). At the same time, given rapid inflation real wages would be a more reliable proxy for per capita income. This poses the question of how good the nominal wage is as a proxy for per capita income.

To evaluate it, we examined the respective variables in the dataset on Russian regions. Dividing the nominal wage by the basket price we got a real wage. On these regional data Pearson and Spearman correlations between nominal and real wages in 2011 are 0.92 and 0.94 for all 83 regions; and between nominal wage and per capita income in 2010 the

Table 2. Summary statistics of wage for several years

Years	obs	mean	sd	min	max
1991	944	0.572	0.276	0.236	6.760
1998	1,056	990.5	601.4	300	5,359
2001	1,054	2,888	1,979	844	17,012
2008	1,057	14,565	7,133	4,915	80,299
2011	1,054	19,867	9,096	7,957	72,551

same correlations are equal 0.88 and 0.89 for 80 regions excluding three subregions of Tumen and Archangelsk regions for which GRP is absent in the dataset. These values basically mean that even if wages comprise only a smaller part of all incomes it is quite representative in that the relationship between wages and other incomes is approximately the same across the regions.

Using nominal wages rather than real ones is not to affect substantially our results. The matter of interest for us is cross-city income differences, and cross-city differences in the price index do not offset the nominal incomes gaps. It can be seen if one checks the coefficients of variation and the respective correlations in the dataset on Russian regions. The coefficients of variation for the average wage and price index in 2010 are 0.46 and 0.20, respectively, and Pierson's and Spearman's correlation coefficients of nominal wage and real wage for 2010 are 0.92 and 0.94, respectively. What we get when shifting from nominal wages to real ones is a lowering of their variation. The coefficient of variation for the real wage is 0.25 instead of 0.46 for the nominal wage, but the spatial distribution remains basically the same. Dynamic regressions, in which inflation can be a major compounding factor, can be controlled for by time fixed effects dummies. Hence, we can treat nominal wages as a good proxy for per capita regional income in Russia.

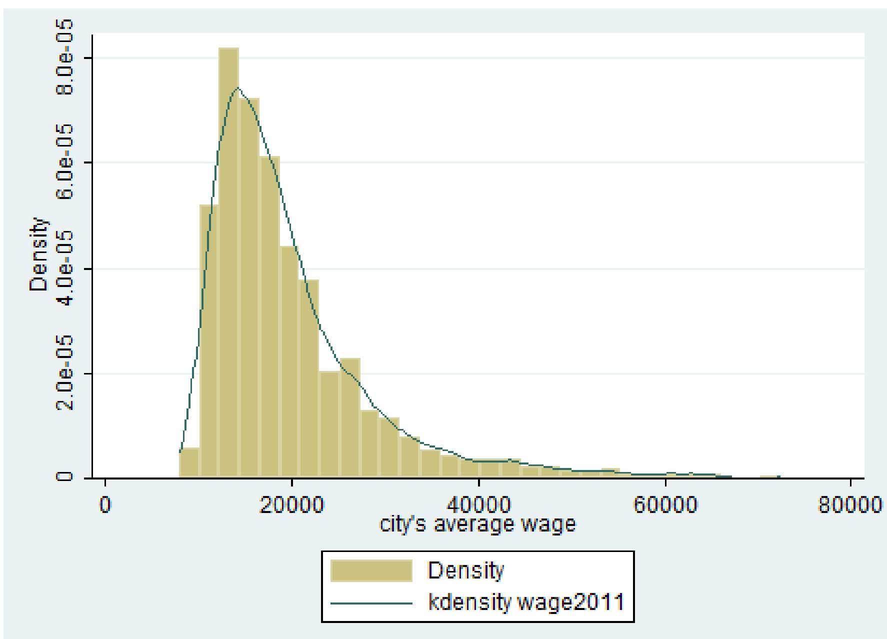

Figure 4. Kernel density of wage, 2011

Summary statistics for the wage are presented in Table 2. Mean, min and max values for the subsequent years reveal a strong upward trend related to the hyperinflation of 1990s and the economic recovery in 2000s. One can note that wage variation was highest for 1998 and 2001, toward the end of the recession period. It implies that income gaps and overall GDP were negatively correlated, and this is in line with the empirical evidence for other countries. At the same time, wage variation is much lower than for age.

The histogram in Figure 4 reveals a heavy right tail in distributions of wage making it similar to that of the age. As a whole, it means that though there are relatively old and rich cities, the most typical ages and wages are relatively low. As mentioned, during the post-Soviet era Russian cities tended to diverge in their relative per capita incomes, in which Russia seems to replicate relative income dynamics other countries displayed at early stages of their industrial market-oriented development (Kuznets 1955; Ayuda et al. 2010).

Other variables with their names and panel length which we used as controls or as dependent variables in specifications designed to test alternative hypotheses, are listed in Table 3.

Table 3. Variables definition and available years in the dataset

Variables	Years
Geographical controls	
Latitude divided by 90	
Longitude divided by 180	
Log (1+distance to a nearest railroad station in km.)	2011
Log (1+distance to a nearest dock in km.)	2011
Social and demographical controls	
Log population size	1991-2011
Population density	1991-2011
Net migration to population ratio	1991-2011
Labor to population ratio	1991-2011
Students to population ratio	1991-2011
Doctors to population	1991-2011
Economic controls	
Unemployment rate	2001-2011
Firms to all organizations ratio	1997-2011
Big firms to all firms ratio	1997-2004
Discharge from stationary sources to area ratio	1991-2011
Industrial controls	
Manufacturing output to labor ratio	2005-2011
Production and distribution of electricity, gas, and water to labor ratio	2005-2011
Extractive output to labor ratio	2005-2011
Fuel output	1991-2004

4. Spatial reversal of fortune

4.1. Specification and estimation

This section presents the link that is the main starting point in our discussion of the hypotheses as to spatial patterns of economic activities and results. This link is a result obtained when estimating the following mean regression:

$$\log wage_{it} = \log age_{it}\alpha_1 + \mathbf{X}_{it}\boldsymbol{\alpha}_2 + \rho_i + \varrho_t + \varepsilon_{it} \tag{1}$$

where $\log wage_{it}$ is the logarithm of a city's average wage, $\log age_{it}$ is the logarithm of a city's age which is our interest variable, $\mathbf{X}_{it}$ denotes the vector of an intercept and controlling variables (for their list see Table 3), ρ_i is the ith city fixed effect, ϱ_t is the tth year fixed effect and ε_{it} is an error term. Our main interest here lies in the relationship between wage and age, therefore in all the specifications considered in this section log wage is used as a dependent variable and log age and the intercept are the right side variables.

We estimate the interest link in logarithms. The reasons for this are related to the variable distributions as they are presented in Figures 3 and 4. Both age and wage are skewedly distributed to the right. Their transformation to log form makes their distribution much more like a normal distribution. It is quite possible that transformation makes the regression better fit the data. Goodness of fit, when it makes sense to compare R-squareds, is better for log-log model versus log-linear model and for linear-log model versus linear one. One more benefit is that the log specification is more convenient for interpretation. Units measuring city age and wage, years and still more rubles, are too little to affect anything. At the same time, because of rapid inflation the ruble radically changed its value. For the age an additional year as a rule matters much more for younger cities. So, the respective slope coefficients as age elasticity of wage rather than just marginal effects make much more sense.

As for other variables they serve as controls and are used in different combinations from one specification to another. In most of the specifications the control variables include a full set of regional dummies.[6] One of the considerations in favor of using regional dummies is that the link can be compounded by so called "northern bonuses". The latter are special rewards of higher wages for living and working in the severe conditions of the extreme North, which are inherited from the Soviet times. In such regions wage is multiplied by the coefficient in a range of 1.15 and 2 depending on the region.[7]

An immediate inclusion of these bonuses is hardly possible as they are, in fact, used in almost all the regions and are very numerous and different from one profession to another. So, what can be taken into account is that regions differ from one to another with respect to various conditions of life such as weather, pollutions, or the institutions quality which are reflected in the ultimate wage differentials across regions (Rappaport and Sachs 2003), including those resulting from the compensating bonuses. In our regressions this fact is captured by the inclusion of the full set of regional dummies. As is seen in Figure 1, the link between age and wage has a major regional component. If this component is the only one, the link will vanish after controlling the regional assignments.

Another important control variable we use is the dependent variable spatial lag. Like its time counterpart, the dependent variable spatial lag may correlate with a number of unobservable characteristics of a location, so that its inclusion would allow one, at least in part, to control for them. Also this control variable allows one to control for the spatial dependency and spatial heterogeneity in the data (Anselin 1988), the existence of which is quite possible. Again, note the important regional component in our link. Figure 1 shows a regional grouping of cities by both their age and wage. A similar grouping by neighborhood rather than only a regional assignment is possible as well. The exact link between age and wage is different in different parts of Russia. For example, Kursk Oblast has a strong negative link between them (Figure 2), while there are regions with no link or a positive one. All these potentially compounding factors are to some extent controlled for by our dependent variable spatial lag.

As a spatial weights matrix we use a row-normalized matrix of the five nearest neighbors. Our choice in favor of this kind of spatial weights matrix rather than its most widespread alternative – (a function of) inverse distance matrix – is related to Russia's size and that it is very sparsely and heterogeneously populated in terms of both people and cities.

The minimal geographical distance between nearest neighboring cities is less than 3 km., and these cities are Korolyov and Yubileyny in Moscow Oblast, which is not surprising for this is the most densely populated region in terms of cities in Russia. At the same time, the maximum distance from a city to its nearest neighboring city is 644.5 km; it is a distance between Verkhoyansk – the coldest city in Russia (see, e.g., Pipes 1973) – and capital of Sakha Republic, Yakutsk. Another similar example is the distance from the administrative center of Chukotka Autonomous Okrug, Anadyr, to its nearest neighboring city Bilibino, 616.3 km.

As a whole, this characteristic of the spatial allocation of the cities varies greatly. The coefficient of variation is 1.32, which reflects very different mutual allocations of the cities in various segments of Russian territory. In radius of 644.5 km. from Yubileyny there are 433 cities in Russia versus no any such city in the same radius of Verkhoyansk, so that 644.5 km is less significant distance in sparsely populated areas than for more densely populated areas. And the more so when it comes to distance between very distant cities as those in western and far eastern parts of the country; the maximum geographical distance between any pair of the cities is 7894.7 km.

Given this characteristic of the cities allocation, we have chosen the nearest neighbors spatial weights matrix. The rows of this matrix contain unities for cities that are among five nearest ones to a row city, and zeros for other cities. The main diagonal contains zeros. Also this matrix was normalized to unity to make the results for the dependent variable spatial lag within range $[-1, 1]$ and thereby make it easier to interpret them. Thus, the matrix is comprised of rows each of which contains five 0.2s for the nearest neighbors and the remaining zeros.

[6]As of 2011 among the 80 Russian regions in Regiony Rossii (2012) three regions included autonomous districts that we treated as distinct regions. Thus, we used 82 regional dummies using one of the regions as a reference one. There are several regions that contain only one city. For example, Moscow and Saint-Petersburg are cities that have the status of distinct regions. Therefore the actual number of dummies used in the regression analysis was slightly fewer.

[7]Federal law No 4520-1, 19.02.93 "About the state warranties and compensations for those living in areas of the extreme North and areas that are equivalent to them".

Finally, we use a number of geographical, demographic, social, environmental, economic and industrial controls (see Table 3). All these factors potentially underlie the link we explore here. For example, the list of geographical controls includes latitude and longitude, both normalized to unity concerning the scale of the respective hemisphere. These variables, along with the respective regional assignments, allow us to control for the "northern bonuses". Other controls capture important covariates of relative prosperity such as unemployment rate, population size and composition, industrial concentration.

To check the robustness of the established link, we have obtained estimates for various specifications different from each other by the value of T, combination of control variables, the age variables used, the censuring of the sample for population size, and estimation procedures. We estimated regressions with $T = 1, 7, 14$, and 21 panels depending on the available data by years and the included controls. Then, allowing for the official definition of a city as a location with population over 12,000 we estimated the regressions on a subsample of cities that meet this criterion.

We chose estimation procedures depending on the specification. The simplest cross-section version was estimated by OLS with the robust standard errors obtained by the sandwich estimator. For a cross-section specification with the regional dummies, robust clustered standard errors were obtained to account for the regional dependence of observations. Finally, cross-section specifications with the dependent variable spatial lag were estimated by 2SLS to address the endogeneity issue. As excluded instruments, spatial lags of all the exogenous regressors were used (Kelejian and Prucha 1998). Finally the GLS three step estimator was used (Kelejian and Prucha 1998) to allow for spatially dependent errors, and the endogenous spatial lag of the dependent variable. The three steps are obtaining initial estimates by standard 2SLS, obtaining the slope coefficient of the residual term spatial lag, and once more obtaining 2SLS estimates corrected by the Cochrain-Orcutt transformation.

An alternative estimator in this case is the ML estimator. It is an effective estimator and given a distributional assumption it is also a consistent one. We did not use the ML, because it is prohibitively complicated computationally. Also the strict assumption of error normality distribution it imposes may result in inconsistent estimates if it actually does not hold. Neither of these drawbacks are there when using the 2SLS estimator. Unlike the ML, the latter is consistent estimator, whether the distributional assumption holds or not, and quite feasible computationally.

When estimating regressions with $T > 1$, i.e., on panel data, we face the complication that the interest variable is time-invariant, which prevents us using a consistent fixed effects estimator. The standard solution to this problem is the Hausman-Taylor estimator. The problem with that is the choice of truly exogenous time-invariant and time-varying variables when it comes to spatial data. An obvious candidate as an exogenous time-invariant variable is the age variable. However, it potentially correlates with any other spatial characteristic and thereby is an invalid instrument. The same can be said about potential exogenous time-varying variables. To avoid the difficult choice between exogenous and endogenous covariates, we used GLS specifications with as many controls as possible. First, we estimated panel regressions with a full set of regional and time fixed effects. Second, we included the dependent variable spatial lag, which, as mentioned, enables us to capture many unobservable spatial characteristics. The five groups of control variables were included as well.

Again, the inclusion of the dependent variable spatial lag poses an endogeneity issue that is addressed by the 2SLS estimator. And once more we estimated the panel regression with the dependent variable spatial lag instrumented by the full set of lagged exogenous variables. As for standard errors we allowed for time-wise autocorrelation by means of the standard GLS.[8]

4.2. The link in cross-section regressions

We began with estimating a number of cross-section regressions differing from each other by the number of included controls, the ensuing estimation methods, and the cross-section used. In Table 3 we report the results for the 2011

[8]The same regression was estimated also by the GM estimator proposed in Kapoor et al. (2007) that allows for time-wise as well as spatial autocorrelated errors, but its using is a cumbersome task when one tries to include regional fixed effects. Without the latter we used both estimators, and the results as for the age variables are basically the same, if not mention much higher statistical significance of estimates obtained by the latter estimator.

cross-section. Our initial cross-section specification for 2011 just captures the correlation between age and wage with the estimation results in the first column. Here we have a highly significant negative link between age and wage. The specification in logarithms implies a slope coefficient as roughly an age elasticity of the wage. A one percent increase in age is consistent with a 0.14% fall of average wage. Given that the mean wage for 2011 is 19,866.5 rubles, at this mean point a one percent increase of city age results in fall in predicted wage of 28.4 rubles. As to bigger age differences, for example, Vysokovsk, aged 128, compared to Salavat, aged 63, is to have wage as a share of the latter city's equal to $\exp(-0.143 \log(128/63)) \approx 0.9$. In other words, a twofold difference in age implies a 10% difference in the average wage in favor of younger city. And this basically equals to actual difference in their 2011 wage, namely $19,573$ rubles$/21,682$ rubles ≈ 0.9. Given the range of the first age variable after dropping outliers is about 600 years, not to mention the range for the whole sample 2550 years (Table 2), it implies substantial income gaps between cities predicted by their age differences.

The second column contains the results after inclusion of the regional dummies. The absolute value of the slope coefficient for the age variable is much lower now, but the negative link is still highly significant. This result confirms our initial observation when examining the maps in Figures 1 and 2. Though the negative link between the age and wage is for the most part related to inter-regional differences, this link exists within regions as well. A similar control is the dependent variable spatial lag, which is added in the specification presented in the third column. There is a clear spatial pattern of per capita incomes and therefore wages, and one might suppose that high wages in many young cities are explained by their proximity to big and rich cities. This is the case for Moscow Oblast where around Moscow there is a constellation of new cities that are relatively well off due to their position as near neighbors of the capital. There are similar allocations of subordinate cities around other metropolises as well. However, this effect explains the link only in part, if it does at all: the slope coefficient remains negative and highly significant after the inclusion of the spatial lag.

The same can be said about various controlling variables. Our geographical controls including normalized latitude and longitude as well as log distances to railroads and docks may explain some variation of wage and, again, be responsible for the link. As mentioned, geographical latitude and longitude may be the basis for paying the "northern bonuses", and at the same time some cities might be founded relatively recently in areas with severe natural conditions. Proximity to railroads and docks also may correlate with both the age and wage, and thereby explain some part of the link. In columns 3-9 we can see that among the four geographical controls only one, latitude, matters for urban wages. The latter are higher in the northern direction, which is most probably related to the "northern bonuses".

Other controls are social and demographical variables, which include log population size, population density, and four relative variables, i.e., those divided by population size: net migration, labor force, numbers of students and doctors. All these controls are obvious covariates of both the age and wage. Population size and density, other things being equal, are higher in older cities (Davis and Weinstein 2002; Giesenand and Suedekum 2012) and they are to correlate with wage (Combes et al. 2008). The other variables measure various aspects of population dynamics and composition. Net migration may reflect the relative attractiveness of a location, which can include its comparative economic position. The shares of able-bodied people, students, and doctors in the population characterize composition of population with respect to its labor force, human capital (proxied by its prospective owners), and social welfare.

Particularly, the relative labor force is an important control. On the one hand, it is to negatively correlate with age, because, other things being equal, in older cities there are to be more pensioners. On the other hand, retired aged people often still work. Since their additional earnings are usually quite small, their proportion is to have a downward effect on the average wage. Thus, the relative labor force may explain the negative link between the age and wage. Log population size and the relative labor force have a significantly positive effect on wage in all the specifications. This is in line with the increasing returns hypothesis and indicates the important role of the relative quantities of able-bodied people and thereby burden population, viz., elderly and child population for per capita income. The only control that is insignificant in all the specification is the relative number of students. The remaining are significant in some of the specifications. It is obviously explained by these variables closely correlating with other measures of large agglomeration.

Table 4. Base-line regression of log wage 2011

	1st age var					2nd age var		3rd age var	4th age var
	Base-line regression	+ reg FE	+ main controls	+ industrial controls	+ rest. to >12000	all controls	+ rest. to >12000	all controls	all controls
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Age									
1st age var	-0.1427***	-0.0366***	-0.0539***	-0.0451***	-0.0396***				
	[0.0135]	[0.0115]	[0.0074]	[0.0073]	[0.0077]				
2nd age var						-0.0288***	-0.0296***		
						[0.0066]	[0.0073]		
3rd age var								-0.0270***	
								[0.0063]	
4th age var									-0.0274***
									[0.0064]
Geographical controls									
Latitude			1.8179***	1.6987***	1.8806***	1.7108***	1.8301***	1.8179***	1.8216***
norm.			[0.3524]	[0.3650]	[0.4348]	[0.3778]	[0.4479]	[0.3876]	[0.3875]
Longitude			-0.1761	-0.1560	-0.1867	-0.1536	-0.1880	-0.1578	-0.1477
norm.			[0.1487]	[0.1394]	[0.1666]	[0.1450]	[0.1726]	[0.1501]	[0.1502]
Distance to			-0.0002	-0.0015	-0.0033	-0.0013	-0.0035	-0.0020	-0.0020
the nearest station			[0.0032]	[0.0032]	[0.0038]	[0.0033]	[0.0038]	[0.0033]	[0.0033]
Distance to			0.0003	0.0022	0.0056*	0.0026	0.0055*	0.0024	0.0023
the nearest dock			[0.0032]	[0.0030]	[0.0032]	[0.0030]	[0.0033]	[0.0031]	[0.0031]
Demographic and social controls									
Log population size			0.0577***	0.0512***	0.0327***	0.0534***	0.0362***	0.0537***	0.0541***
			[0.0079]	[0.0081]	[0.0104]	[0.0083]	[0.0106]	[0.0084]	[0.0084]
Density			0.0001	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003
			[0.0006]	[0.0006]	[0.0006]	[0.0006]	[0.0006]	[0.0006]	[0.0006]
Relative net			-0.0591	0.5958	1.1064**	0.5292	0.9298*	0.5292	0.5219
migration			[0.4514]	[0.4425]	[0.5389]	[0.4475]	[0.5429]	[0.4578]	[0.4576]
% of labor			1.4941***	1.2675***	0.9487***	1.3314***	1.0074***	1.3010***	1.3029***
			[0.2270]	[0.2280]	[0.2685]	[0.2314]	[0.2715]	[0.2359]	[0.2358]
% of students			-0.1730	-0.0158	0.3965	-0.0546	0.3973	-0.1546	-0.1595
			[0.2663]	[0.2509]	[0.3254]	[0.2560]	[0.3311]	[0.2569]	[0.2566]
% of doctors			11.0019***	10.6520***	3.2747	9.3978**	2.4418	10.0452***	10.0315***
			[3.3591]	[3.6333]	[4.1218]	[3.6648]	[4.1411]	[3.7320]	[3.7292]
Economic controls									
Air pollutions			10.0589***	3.8362**	2.3073	4.7076**	2.7733	4.7519**	4.7060**
			[1.8876]	[1.8899]	[2.1406]	[1.9025]	[2.1496]	[2.0202]	[2.0202]
% of firms			0.3847***	0.3373***	0.4374***	0.3402***	0.4359***	0.3465***	0.3461***
			[0.0846]	[0.0826]	[0.0941]	[0.0841]	[0.0953]	[0.0854]	[0.0853]
% of big firms			-0.5603***	-0.3787**	-0.8777***	-0.3460*	-0.8344***	-0.3560*	-0.3523*
			[0.1833]	[0.1873]	[0.2880]	[0.1898]	[0.2921]	[0.1946]	[0.1946]
Relative			-0.7143**	-0.8859**	-0.9259**	-0.9404**	-1.0012**	-0.9948**	-0.9970**
unemployment			[0.3067]	[0.4020]	[0.4522]	[0.4059]	[0.4541]	[0.4115]	[0.4114]
Industrial controls									
Relative energy				0.0003***	0.0003***	0.0004***	0.0004***	0.0004***	0.0004***
output				[0.0000]	[0.0000]	[0.0000]	[0.0000]	[0.0000]	[0.0000]
Relative manufacturing				0.0001***	0.0001***	0.0001***	0.0001***	0.0001***	0.0001***
output				[0.0000]	[0.0000]	[0.0000]	[0.0000]	[0.0000]	[0.0000]
Log wage spatial lag			0.2939***	0.2763***	0.2435***	0.2646***	0.2303***	0.2490***	0.2492***
			[0.0485]	[0.0460]	[0.0538]	[0.0478]	[0.0554]	[0.0485]	[0.0484]
Obs.	1,054	1,054	911	825	692	825	692	806	806
Regional FE	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R-sq. adj.	0.110	0.665							
R-sq. adj. (Buse 1973)			0.783	0.818	0.805	0.823	0.814	0.825	0.825

Robust standard errors in brackets. *** p<0.01, ** p<0.05, * p<0.1

One more group of controls are economic, which includes the ratio of air pollutions to area; the ratio of firms to all organizations; the industrial concentration; and the unemployment rate. All the economic controls are significant factors for wage in most specifications and have the signs predicted by conventional economic theory. The relative air pollutions from stationary sources reflect the intensity of industrial production and are an obvious covariate of employment and output. During the recession of 1990s, many industrial enterprises decreased or even stopped production which resulted in an improvement of the environmental conditions in the respective areas. At the same time, there are a number of cities with relatively high wage and poor, production-related environmental conditions. An example is Cherepovets, where in 2011 the average wage was 29,324 rubles, which is some 1.5 times the average across the country, while the air pollution was 0.27 tons per square kilometer which is more than 26 times the average across Russia. The positive link between the relative air pollutions and wage vanishes when small towns are excluded, i.e., those with a population under 12,000. This can be explained by the comparatively greater significance of industrial production for small towns.

The relative number of firms is significantly positive in all the specifications. This variable is an important covariate of the quality of institutions (North et al. 2009). Specifically, it reflects the opportunities for doing business, which substantially vary across the cities. Industrial concentration was proxied by the ratio of big firms to all firms as of 2004. Though, due to a lack of other data, we used such a crude proxy, it turned out significant and gave the expected results in all the specifications, viz., the share of big firms was a negative covariate of wage. This link results from the ongoing performance of the mono-cities inherited from Soviet times. These cities feature a small number of big firms, or frequently just a single firm, providing jobs for most of the population of the respective cities. Poor wages in these cities are explained by the lack of incentives on the part of the management to pay more, which is related to a position of the big firms as monopsonies in the respective labor markets (Williamson 1985). On the other hand, many such firms are inefficient and continue working for social rather than economic reasons. In some cases the relatively low wellbeing of these cities has the same reasons as those that led to the decline of Detroit (Glaeser 2011, Glaeser et al. 2010, Breinlich et al. 2014). Their narrow specialization, though it might have been reasonable in the Soviet times, restricts their development in new conditions. Expectedly, unemployment is a negatively significant covariate in all the specifications. Similarly, industrial controls, relative energy output and manufacturing production are highly significant positive covariates of wage.

The interest link is robust to changing specifications and remains highly statistically significant, namely, at less than 0.01 level. The negative association between age and wage is robust to the inclusion of a full set of regional dummies, the spatial lag of the dependent variable and many statistically significant controls. Also the link is robust to different age variables. Columns 6 and 7 contain results for the second age variables and columns 8 and 9 are for third and fourth ones, respectively. Finally, the link is robust to censuring the sample by the cities whose population meets the above-mentioned criteria, namely, is in excess of 12,000. Measures of goodness of fit, adjusted R-squared and that proposed by Buse (1973) for the dependent variable spatial lag regressions indicate the high quality of the regressions and the non-trivial explanatory capacity of the controls. After the inclusion of regional dummies adjusted R-squared jumped from 0.11 to 0.665, which suggests that the bulk of wage inequalities across the cities can be ascribed to the respective inequalities across the regions. Though we cannot directly compare the usual R-squared and Buse's R-squared, higher values of the latter after the inclusion of spatial lag and other controls along with the highly statistically significant coefficient on the lag as well as the significant coefficients on other controls allow us to infer that they contribute much to the quality of our regression. When proceeding from 3 to 4 one can see also that the important industrial controls add explanatory power. At the same, if we compare the results for the full sample with those for the censured sample, namely 4 and 5 for the first age variable, and 6 and 7 for the second age variable, we can conclude that the included controls have more explanatory power when it comes to small towns.

4.3. The link in panel regressions

To make sure that the negative link between age and wage holds not only in 2011, but also in any other year during the post-Soviet period, we used all the data available, 1991-2011, though not all the variables we use are available for all these 21 years. We estimated several panel regressions differing from each other by the included controls and thereby by panel length as well as starting year. At the same time, in all the specifications a full set of the regional dummies and time fixed effects as well as most of the controls were used. The estimation results are presented in Table 5.

Again, here there is a highly significant negative link between age and wage in all the specifications and for the two main age variables. As in the cross-section counterparts, in the panel specifications the link is robust to the inclusion of the full set of regional dummies and the geographical, demographic, social, economic, and industrial controls. Further, what the panel regressions add is that they reveal robustness of the link to the years analyzed, which allows us to conclude that the existence of the negative link between age and wage dates back to at least 1991.

Table 5. Panel regressions, 1st and 2nd age variables

	log wage 1991-2011	log wage 1991-2011 + spatial lag	log wage 1997-2004 + economic controls	log wage 2005-2011 + industrial controls
	(1)	(2)	(3)	(4)
	Panel A			
1st age variable (log)	-0.0701***	-0.0603***	-0.0794***	-0.0534***
	[0.0077]	[0.0103]	[0.0101]	[0.0079]
Log wage spatial lag		0.6330***	0.9708***	0.4024***
		[0.0426]	[0.0068]	[0.0455]
Obs.	16,790	9,849	6,504	4,781
Ids.	863			
Regional and time FE	Yes	Yes	Yes	Yes
R-sq within	0.996			
R-sq between	0.927			
R-sq overall	0.992			
R-sq adj. (Buse 1973)		0.996	0.975	0.972
	Panel B			
2nd age variable (log)	-0.0665***	-0.0517***	-0.0648***	-0.0424***
	[0.0074]	[0.0095]	[0.0092]	[0.0071]
Log wage spatial lag		0.6261***	0.9710***	0.3967***
		[0.0427]	[0.0068]	[0.0460]
Obs.	16,790	9,849	6,504	4,781
Ids.	863			
Regional and time FE	Yes	Yes	Yes	Yes
R-sq within	0.996			
R-sq between	0.926			
R-sq overall	0.992			
R-sq adj. (Buse 1973)		0.996	0.974	0.972

Standard errors in brackets. *** p<0.01, ** p<0.05, * p<0.1

The estimation results for the full 21 year panel are presented in 1 and 2, the second one differing from the first one by the inclusion of the dependent variable spatial lag. Comparing specifications 1 and 2 one can note robustness of the link to not only inclusion of the spatial lag, but also to varying samples. The first specification uses many more observations than the second, though both specifications use the same panels by length. This difference in sample size is related to the estimation method used for the spatial lag specification that requires a fully balanced panel structure, so that when estimating the specification 2 we had to drop observations with unbalanced panels. The specifications 3 and 4 are also the spatial lag ones, but they use shorter panels as a result of the inclusion of controls for which data are available for shorter periods. In particular, specification 3 is for 1997-2004, for which data about the industrial concentration and the number of firms are available, and 4 is for 2005-2011, for which period data about output by industries are available.

Coefficients on the interest variable are somewhat different from one specification to another, which suggests that the link between age and wage, though unambiguously negative over all the years, varies to some extent from year to year. If we compare the interest coefficients in 3 and 4 of Table 4 with their panel counterpart in 2 and 4 of Table 5, we can see that in panel specifications coefficients on the interest variable are lower than those in the cross-section specifications. This comparison along with a comparison of 3 and 4 of Table 5 suggest that at least since the late 1990s the negative age elasticity of wage has been decreasing. Panel specifications, like their cross-section counterparts, reveal a highly significant positive effect of the dependent variable spatial lag, but the coefficients are much higher than those in the cross-section specification. Again, if we take into account that the effect of the spatial lag is decreasing from 1997-2004 to 2005-2011, as is seen in 3 and 4 of Table 5, it suggests that since the end of the previous century the positive effect of the spatial lag has been decreasing, as with the negative effect of the age. Thus, over the last decade spatial dependence as to wage has been weakening simultaneously with a similar weakening of the negative effect of age. It is an additional evidence in favor of the spatial lag being an important control.

Finally, it is worth noting the very high values of the measures of goodness of fit. Values of within, between, and overall R-squareds in specifications 1 of well above 0.9 suggest the very high quality of the regression model. In particular, it explains almost comprehensively variation of wage both within and between the groups. The same can be said about

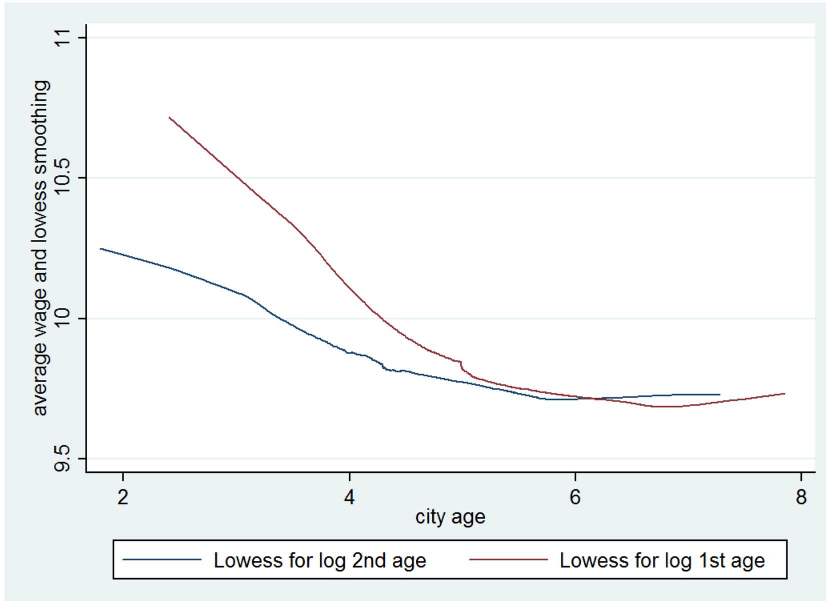

Figure 5. Log wage and log age, simple locally weighted regression

the quality of the spatial lag regressions measured by Buse's adjusted R-squared. The high quality of the regressions is explained by the number of the controls included. At the same time, with the same controls goodness of fit of the panel regressions is higher than that of the cross-section counterparts. This can be explained by the inclusion of time fixed effects in the panel regressions that capture the trend effect as well as other time effects on our dependent variable.

4.4. Localization of the link in the distribution quantile and time

The logarithmic specification we use also implies the diminishing marginal impact of age. The relative predicted difference in average wage between cities founded 100 and 1000 years ago is the same as that between cities founded 7 and 70 years ago. Thus, the marginal effect of an additional year of age is stronger for younger cities. And highly significant estimates and high values of goodness of fit for the log specifications imply that they are correct. However, it is possible that the nonlinearity in our interest link is more than that captured by the logarithmic specification. To trace the link more closely, we have localized the link on the scale of age using a locally weighted regression of log wage on two of the log age variables. Unlike the usual regression predicted values, the smoothed values of the locally weighted regression tend to follow the data and therefore give a more accurate picture of empirical link between the dependent variable and a regressor (Cleveland 1979). The results of the locally weighted scatterplot smoothing with bandwidth 0.8 for the two age regressors are presented in Figure 5.

We see that the slope of the curve is much stronger for younger ages and the curve becomes almost horizontal, if not positively sloped, for older ages. For the first age variable the slope becomes comparatively weaker after $\log age_1 \approx 5.5$, and after $\log age_1 \approx 6.5$ the slope becomes horizontal and still further even upward. For the second age variable the slope begins to flatten approximately after $\log age_2 \approx 4.5$ and becomes almost horizontal after $\log age_2 \approx 5.75$. Thus, range of observations on which the strong negative slope is seen is restricted from above by $age_1 = \exp(5.5) \approx 244.7$ and $age_2 = \exp(4.5) \approx 90$ years, and the range on which some negative slope is observed is restricted from above by 665.1 and 314.2 years, respectively. Thus, substantial negative correlation between age and wage is shown up only for the cities founded under the Catherine the Great and given city status in the Soviet period. Negative correlations between age and wage is absent for most cities founded in Middle Ages and given city status under Peter the Great or earlier. The age limits within which the negative correlation is seen suggest that it is related to the industrial development of Russia, which

Table 6. Cross-section regressions with various sample restriction by age, 2011

	all	<=1000	<=800	<=600	<=400	<=200	<=50
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Panel A						
1st age variable (log)	-0.1427***	-0.1539***	-0.1671***	-0.1869***	-0.2025***	-0.3446***	-0.3935
	[0.0135]	[0.0143]	[0.0156]	[0.0185]	[0.0220]	[0.0413]	[0.3226]
Obs.	1,054	1,030	985	914	812	515	58
Regional FE	No	No	No	No	No	No	No
Full list of controls	No	No	No	No	No	No	No
R-sq adj.	0.110	0.116	0.120	0.121	0.115	0.141	0.0157
	Panel B						
2nd age variable (log)	-0.0940***	-0.0960***	-0.1015***	-0.1196***	-0.1365***	-0.1465***	-0.2474***
	[0.0115]	[0.0119]	[0.0127]	[0.0154]	[0.0171]	[0.0321]	[0.0669]
Obs.	1,054	1,045	1,018	962	921	705	202
Regional FE	No	No	No	No	No	No	No
Full list of controls	No	No	No	No	No	No	No
R-sq adj.	0.0613	0.0613	0.0611	0.0648	0.0719	0.0354	0.0585
	Panel C						
1st age variable (log)	-0.0451***	-0.0537***	-0.0562***	-0.0599***	-0.0611***	-0.0597**	0.1321
	[0.0073]	[0.0079]	[0.0085]	[0.0096]	[0.0115]	[0.0233]	[0.2571]
Obs.	825	803	769	716	641	396	43
Regional FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Full list of controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R-sq adj. (Buse 1973)	0.818	0.820	0.808	0.805	0.797	0.753	
R-sq (Buse 1973)							0.053
	Panel D						
2nd age variable (log)	-0.0288***	-0.0293***	-0.0309***	-0.0372***	-0.0434***	-0.0399**	0.0069
	[0.0066]	[0.0068]	[0.0073]	[0.0083]	[0.0092]	[0.0166]	[0.0308]
Obs.	825	817	796	752	718	548	158
Regional FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Full list of controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R-sq adj. (Buse 1973)	0.823	0.822	0.815	0.809	0.810	0.791	0.897

Robust standard errors in brackets. *** p<0.01, ** p<0.05, * p<0.1

was giving birth to new cities. The most unambiguous negative association between age and wage is observed for cities founded or given city status under Soviet rule. At the same time, the logarithmic specification means that within any age range an additional year of age contributes more to a wage decrease for younger cities.

To clarify this relationship between the correlation and the sample restriction by the age, we have obtained OLS estimates of the regressions with various restrictions for the upper limit of city age. These censured regressions were estimated for the two dependent variables and with only one interest regressor and full set of the controls which were used in our most comprehensive cross-section specification (see Table 4). For all the regressions robust standard errors were estimated, viz., for the simple regressions the sandwich standard errors were estimated, and for the long regressions, spatially lagged robust standard errors were estimated. The results are presented in Table 6 (output for the controls are dropped).

As is seen in Panels A and B of Table 6, the negative slope is monotonously becoming steeper along with restricting the sample by younger cities for both age variables, though the interest coefficient in column 7 of Panel A is insignificant due to much fewer observations. For the subsample of cities younger than 200 by their origin, the slope coefficient is equal to -0.3446. If we compare wage in, say, Suoyarvi founded in 1918 and Alexeevka founded in 1829, their wage difference is to be $\exp(-0.3446\log(182/93)) \approx 0.79$, i.e., in this subsample doubling the difference in age is compatible with a more than 20 percent reduction in average wage. (In this case it is quite near to actual ratio of their wages, 20,338 rubles to 16,182 rubles.) The regressions with the second age variable display the still more unambiguous tendency of the decreasing slope coefficient up to the sharp jump from column 6 to column 7 meaning that cities which were given city status after 1961 display the strongest negative correlation between their age and wage. Also there is a slight tendency of increasing goodness of fit from larger subsamples to those more restricted by the upper age limits, meaning that more variation in the dependent variable is explained by age for younger cities.

The long regressions, presented in Panels C and D of Table 6, show a similar tendency, as one restricts samples

Table 7. Quantile regressions, q0.5, q0.25, q0.75. Cross-section 2011

	q0.5	q0.25	q0.75	q0.5 all controls	q0.25 all controls	q0.75 all controls
	(1)	(2)	(3)	(4)	(5)	(6)
	Panel A					
1st age variable (log)	-0.1326***	-0.0872***	-0.1838***	-0.0349***	-0.0332***	-0.0533***
	[0.0147]	[0.0142]	[0.0213]	[0.0087]	[0.0063]	[0.0130]
Obs.	1,054	1,054	1,054	825	825	825
Regional FE	No	No	No	Yes	Yes	Yes
Pseudo R-sq	0.048	0.030	0.074	0.659	0.668	0.669
	Panel B					
2nd age variable (log)	-0.0804***	-0.0381***	-0.1365***	-0.0202***	-0.0169**	-0.0347***
	[0.0135]	[0.0137]	[0.0188]	[0.0076]	[0.0067]	[0.0119]
Obs.	1,054	1,054	1,054	825	825	825
Regional FE	No	No	No	Yes	Yes	Yes
Pseudo R-sq	0.025	0.007	0.048	0.656	0.663	0.665

Standard errors in brackets. *** p<0.01, ** p<0.05, * p<0.1

by younger cities, but this stops a step earlier, and now the last column contains insignificant coefficients for both age variables. Thus, the picture is basically the same, though the inclusion of the controls makes the tendency of the decreasing of the slope coefficients somewhat less unambiguous meaning that part of this tendency is due to some correlations between the age variables and the controls. As for the measure of goodness of fit it decreases slightly when censuring the data by younger cities, which may mean that the full set of the controls have higher explanatory power for wage in older cities.

Combining the results of the locally weighted regressions and the censured regressions we can conclude that, though the most ancient cities are much poorer than the youngest cities, the inverse link between age and wage for the most part is displayed by relatively young cities. In most spectacular form, the negative association between age and wage is shown up by cities founded or given city status less than a century ago, viz., since launch of the Soviet industrialization in 1930 or even much later. Recall that more than half of Russian cities were founded after 1930, and it is due to these cities that this negative association is observed.

The results of the nonparametric regressions imply also that the subsample of relatively rich cities are to display a stronger negative association between age and wage. To make sure of it, we estimated a number of quantile regressions

$$q_\alpha(\log wage_i | \log age_i, \mathbf{X}_i) = \log age_i \beta_1 + \mathbf{X}_i \boldsymbol{\beta}_2 + \epsilon_i \tag{2}$$

where q_α is a quantile of level α of $\log wage_i$ conditional on $\log age_i$ and $\mathbf{X}_i$. We estimated quantile regressions for both age variables without and with the full list of controls (excluding the spatial lags) and for α equal to 0.5, 0.25, and 0.75. The results are presented in Table 7.

Again, we see a highly significant negative link between age and wage in all the specifications. Thus, we can conclude that this link exists not only for mean wage, but for various quantiles of it, meaning that age matters both for relatively poor and rich cities. Comparing the results in columns 1 in Panels A and B of Tables 6 and 7 we can see that the result of the OLS estimation of the simple mean regression is somewhat higher than that for the median regression, which is likely to be explained by outliers which do not affect the median regression, but do affect the mean regression. As is seen in Figure 4, the distribution of wage has a heavy right tail, and, though it was eliminated by the log transformation, a comparison of the results for the mean and median regressions indicate the presence of some right tail in the distribution of even log wage.

The differences between quantiles for the interest coefficients have the same order in various specifications and for both variables. The absolute values of the slope coefficient are the highest for the 75 percent quantile, lowest for the 25 percent quantile, and for the median they have intermediate values. These differences between the quantiles are quite substantial when estimating simple regressions. Specifically, the absolute value of the slope coefficient for the first age variable for the 75 percent quantile is more than twice that for the 25 percent quantile; for the second age variable the same ratio is well above the three. When estimating long regressions the coefficients differed from each other in the same way, though more modestly. Thus, the negative link is more intensive for a higher distribution quantiles of the urban wages, and it less intensive and even tends to vanish when it comes to quite lower quantiles. We estimated other quantile regressions up to 0.01 and 0.99 quantiles, and the coefficient tends to become insignificant when estimating regressions with a lower quantile, but it becomes more significant for higher quantiles. Consequently, the downward effect of age is very strong for rich cities, it is weaker for poorer cities, and this effect actually absent for the poorest cities.

The relationships between the slope coefficients for various quantiles are compatible with the respective relationships between measures of goodness of fit. For the simple regressions, age alone gives the largest share of the explained variation of the 75 percent quantile, this share is least for the 25 percent quantile, and for the median it has an intermediate value. And this is true for both age variables. For the long regressions these relationships in goodness of fit hold in much less extent only when comparing the 75 percent quantiles with other quantiles. Thus, age contributes much to explaining the variation of comparatively high wages, while variation of low wages are much less explained by age even if the latter is highly significant. The relationships between the coefficients for the high and low quantiles remain in the long regressions to a much less extent, so that some part of them is explained by the controls. The differences between the age variables for the interest coefficient and goodness of fit in both the short and long regressions for the same quantiles are significant. In all cases, for the first age variable a stronger negative link and higher regression quality are observed than for the second variable. It suggests that age matters more when measured by foundation year than by a year of gaining city status.

The localization of the link between age and wage has an implication for the dynamics of the relative wage. If the established link between age and wage is persistent across time, there is to be observed a downward trend of relative wage for any given city, because as time goes by it gets older and, as the link implies, relatively poorer. At the same time, the fact that the link holds for logarithmic specifications and is stronger for younger cities means that an additional year contributes to the decreasing relative wage much more for younger cities. As we have seen, starting from some age the latter actually does not have any downward effect on wage, and even may have positive effect. It suggests an income convergence among Russian cities if we use a lower age limit. The dynamics may take such a form that as a city crosses some age thresholds its wage differential compared to other cities becomes lower and lower. Or, more generally, for any given subsample, younger cities are relatively rich, but their wage advantage falls with time. For slightly older cities the same pattern is to be observed, but to a lesser extent. Finally, for sufficiently old cities the situation is reversed: they are poor, but their relative wage is constant or even rising as a result of the comparatively rapid falling wage in younger cities.

4.5. Bottom line

We have established a number of facts concerning the link between age and wage. There is a highly significant negative link between log city age and log wage whether it is estimated on cross-section data or panel data. The link is highly robust to the inclusion of a number of controls, namely, a full list of regional and time fixed effects, the spatial lag of the dependent variable, geographical, demographic, social, economic, and industrial controls. The link is also robust to various definitions of city age, or a combination of them. Finally, the link is observed in various quantiles of the dependent variable. A substantial part of the link observed in simple regressions vanishes after the inclusion of the controls, meaning that this part results from the respective omitted variables bias. Nevertheless, in the long regressions highly significant negative coefficients on age are observed. We conclude that age is an important covariate of wage that does not disappear after the inclusion of any controls.

This link is more intensive, the higher the wage quantile considered, or, the younger the subsample of cities considered. As a whole, younger or richer cities are more sensitive to an additional year of age as a negative factor of wage; and vice versa old or poor cities are relatively insensitive to an additional year of age up to the total disappearance of the negative correlation between age and wage. Such a pattern follows from the interpretation of the logarithmic specification. This pattern is reinforced by running simple non-parametric regressions as well as short and long quantile and censured regressions. Specifically, the link for the 75 percent quantile is much more intensive compared with that obtained by median regression; and even more so when comparing the 90 percent quantile with the 10 percent quantile. When running the regression with various restrictions to age, the intensity of the negative link is the more intensive, the younger the city.

Thus, the link shows convergent dynamics. Given the fixed subsample of the cities by their age, wage gaps between youngest and oldest cities decline as time goes by. In particular, the ratio of average wage to overall average wage declines with time for the youngest cities, but is persistent or even grows for relatively old ones. These established results as to the relationship between age and wage make an interesting pattern of spatial income distribution. Our ultimate goal is to determine what hypothesis among those designed to explain spatial patterns of economic activities and their rewards can give the most reliable explanation of these facts. In the first place, we develop a model based on the sophisticated geography hypothesis to present a potential mechanism behind our empirical regularities. Then the task will be to test the predictions following from the model.

5. The model

5.1. An approach

As mentioned, the sophisticated geography hypothesis suggests that geographical characteristics can change with time in their relative importance and in their presence. Such a characteristic is an exhaustible resource. As with any other resource, it can change in its nominal value depending on long term economic developments and varying market conditions. At the same time, as the resource is being depleted, its stock as a local characteristic is vanishing. There are a number of results in the field of the economics of exhaustible resources to use when constructing a mechanism underlying our regularities.

Issues addressed in this literature are mostly related to long term patterns of using an exhaustible resource. Hotelling in his classical paper (1931) elaborated a rule underlying the optimal consumption path for a consuming unit on an infinite horizon. As remaining resource stock is decreasing, so is the consumption due to the increase of the resource's marginal value. According to the Hotelling rule, the growth of the resource marginal value is to accord with interest rate. This would make various units of the resource consumed at times equal by their discounted marginal values. A similar issue is the intergenerational equity in terms of the consumption on an infinite horizon. The nuance is that there are a number of consumers over time, and less resource stocks for future generations cannot be compensated for by a discount factor, as is the case with an individual consumer, but it can by other goods, in particular, hand-made capital. Hartwick (1977) assuming substitutability between natural resources and man-made capital argued that proper capital accumulation could result in the extraction of that resource not harming future generations. The same issue under constant or increasing population, and with or without technological progress was discussed by Solow (1974). Apart from the investment rule, another issue is related to the production function that would enable steady growth. According to an important result for the issue, a condition of constant growth of consumption given constant population is increasing the return on capital, which can offset scarcity (Hartwick 1978); otherwise, under constant returns, consumption is declining (Dasgupta and Heal 1974). Finally, an important issue is that of the resultant value of the exhaustible asset (Miller and Upton 1985). The Hotelling rule suggests positive price growth as a resource is getting scarcer. The ultimate value of the asset depends on both the resource exhaustion and increasing its marginal value. Mine appreciation due to price increase may outweigh mine depreciation due to resource depletion. In particular, if demand is inelastic, mine appreciation can even grow, because the price is increasing more rapidly than the resource is being depleted (Lozada 1995).

When constructing a model our question is different from those addressed in this literature in that our main interest is in the changing relative wellbeing of various locations rather than just the intertemporal allocation of the resource for a spatial unit. Instead of one site that is to live off its resource, we assume a number of sites that come to participate in the overall activity within a larger economic body in different and exogenously given moments after the discovery of their resource deposits.[9] Further, the resource is not essential for living in a location; it is rather a source of bonus. This assumption is

[9] This pattern of developing territory and establishing new locations fits the actual urban history of Russia. In the past Russians pushed towards the North and the East motivated by their search for valuable furs. When arriving at a fur-rich territory they built up zimovie – a fortified establishment designed to enable the newcomers to impose yasak, i.e., tribute in furs on the native population. These establishments gave birth to many Russian cities

based on a mere empirical fact that the income of most modern cities is comprised of receipts from various activities that may or may not include those from extractive production. In Russia a lot of cities were founded on resource-rich territories that subsequently, after their having been exhausted or having lost their previous relative economic importance, ceased to play any role in generating their incomes. Treating the resource income as a bonus enables us to use an additive function of overall income.

From the standpoint of a social planner, the problem is an optimum path for the populations of the locations after their coming into being at their particular moments. An objective function is related to economic growth, i.e., it is a total income, rather than per capita income, over a period that is to be maximized. Hence, the model admits that the solution to the social planner problem is possible only on a finite horizon and conducive to constant or even falling per capita income. In this respect, our model is in the spirit of the Solow model in that per capita income growth is possible only via positive dynamics of exogenously given technological progress (Acemoglu 2009, p. 69). Apart from the precedents in economic theory, this assumption is justified by the empirics of a number of developing countries with constant or falling per capita income due to their rapidly growing population (see, e.g., World Development Indicators 2014). Countries with higher population growth than economic growth exist in the Arab world, Latin America, Sub-Saharan Africa, etc. Moreover, as a social planner problem it corresponds to current Russian demographic policies designed to promote birth rate and reducing mortality, which suggests that population growth is considered to be as an important goal along with income growth.

The Hotelling model suggests that a resource price rises because of its growing scarcity (1931; Lozada 1995). In our case the price is given by overall market conditions, rather than by the relative resource scarcity in a particular location. At the same time, taking technological progress as constant we can assume that the resource price is time-invariant. (However, to grasp the potential effect of the price increases, we examined them in our numerical example.) Whereas other models in the field of exhaustible resource economics assume that resource exhaustion is offset by capital accumulation, in our model the compensation comes partly from the manufacturing industry. Finally, we assume that the two industries operate with diminishing and constant returns, respectively. The former corresponds to the diminishing return in agriculture which was traditionally assumed in economics (e.g., Malthus 1798; Ricardo 1817; Marshall 1890). The constant scale return in the manufacturing industry is also one of the oldest assumptions (e.g., Marshall 1890). As for more recent developments, one might assume the combination of agglomeration effects and congestion costs (see, e.g., Combes et al. 2008), which offset each other resulting in a constant return at the local level. Actually, this assumption is also needed to meet the second order condition for solving the social planner problem. In this respect our assumption of the diminishing and constant returns is a counterpart of the constant and increasing returns in the Dixit-Stiglitz model (1977).

5.2. General setup

We consider an economy dependent on an exhaustible resource. Its territory is a set of n plots endowed with the homogeneous exhaustible resource. The life expectancies of the economy and its locations are finite. Plots are settled as a result of discoveries of their resource deposits. (This assumption may be widened to include any other features that would make it potentially attractive place for residence after their discovery.) Hence, the ith plot begins to be settled only since a moment t_i^* of the respective discovery. It implies that

$$L_i(t) = 0 \text{ if } t < t_i^*$$

where L_i stands for a population size of the ith plot and t_i^* is the birth date of a city on the ith plot. It means that it is the age of a city that is the time span during which the dynamics of a city's population size and per capita income is determined. For a particular location its birth date as a city is a constant, so that t and t_i^* are independent. This coupled with the definition of city age as $\tau_i = t - t_i^*$ enable us to argue that

(see, e.g., Kluchevsky 1911; Lappo 1998; 2006). In the Soviet times the foundation and development of the bulk of the new cities were motivated by only resource deposits to be developed (Lappo 1998; 2006).

$$\frac{\partial \tau_i}{\partial t} > 0 \Rightarrow \frac{\partial t}{\partial \tau_i} > 0$$

from whence

$$\text{sign}\left(\frac{\partial x}{\partial t}\right) = \text{sign}\left(\frac{\partial x}{\partial \tau_i}\right) \tag{3}$$

where x is some variable.

Once settled, a city has two industries, an extractive industry, which is related to the existing resource, and a manufacturing one. The technology used in each industry is the same across all cities, and the only input to both industries is labor L. An extractive production function $Y(\cdot)$ is a thrice continuously differentiable, and is featured by a diminishing rate of return owing to the growing unit cost of extracting when one exploits less and less convenient parts of a plot. However, as is conventionally assumed, the extractive marginal product function is a convex one. The manufacturing industry has a constant rate of return. The production functions in both industries are then defined as follows

$$Y(0) = 0,\; Y'(L_i) > 0,\; Y''(L_i) < 0,\; Y'''(L_i) > 0, \tag{4}$$

$$G(0) = 0,\; G'(L_i) > 0,\; G''(L_i) = 0 \tag{5}$$

where Y and G denote outputs of the extractive and manufacturing industries, respectively.[10] Since labor is the only input, neither industry can produce anything without a population in a location.

A resource in the ith plot begins to grow depleted once a city occurs there and its extractive industry begins to operate. Thus, the dynamics of remaining stock of the ith resource is given by the following standard rule

$$\dot{S}_i = -Y(L_i) \tag{6}$$

where S_i is a remaining stock of a resource in the ith plot.

The manufacturing industry's produce is homogeneous across the cities and serves as a numeraire. A nominal value of a unit of the resource $P(t)$ is subject to the state of technology. Since technological progress is a long-term process, the nominal value changes in the very long run, i.e., slowly enough to assume its almost zero time derivative, so that

$$\dot{P}(t) \approx 0. \tag{7}$$

Then the income of the ith city is comprised of the produce' value made by the extractive and manufacturing industries

$$I_i = P(t)Y(L_i) + G(L_i). \tag{8}$$

A key feature of the economy we deal with is the dependency – direct for the extractive industry and indirect for the manufacturing industry – on an exhaustible resource, which makes the economy's life expectancy potentially finite. Unlike the standard problem of an exhaustible resource consumption (Hotelling 1931), we assume the presence of the two industries that is needed to make our problem tractable. (As is seen below, specifically in (13), under one industry we could not find a solution for the population path.) Finally, we assume an income, rather than per capita income, as an objective function to be maximized, i.e., the economy is driven by aiming at welfare coupled with population size, rather than just social welfare. Thus, the economy is to develop extensively as in pre-industrial and many industrializing countries.

[10] Set of assumptions as to $Y(L)$ here is quite standard. The Cobb-Douglass production function is one example of a technology relating a particular input and output as it is implied in (4).

5.3. The analysis of dynamics

To determine the dynamics of the ith location's per capita income, take the time derivative of (8) divided by L_i, and simplify it to obtain

$$(\iota_i)_t = \left(\frac{I_i}{L_i}\right)_t = \dot{P}\bar{Y}_i + \frac{\dot{L}_i}{L_i}\{P(Y_i' - \bar{Y}_i) + (G_i' - \bar{G}_i)\} \tag{9}$$

where $\bar{Y}_i = Y/L_i$ and $\bar{G}_i = G/L_i$. The first term of (9) is near to zero as assumed in (7). As for the second term, using (4)-(5) we have $(Y_i' - \bar{Y}_i) < 0$ and $(G_i' - \bar{G}_i) = 0$ so that sum of the terms in the brackets is negative. Thus, the sign of the second term of (9) depends on the sign of population growth $\dot{L}_i$.

To determine that sign, a social planner problem can be solved. The dynamics of the population in n plots can be thought of in terms of maximizing the social welfare. The problem would be as follows

$$\max_{L_i(t)} \int_{T\in[0,\infty[} \exp(-\rho t)\sum_{i=1}^{n}[P(t)Y(L_i) + G(L_i)]dt \tag{10}$$

subject to (7) and (6)

$$S_i(t) \geq 0,\ S_i(0) = \underline{S_i} > 0,$$

where $\rho > 0$ is a time-constant rate of intertemporal preference of the society. As was said, our assumption is that the social planner maximizes total income.

Given that the integrand in (10) is a divergent one, the problem is a free endpoint one. In particular, the time span for operating the economy T is to be chosen endogenously along with $L_i(t)$. And the respective trasversallity condition is given by

$$\bar{S}_i\bar{\nu}_i = S_i(\bar{t}_i) = 0, \tag{11}$$

where $\nu_i(t) > 0$ is a costate variable for the ith location, which is conventionally treated as the time-varying marginal value of the ith plot's resource stock.

The current-valued Hamiltonian for this problem is

$$\mathcal{H} = \exp(-\rho t)H = \sum_{i=1}^{n}[P(t)Y(L_i) + G(L_i)] + \mu\dot{P}(t) - \nu_i Y(L_i) \tag{12}$$

where $\mu(t)$ is the time-varying marginal value of the resource price in units of social welfare. The second order conditions are ensured by (4) and (5) which imply the concavity of the integrand of (10). Then the optimal interium paths of the locations' populations are given by FOC for L_i whereof one gets an expression for the marginal labor product in the extractive industry

$$Y'(L_i)\,(\nu_i - P) = G'(L_i). \tag{13}$$

The FOC in (13) sets a dynamic marginal principle for the marginal products in the two industries. The optimal population size implies an equality between the industries marginal products adjusted for the opportunity cost of the resource and the difference between their prices. FOC for S_i after solving for $\dot{\nu}_i$ gives us

$$\dot{\nu}_i = \rho\nu_i \tag{14}$$

which gives a solution for ν_i

$$\nu_i = C_i \exp(\rho t) \tag{15}$$

where C_i is a constant of integration or an initial value of the ith costate variable.

Taking the time derivative of (13), rearranging it, and solving for labor dynamics we have

$$\dot{L}_i = \frac{G''(\nu_i - P) - G_i'(\dot{\nu}_i - \dot{P})}{Y''(\nu_i - P)^2}.$$

which using (4), (5), and (14) results in

$$\dot{L}_i = -\frac{G'\dot{\nu}_i}{Y''(\nu_i - P)^2} > 0. \tag{16}$$

Thus, the optimal time path for population implies positive growth that renders a negative sign of (9).

Starting from this result we can make the following propositions.

Proposition 1. *The age of a resource extracting location is inversely related to its per capita income.*

Proof. From (7) and (16) we established the negative sign of (9). Then using (3) we can establish the sign of the respective derivative

$$\frac{\partial \iota_i}{\partial \tau_i} = \frac{\partial \iota_i}{\partial t} \times \frac{\partial t}{\partial \tau_i} = \frac{\dot{L}_i}{L_i}\{P(Y_i' - \bar{Y}_i) + (G_i' - \bar{G}_i)\} \times \frac{\delta t}{\delta \tau_i} < 0. \tag{17}$$

Thus, per capita income goes down with age. At the same time, if we compare the per capita incomes of two locations with different start times, we will have

$$\iota_i - \iota_j = P(\bar{Y}_i - \bar{Y}_j) + (\bar{G}_i - \bar{G}_j) \lesseqqgtr 0 \text{ for } (\tau_i - \tau_j) \gtreqqless 0 \quad \forall \tau_i, \tau_j > 0 \tag{18}$$

The signs are established from (3)-(5). □

Proposition 2. *The age of a resource extracting location is inversely related to its per capita extractive output.*

Proof. If we compare the per capita extractive outputs of two locations, we will have

$$\frac{Y_i}{L_i} - \frac{Y_j}{L_j} = \bar{Y}_i - \bar{Y}_j \lesseqqgtr 0 \text{ for } (\tau_i - \tau_j) \gtreqqless 0 \quad \forall \tau_i, \tau_j > 0$$

where the sign is established using (3). Similarly, the per capita extractive output goes down with time and thereby with age. This is shown in the sign of the time derivative of the ith per capita extractive output which, after simplifying, is

$$\left(\frac{Y_i}{L_i}\right)_t = \frac{\dot{L}_i}{L_i}\left\{Y_i' - \bar{Y}_i\right\} < 0$$

where the sign is established from (4). □

Proposition 3. *The difference between the per capita incomes of locations with different ages goes down with time. Thus, a convergence in terms of per capita income takes place among a fixed number of locations.*

Proof. Taking the time derivative of (18) and simplifying it we have

$$(\iota_i - \iota_j)_t = \dot{P}(\bar{Y}_i - \bar{Y}_j) + P\left[\frac{\dot{L}_i}{L_i}\left\{Y_i' - \bar{Y}_i\right\} + \frac{\dot{L}_j}{L_j}\left\{Y_j' - \bar{Y}_j\right\}\right] + \frac{\dot{L}_i}{L_i}\left\{G_i' - \bar{G}_i\right\} + \frac{\dot{L}_j}{L_j}\left\{G_j' - \bar{G}_j\right\} \tag{19}$$

From (7), (4), and (5) the first, third, and fourth terms are near or equal to zero. From (3) one can posit that

$$|Y_i' - \bar{Y}_i| - |Y_j' - \bar{Y}_j| \gtreqless 0 \text{ for } (\tau_i - \tau_j) \lesseqgtr 0 \quad \forall \tau_i, \tau_j > 0. \tag{20}$$

Thus, the sufficient condition for (19) to be less than zero is $\dot{L}_i/L_i \geq \dot{L}_j/L_j$. This inequality would hold if $(\dot{L}_i/L_i)_t \leq 0$. To establish the sign of this inequality, take the time derivative and simplify it as follows

$$\left(\frac{\dot{L}_i}{L_i}\right)_t = \frac{G'(L_i)}{Y''(L_i)(\nu_i - P)^2 L_i}\left[\frac{2(\dot{\nu}_i - \dot{P})^2}{(\nu_i - P)} - (\ddot{\nu}_i - \ddot{P}) + \frac{\dot{L}_i}{L_i}(\nu_i - P) + \frac{Y'''(L_i)}{Y''(L_i)}\dot{L}_i(\nu_i - P)\right]. \tag{21}$$

The derivative in (21) will be negative if the expression in brackets is positive. To establish the sign of that expression, let us rewrite it using (7), and (14)-(16) as

$$\frac{\ddot{\nu}_i(\nu_i + P)}{(\nu_i - P)} - \frac{Y'(L_i)\dot{\nu}_i}{Y''(L_i)(\nu_i - P)}\left[\frac{Y'''(L_i)}{Y''(L_i)} + \frac{1}{L_i}\right]$$

that amounts to

$$\frac{\dot{\nu}_i}{\nu - P}\left\{\rho(\nu_i + P) - \frac{Y'(L_i)Y'''(L_i)}{Y''(L_i)^2} - \frac{Y'''(L_i)}{Y''(L_i)L_i}\right\}.$$

In the brackets one of the three terms $Y'Y'''/\left(Y''\right)^2$ is negative, and it is a constant, while the first term $\rho(\nu_i + P)$ is an exponential function as is seen from (15). Hence, at least starting from some moment t the expression in the brackets is positive and thereby the time derivative in (21) is negative. This in turn makes (19) negative as well. The negative sign of the time derivative (19) means that difference between per capita incomes of locations distinct in terms of their ages goes down with time. □

Proposition 4. *The difference between the per capita extractive outputs of locations with different ages goes down with time. A convergence in terms of per capita outputs takes place among a fixed number of locations.*

Proof. Take time derivatives of the difference between per capita extractive outputs of different locations and simplify them to get

$$(\bar{Y}_i - \bar{Y}_j)_t = \left[\frac{\dot{L}_i}{L_i}\left\{Y_i' - \bar{Y}_i\right\} + \frac{\dot{L}_j}{L_j}\left\{Y_j' - \bar{Y}_j\right\}\right] < 0. \tag{22}$$

The sign of the right side of (22) has been established in the proof of Proposition 3. □

An additional conclusion about the initial resource stock can be made. Specifically, the ith location's initial resource stock is directly related to its initial relative per capita income and influences its subsequent dynamics. Let us first clarify the relationship between the initial resource stock and the initial value of the respective costate variable. For that we divide (14) by (6) to obtain the following derivative of the ith costate variable with respect to the ith resource stock

$$\frac{\partial \nu_i}{\partial S_i} = -\frac{\rho \nu_i}{Y(L_i)} \tag{23}$$

which implies that, other things being equal,

$$\underline{S_i} - \underline{S_j} \gtreqless 0 \text{ if } C_i - C_j \lesseqgtr 0,$$

i.e., a difference in the initial values of the costate variables between locations which means the opposite difference in the respective resource endowments.

This implication also affects the optimal life expectancy of a city. To see that, we present a function of population using (15) and (16) as follows

$$L_i = \phi(C_i \rho \exp(\rho t)), \text{ ceteris paribus, } C_i - C_j > 0 \Longrightarrow \phi(C_i) - \phi(C_j) > 0 \tag{24}$$

where the sign of the difference is established using the (4) and (13). Then if we solve (6) using (11), we obtain

$$\bar{S}_i = \underline{S_i} - \int\limits_{T_i \in T} Y_i(\phi(C_i \exp(\rho t)))dt = 0. \tag{25}$$

From (24) and (23) follows that a higher initial value of the ith costate variable is compatible with a lower value of $\underline{S_i}$, but a higher value of the integrand in (25). This means that (25) can hold only provided that the integration interval is narrower for lower $\underline{S_i}$, which amounts to a shorter life expectancy of the ith location T_i. In other words, a smaller initial resource stock implies that the income losses imposed by the life shortening are to be heavier than benefits related to the higher instantaneous incomes.

5.4. A numerical example

To better understand the dynamics of relative per capita income implied by the model, let us consider an example with the following specifications:

$$Y(L_i) = L_i^{\kappa}, \quad \kappa \in]0, 1[\tag{26}$$

$$G(L_i) = \lambda L_i, \quad \lambda \in]0, \infty[\tag{27}$$

The specifications in (26) and (27) meet the constraints imposed on the production functions by (4)-(5). Using these specifications one can integrate (16) to get the exact function of a location's population

$$L_i = \left[\frac{\exp(\rho t) - P}{\lambda(1-\kappa)} \right]^{\frac{1}{1-\kappa}} \tag{28}$$

The solution (28) suggests that a location's population size goes hand in hand with the local opportunity cost of the resource and in the opposite direction to the remaining stock of the resource. Another positive factor for the local population is the labor elasticity of production κ meaning that the diminishing return in one of the industries hampers population growth. The two negative factors are the marginal labor product in the manufacturing industry λ and the resource price P. The first factor sets a minimum value of the marginal labor product in the extractive industry as it is seen in (13), and the higher its minimum value, the lower the population size of the location to meet the requirement (13) because of the negative relationship between labor and the extractive industry's marginal product (4). As for the price, the marginal principle in (13) indicates that the price offsets the growing opportunity cost of the resource a location faces, making it less necessary to increase its population to support the dynamic balance between the net nominal marginal labor products in the two industries. Also we examine the impact of a changing resource price. In particular, we took an exponential positive time function and a periodic time function.

Based on these specifications, we can get the dynamics of the ratio of per capita incomes of two locations with different ages. In the Figure 6 a plot is presented with the following parameters $\kappa = \lambda = 0.5,\ \rho = 0.1$ and $C_1 = C_2 = P = 1$ in A; $C_1 = 1,\ C_2 = 1.4,\ P = 1$ in B; $C_1 = 1,\ C_2 = 1.5,\ P = 1$ in C; $C_1 = C_2 = 1,\ P = 1.2$ in D. Finally, in E and F of Figure 6 the resource price is given as a time function, namely $P = \exp(0.05t)$ in E and $P = 1 + 0.1\sin(t)$ in F, while the other parameters are as in A. In both cases we took the coefficients to exclude negative values of labor and price.

Given the equal initial resource stocks, a new born city has a much higher per capita income. However, this advantage rapidly disappears resulting in the convergence of the per capita incomes of cities with different start times. If a new born city has a less resource endowment than an existing city, its advantage in per capita income is lower (B); under still less

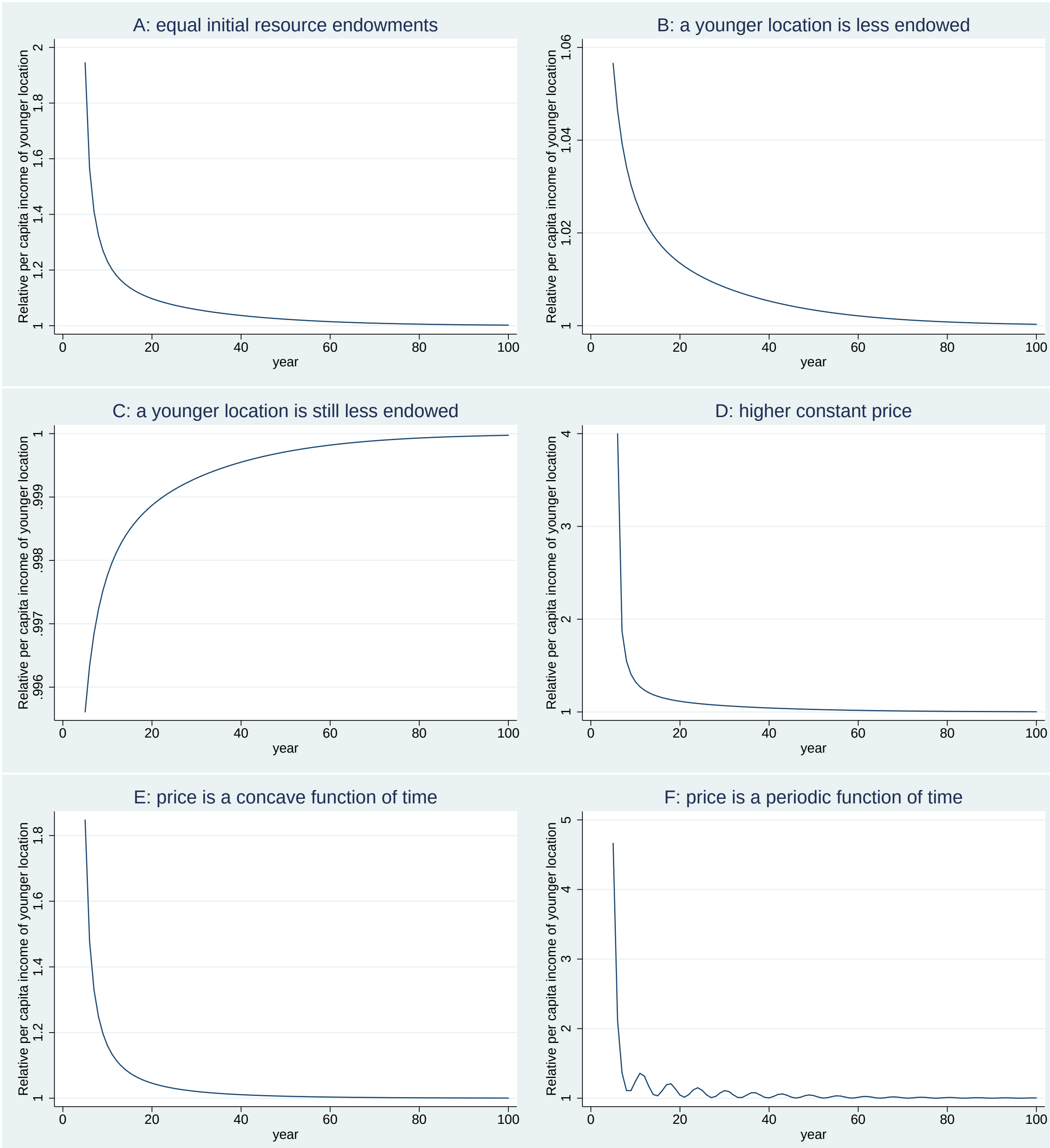

Figure 6. Ratio of per capita income of younger location to that of older one (five year difference in age)

resource endowment of a new born city compared to that of an existing city this may fully offset its initial advantage and even make its initial per capita income lower than that of the existing city as it is seen in C of Figure 6. However, this again implies a convergence in terms of per capita incomes, but from the opposite side. Actually, the advantage in per capita income is due to the difference in the remaining resource stocks, and the latter depend on both the initial endowments and how long this endowment is exploited. If a location is much richer in its initial endowment than other, it may still have more even if the other location began to use its stock later. Any comparative advantage in resource stock, whether this advantage results from the endowment differential or from differences in depletion, vanishes with time.

A higher price implies a higher initial advantage in terms of per capita income for a location with higher remaining resource stock (D). If price increases with diminishing marginal jumps per time unit, the relative per capita income falls more slowly, because the diminishing return in the extractive industry is partly offset by the increasing price. Given a less concave or even convex time function of price, the relative per capita income of a younger location may initially even grow. In our particular example (E) the exponential time function of the resource price is still compatible with the downward dynamics of the relative per capita income. Finally, there may be price fluctuations, in which case the long term trend is still downward, though this overall tendency is going through temporary jumps and slumps (F).

5.5. The implications and predictions

The model outlines a growth pattern of n-city and a two-sector economy with resource depletion. This enables one to derive a number of features of economy dependent on an exhaustible resource and comprised of extractive and manufacturing industries with technologies conventionally attached to them. Such an economy displays extensive growth, meaning that growth in terms of total income may be equal or less than that in terms of population size. An optimal path implies a positive absolute (but not relative) population growth in every location once it is settled.[11] An individual location begins with its maximum per capita income, which then decreases. The same is true for the per capita extractive output. Thus, a younger city is better-off and has a higher per capita extractive output than an older one.

What about the dynamics of per capita income for the whole economy? A particular city faces a diminishing per capita income. However, it does not necessarily mean that per capita income for the whole economy is to be diminishing, as decreasing per capita incomes in existing cities can be offsetting by the relatively high per capita incomes of new born cities. At the same time, the ultimate dynamics of the economy's per capita income depends on how rapidly per capita incomes of existing cities are falling. Positive growth of the economy's per capita income can be facilitated by that new born cities have higher per capita incomes and the decreasing of per capita incomes of existing cities is slowing down. That this is true is established in Propositions 1 and 3.

Finally, a higher initial resource stock of a plot implies lower initial and subsequent population sizes, higher per capita income, and longer life expectancy for the city.

6. Hypotheses and their testing

The model derived in the previous section presents a data-generating process which may underly the negative link between age and wage. According to the model, a major factor of relative per capita income is a remaining resource stock; naturally, the latter is inversely related to the age of a location; hence, what underlies the link is the negative correlation between the age of a location and its remaining resource stock.

The process that actually generates the link is to bring about additional regularities. If x is a key force behind the data-generating process, then we should observe a certain pattern of slope coefficients in the following regressions

$$\log wage_{it} = x_{it}a_1 + \mathbf{X}_{it}\boldsymbol{a}_2 + e_{it} \tag{29}$$

$$x_{it} = \log age_{it}b_1 + \mathbf{X}_{it}\boldsymbol{b}_2 + u_{it} \tag{30}$$

Specifically, the key slope coefficients in these regressions are to obey the following rule

$$\text{sign}(a_1 b_1) = \text{sign}(\beta_1) \tag{31}$$

[11] Though in Russia over the post-Soviet period a whole population declined (World Development Indicators 2014), the resource extracting cities actually grew. Specifically, according to Economica gorodov Rossii (2013), among 1005 cities with the available data on their population size for 2001-2011 mean population growth across 683 cities that did not produce any extractive output in 2011 was equal -160 persons, while in remaining 322 extracting cities this figure was 853 persons. Thus, an incidental result from the model that extractive production is compatible with positive population growth fits the evidence from Russia.

where β_1, let us remind, is the slope coefficient for the age in the base-line regression (1). Thus, for x to be the driving force of the link, the sign of the interacted slope coefficients is to be the same as that for the base-line regression. For example, among possible mechanisms existing in the literature, this requirement would meet by the institutions hypothesis if the city age decreased the institutions quality while the latter increased wages; or an environmental hypothesis would be supported if the city age increased the pollution and this decreased wages.

At the same time, to correspond to the link, x is to follow similar dynamics. To clarify this requirement, let us give the dynamics implied by the link as follows

$$rw_{it} = year_t \gamma_1 + \mathbf{d}\boldsymbol{\gamma}_2 + year_t \times \mathbf{d}\boldsymbol{\gamma}_3 + \mathbf{X}_{it}\boldsymbol{\gamma}_4 + \eta_{it}, \tag{32}$$

$$rx_{it} = year_t g_1 + \mathbf{d}\boldsymbol{g}_2 + year_t \times \mathbf{d}\boldsymbol{g}_3 + \mathbf{X}_{it}\boldsymbol{g}_4 + v_{it} \tag{33}$$

where the dependent variables are defined as

$$rw_{it} = \frac{wage_{it}}{\frac{1}{N}\sum_{i=1}^{N} wage_{it}},$$

$$rx_{it} = \frac{x_{it}}{\frac{1}{N}\sum_{i=1}^{N} x_{it}},$$

i.e., rw_{it} and rx_{it} denote the relative wage and relative value of x in the ith city for the tth year, respectively, $year$ is a trend variable, $\mathbf{d} = (d_j)_{j=1,2,...k}$ is a vector of k age dummy variables from younger to older. If x is a source of the link, we should obtain the estimates satisfying the following condition

$$\text{sign}(\gamma_{4j} - \gamma_{4j-1}) = \text{sign}(g_{4j} - g_{4j-1}) \tag{34}$$

The condition (34) means that the same age groups of the cities are to show similar dynamics of their relative wage and their relative values of x. Naturally, the following condition also must hold

$$\text{sign}(\gamma_{2j} - \gamma_{2j-1}) = \text{sign}(g_{2j} - g_{2j-1}), \tag{35}$$

but this condition is implicitly contained in (31) as its special case, in which x is negatively linked to age, as wage is.

If the model really describes a data-generating process behind the negative link between age and wage, the latter link is to go hand in hand with the negative link between age and per capita resource extraction; at the same time, per capita resource extraction is to be a positive factor for wage. Also the relative dynamics of per capita resource extraction by city age are to replicate those for wage. Thus, our model gives us two hypotheses: a static hypothesis given in (29)-(31) and a dynamic one given in (32)-(34). As the model suggests, x within our hypothesis is a measure of extractive production. Then the static hypothesis predicts that $a_1 > 0$, and $b_1 < 0$, i.e., that the extractive production is a positive factor of wage, but is inversely related to age. The same static hypothesis would be supported when running the dynamic equation (33) if $g_{2j} - g_{2j-1} < 0$, i.e., younger cities on average earn more, which is reflected in the higher coefficients on the respective age dummies. The dynamic hypothesis predicts that $g_{4j} - g_{4j-1} > 0$ meaning that the downward dynamics of the relative wage is stronger for younger cities. Generally, our hypotheses imply that extractive output is higher in younger cities making them richer, and resource exhaustion brings about the downward dynamics. At the same time, cities with less remaining resources are poorer, but are less dependent on their resource extraction making their per capita incomes less subject to the resource exhaustion. Thus, the younger the city, the higher its wage and the more downward its wage dynamics.

Thus, here we exploit a standard method of differences-in-differences that allows groups to be distinguished by the effect of a treatment on them. As the model suggests, in this case time has a treatment effect. Treatment groups are two age

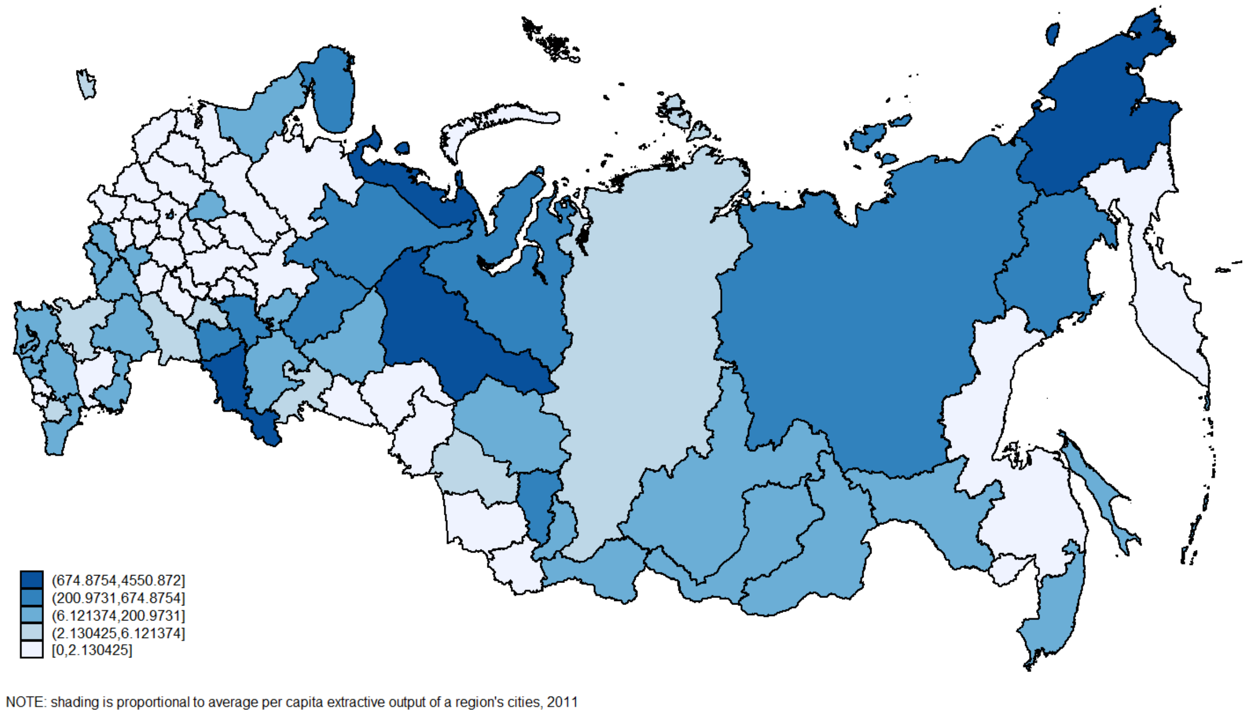

Figure 7. Russian regions shaded by their extractive output per capita

groups differing from the control group by their relative youth. The prediction of the model would be fulfilled if younger groups were richer with respect both to their wage and their per capita resource output, while the dynamics of their relative wage and per capita output were more downward.

7. Results

7.1. Map analysis

Our first task is to see whether there is a relationship between age and a measure of the extractive output similar to that between age and wage. A first impression can be taken when examining the map. Recall that the link between age and wage is readily seen in the map when comparing the regional map colored to indicate relative age of regions' cities with those colored to indicate the relative wage (Figure 1). Regions colored with darker blue on the first map are usually those colored with lighter blue on the second map and vice versa, meaning that regions with relatively old cities are featured by relatively low wages. If a measure of extractive output underlies this link, as our model suggests, a similar visual pattern will be seen for extractive output. As measures of the latter we used per capita extractive output averaged across the cities of a region. The respective map is presented in the Figure 7.

Comparing Figure 7 with Figure 1 we can easily check that per capita extractive output basically follows urban wages. When moving from west to east we see the European part of Russia which is relatively poor with respect to its per capita extractive output. The majority of the European regions are among those with least per capita extractive output. Important exceptions are the extreme northern and southern parts of the European Russia: Murmansk Oblast and Orenburg Oblast, respectively. Also there are a number of regions, for the most part in the south, with modest per capita resource extraction. In the northern part the resource measures follow relative wage. The two capital regions, those of Moscow and Saint-Petersburg, feature relatively low resource extraction, but high wages which is explained by their administrative functions and their positions as big agglomerations. In the south, in some cases there is a lack of correspondence between per capita resource extraction and wage. A number of such regions have rich deposits of resources with relatively low market value, for example, chalk, iron ore, cement, salt. In other cases the resource potential is not yet fully realized because of a lack of investment. Finally, many deposits are almost exhausted and their further exploitation imposes substantial cost, which may be offset by the relatively low wages.

The Asiatic part of Russia shows a much better correspondence between per capita resource extraction and wages. In particular, the regions of western Siberia and the northern Far East are characterized by relatively high wages and per capita resource output. The regions of eastern Siberia, in particular the biggest regions of Sakha and Krasnoyarsk Krai, as well as a number of regions of southern Siberia and the southern Far East show full correspondences between wage and per capita resource extraction. Finally, the vast majority of other regions of Asiatic Russia are characterized by modest correspondence between wage and per capita resource output.

If the resource map and the age map are compared, the picture is fully different. The majority of regions with their most ancient Russian cities extract almost none of the resources. At the same time, just those regions that differ from others by their relatively high levels of per capita resource output, in the extreme north and south, are among those with the youngest cities. In some cases the same pattern is observed within regions, as it is across regions. In particular, in Kursk Oblast (Figure 2) the two youngest cities, Kurchatov and Zheleznovodsk, are the richest cities of the region, and it is these cities which are featured by outstanding economic activities in extractive and/or neighboring industries. Kurchatov was founded on the building of a nuclear power plant (i.e., an establishment that since relatively recently became a source of well-being of its location, and at the same time, is that with a constraint resource.) Zheleznovodsk lives off the extraction of iron ore.[12] In the Asiatic part, western Siberia and the northern Far East as a whole feature relatively young and resource-rich cities. There are also regions with relatively low levels of resource extraction and relatively old cities.

Thus, when using a 5-level scale of wage, age, and per capita resource extraction, the vast majority of Russian regions have their wage and per capita resource output at or near the same levels. At the same time, the age of the cities both in European and Asiatic Russia as a whole is higher in the relatively resource-poor regions. We conclude that on the regional level per capita resource extraction is a close covariate of wage in that it correlates positively with wage and negatively with age.

7.2. Results for the static hypothesis

A visual inspection of the associations between the per capita resource extraction and age and wage gives a first impression about regional disparities with respect to these variables. To achieve more reliable results, we estimated a number of regressions using our two alternative age variables and controlling for a full set of regional dummies, spatial lag of the resource dependent variables, and a number of urban characteristics, i.e., the same controls as those used when estimating our base-line regressions. The results are presented in Table 8.

Columns 1 and 2 contain the results for fuel resource extraction per capita as a dependent variable, and columns 3 and 4 are for total resource extraction per capita. Data about the former are available for 1970-2004, which in our case means that relevant available data are for 1991-2004; data about the latter dependent variable are available for 2005-2011. For the whole period, 1991-2011, for establishing the negative association between age and wage and, hence, testing the hypotheses concerning the association, we have to use two dependent variables, differing from each other by what is included in the resource extraction. Apart from the dependent variable used, the regressions in the first two columns differ from those in the last two columns by the controls included. This was again motivated by the available data. For the period 1991-2004 we used the same four geographical controls, which are actually time-invariant, and six social and demographic controls we included in our base-line regressions. For the period 2005-2011 we added four variables: the share of firms in all organizations, the unemployment rate, and per capita outputs in the energy and the manufacturing industries. Columns 1 and 3 are for random effects regressions without the spatial lags, while columns 2 and 4 are for the regressions with the spatial lags. Finally, Panels A and B differ from each other by age variable as a key regressor.

[12] This resource is not of an outstanding market value though, but it suffices to push up relative per capita income within the region.

Table 8. Extractive output and the age

	1991-2004 log fuel output per capita		2005-2011 log ext. output per capita	
	(1)	(2)	(3)	(4)
	Panel A			
1st age variable (log)	-0.1507***	-0.0856***	-0.2905***	-0.3087***
	[0.0292]	[0.0118]	[0.0639]	[0.0777]
Log f. spatial lag		0.0949**		
		[0.0451]		
Log e. spatial lag				0.6017***
				[0.0982]
Obs.	13,310	7,098	6,064	4,781
Ids.	1,057	507	989	683
Regional and time FE	Yes	Yes	Yes	Yes
R-sq within	0,0934		0,0165	
R-sq between	0,4627		0,4621	
R-sq overall	0,3185		0,4313	
R-sq adj. (Buse 1973)		0.235		0.925
	Panel B			
2nd age variable (log)	-0.1215***	-0.0837***	-0.2479***	-0.2300***
	[0.0264]	[0.0110]	[0.0572]	[0.0700]
Log f. spatial lag	[0.0264]	[0,0110]	[0,5711]	[0,0700]
		0.1267***		
Log e. spatial lag		[0.0434]		
				0.5863***
				[0.0985]
Obs.	13,310	7,098	6,064	4,781
Ids.	1,057		989	
Regional and time FE	Yes	Yes	Yes	Yes
R-sq within	0,0935		0,0164	
R-sq between	0,459		0,4613	
R-sq overall	0,3165		0,4299	
R-sq adj. (Buse 1973)		0.241		0.925

Standard errors in brackets. *** p<0.01, ** p<0.05, * p<0.1

From the outset, in all the specifications there is a highly significant negative link between the key regressor and a dependent resource variable. This result firmly supports the visual negative link between age and per capita resource output that was established from the maps. For all that, now the result turns out to be robust to age definition, time period used, the resource dependent variable, regional affiliation, dependent variable spatial lag, year-specific factors, many other relevant controls. To show that the inclusion of the spatial lag and use of the respective estimation methods was reasonable, we report results for the spatial lag. In all the specifications with the spatial lag it is expectedly a significantly positive factor for per capita resource extraction. Hence, we captured both the spatial dependence in the resource extraction and a number of potentially omitted variables that correlate with the spatial lag. The absolute values of the negative slope coefficients on age variables increase from the regressions for 1991-2004 to those for 2005-2011. The reason may be related to non-energy resources being more rapidly get exhausted and to differences between the periods for the age effect.

Our static hypothesis would be supported if per capita resource extraction were negatively linked to age and if it were positively linked to wage. The negative link between the resource and age variables is established and presented in Table 8. To check if the second link holds, we run a number of fixed effects regressions that now are possible because in this case the key variables are time-varying. We estimated four fixed effects regressions differing from each other by the resource variable and the presence of the spatial lag in the list of controls. Among the latter, all time-varying controls that were used in our base-line regressions and in the regressions of the resource variables on age were included. The results are presented in Table 9.

In all the specifications, the results are in favor of our hypothesis. Per capita resource extraction turns out to be a highly significant factor for log wage. This is true whether we use the general resource variable or only the energy resource variable as a key regressor; whether we use the latest years, 2005-2011, or 1991-2004 depending on the available data; whether we use log wage spatial lag or not. The positive link between log per capita resource output and log wage is a result of obtaining consistent estimates from fixed effects regressions, so that the link is established from the within variation. Finally, the link is robust to a number of the key time-varying controls. It can be concluded, therefore, that per capita resource extraction is an unambiguously positive factor for wage in Russia. This result along with the previous

Table 9. Extractive output and wage, fixed effects regressions

	log wage 1991-2004		log wage 2005-2011	
	(1)	(2)	(3)	(4)
Log fuel output per capita	0.0216***	0.1058***		
	[0.0031]	[0.0056]		
Log ext. output per capita			0.0121***	0.0104***
			[0.0024]	[0.0026]
Log wage spatial lag		-0.0120		0.5347***
		[0.0225]		[0.0270]
Obs.	13,225	7,098	6,066	4,781
Ids.	1,070		990	
Regional and time FE	Yes	Yes	Yes	Yes
R-sq within	0.9947		0.9438	
R-sq adj. (Buse 1973)		0.983		0.973

Standard errors in brackets. *** p<0.01, ** p<0.05, * p<0.1

result leads to the same conclusion as that drawn from the maps, namely, that log per capita resource output goes in the same direction as log wage and in the opposite direction to age. In other words, our static hypothesis, which requires opposite signs of a_1 and b_1 in (29) and (30) is strongly supported.

7.3. Results for the dynamics hypothesis

Our dynamic hypothesis predicts that the force driving the data-generating process is to show similar dynamics on the part of different age groups of the cities with that shown up by the relative wages. We saw that relatively young cities are richer, but their relative wage advantages fall with time, and the greater the advantage, the more rapid its decrease. The model describes a mechanism related to per capita resource extraction that could generate these dynamics. Now we should check if this mechanism is actually at work in our data. To do this, we run several regressions such as (32) and (33). Based on the second age variable, we constructed two age dummy variables, for the cities that were given city status not earlier than 1985 and for those given the status between 1970 and 1985.

When choosing these two periods for constructing the dummies we started from some historical considerations. The latter year marks the beginning of Perestroika which was the preliminary period of post-Communism (which we use in our analysis). The next youngest group of cities was placed in the fifteen year period, being roughly the Stagnation Era preceding Perestroika and marked by most intensive development of extractive industries, including energy-related ones reinforced by the oil shocks of 1970s. To check the robustness of our results, we run the same regressions with age dummies based on the first age variable. In this case we chose fifteen year periods thirteen years earlier, so that the first age dummy was for the cities founded not earlier than 1972 and the second one was for those founded between 1957 and 1972. Such a shift in time was chosen based on the average difference between the year of granting city status and the foundation year. We constructed the age dummies based on the first age variable so that the starting and ending years of the period plus the average difference would give 1970-1985.

Thus, when using the second age variable our first age dummy was for the cities that were given their city status not earlier than 1985, and the second age dummy was for those given city status from 1970 to 1984 inclusive. The first group contained 86 cities, and the second one contained 51 cities. When excluding observations with missing values for wage for some years during 1991-2011, the first and the second groups reduce to 20 and 41 observations, respectively. When using the first age variable our first age dummy was for the cities founded not earlier than 1972, and the second one for those founded from 1957 to 1971 inclusive. In the first and second age groups there were 44 and 56 observations, respectively; and after dropping observations with missing values for wage for some years during 1991-2011, the groups reduced to 14 and 36, respectively. Finally, when dropping observations with missing values for some years during 1991-2011 for wage or any of the geographical, social, or demographical controls, the groups reduce to 8 and 19, and 4 and 13, respectively. Thus, we have enough observations by age groups to test the dynamic hypothesis when estimating regressions on an unbalanced panel, but they are quite small if we are forced to use strictly balanced panel.

To test the dynamic hypothesis, we run several panel regressions differing from each other by dependent variables,

Table 10. Dynamics of relative wage and per capita extractive variables by age groups

	Relative wage 1991-2011		Relative fuel output 1991-2004		Relative ext. output 2005-2011	
	(1)	(2)	(3)	(4)	(5)	(6)
	Panel A					
Age1: foundation year>= 1972	42.8490***	34.6822***	872.3918***	-133.1013	4,005.0966***	4,304.5428***
	[5.1698]	[12.8326]	[208.4995]	[170.2137]	[205.4688]	[460.2430]
Age2: foundation year >=1957 & <1972	24.3965***	3.9296	274.3037**	25.1025	962.8715***	1,212.0179***
	[3.4253]	[7.1846]	[137.1582]	[95.3602]	[151.1505]	[338.0406]
Age1*year	-0.0212***	-0.0171***	-0.4309***	0.0683	-1.9909***	-2.1406***
	[0.0026]	[0.0064]	[0.1044]	[0.0852]	[0.1023]	[0.2292]
Age2*year	-0.0121***	-0.0019	-0.1364**	-0.0130	-0.4788***	-0.6032***
	[0.0017]	[0.0036]	[0.0686]	[0.0477]	[0.0753]	[0.1683]
Rel. wage spatial lag		0.0158				
		[0.0228]				
Rel. fuel spatial lag				0.2512***		
				[0.0564]		
Rel. ext. spatial lag						0.0993**
						[0.0433]
Obs.	20,014	10,647	13,343	7,098	6,400	5,607
Ids.	1,081	507	1,067	507	1,003	801
Regional and time FE	Yes	Yes	Yes	Yes	Yes	Yes
R-sq overall	0.631		0.201		0.293	
R-sq adj. (Buse 1973)		0.488		0.0827		0.275
	Panel B					
Age1: city status>= 1985	28.7874***	-1.1888	926.8293***	159.7937	2,002.3479***	2,470.7634***
	[3.7067]	[9.1444]	[152.8803]	[120.9031]	[139.2756]	[321.3081]
Age2: city status >=1970 & <1985	14.0859***	19.0464***	301.7619**	222.4411***	478.8459***	450.1819
	[3.3171]	[6.0211]	[130.2340]	[79.6942]	[151.7043]	[337.1603]
Age1*year	-0.0143***	0.0007	-0.4614***	-0.0794	-0.9964***	-1.2297***
	[0.0019]	[0.0046]	[0.0765]	[0.0605]	[0.0694]	[0.1600]
Age2*year	-0.0069***	-0.0095***	-0.1499**	-0.1107***	-0.2377***	-0.2238
	[0.0017]	[0.0030]	[0.0652]	[0.0399]	[0.0756]	[0.1679]
Rel. wage spatial lag		-0.0265				
		[0.0221]				
Rel. fuel spatial lag				0.2350***		
				[0.0749]		
Rel. ext. spatial lag						0.0524
						[0.0431]
Obs.	20,014	10,647	13,343	7,098	6,400	5,607
Ids.	1,081	507	1,067	507	1,003	801
Regional and time FE	Yes	Yes	Yes	Yes	Yes	Yes
R-sq overall	0.630		0.194		0.272	
R-sq adj. (Buse 1973)		0.485		0.0808		0.260

Standard errors in brackets. *** p<0.01, ** p<0.05, * p<0.1

the age variable as a source of the age dummies, and the inclusion of the spatial lag. Recall that we have data on the two resource variables for 1991-2004 and 2005-2011, respectively, while data on wage are available for 1991-2011. So, we run two regressions for relative wage (32) differing in that one does not include the dependent variable spatial lag and uses unbalanced panel for 1991-2011, while the other specification includes the spatial lag and thereby uses strictly balanced panel for the same period. Also we run two regressions of relative per capita energy resource output and two regressions of relative total per capita resource output (33) with the same differences. As indicated in (34), key variables for testing the dynamic hypothesis are the interaction terms of the trend factor and age dummies. In particular, the slope coefficients on the interaction terms with both age dummies are to be negative, and the coefficient on the interaction with the younger age dummy is to be lower than that for the interaction with older group. At the same time, this regularity is to be observed both for the relative wage and the relative per capita resource extraction. This would mean that both age groups differ from the other cities in their dynamics as to the dependent variables, namely, their relative wage and per capita resource output are decreasing, and in case of the younger group this relative decrease is more rapid. To obtain consistent estimates for the interaction terms, we run fixed effects regressions without the spatial lag and, naturally, time-constant controls. At the same time, to repeatedly check if the static hypothesis as it is defined in (35) is supported in this specification, we run the same specifications by random effects with the spatial lag and time-constant controls. In the latter case we can estimate the effect of the age dummies. The results of estimating the efficient specifications are presented in Table 10.

The columns of Table 10 contain estimation results for the regressions differing from each other by the dependent variable and the inclusion of the dependent variable spatial lag. Columns 1 and 2 are for the relative wage as a dependent variable; columns 3 and 4 are for the relative per capita energy resource output; columns 5 and 6 are for the relative extractive output. Each pair of columns is for panel regressions without and with the spatial lag. Finally, Panels A and B differ from each other by what age variable underlies the age group dummies, viz., the first age variable was used in regressions presented in Panel A and the second age variable was used in Panel B.

In all the specifications without lag there are highly significant estimates on all the dummies and their signs and the relationships are strictly in line with the predictions of the static and dynamics hypotheses. Younger cities earn relatively more and their relative per capita extractive output or, specifically, relative energy resource output per capita are relatively higher, which is predicted by the static hypothesis as stated in (35). At the same time, the relative wage and the relative per capita extractive output, including energy resources, of younger cities are falling with time more rapidly, i.e., the prediction from the dynamic hypothesis stated in (34) is fulfilled.

As to the specifications with the spatial lags, in five of the six specifications at least one of the two age dummies and one of the two interaction terms are significant with the expected signs and relationships between their coefficients. In column 6 of Panel A all the dummies are significant and between the coefficients are the links predicted by the hypotheses. In column 4 of the same Panel all the dummies are insignificant. The reason for the insignificance of some dummies is obviously related to the small number of observations for the age groups when using the spatial lag and thereby dropping all observations with missing values of both dependent variables and the controls for any year of a panel. We can see that when running the spatial lag specification the number of observations drops dramatically. When regressing the relative wage and the relative energy resource variable this number is halved. No wonder that it is when regressing the extractive variable in which case the number of observations drops to a lesser extent after removing the missing values that all the estimates are significant in the spatial lag specification. This is seen only in one of the Panels, which is again explained by the number of observations. Recall that after dropping the missing values the number of observations in the age groups decrease less when the age dummies are based on the first variable. This difference in the remaining observations in the age groups after dropping the missing values turned out decisive for obtaining significant estimates. The results for the spatial lag of relative wage do not show a significant effect, unlike the regressions of log wage. The reason obviously is that the link between wage and its spatial lag is quite non-linear, which is captured by the log specification.

Nevertheless, as a whole, even these results of the spatial lag specifications rather support both the static and the dynamic hypotheses. In all but one specifications, at least, one of the younger age groups displays a higher relative wage and per capita extractive output and more downward dynamics compared with the other cities. It means that cities founded after 1956 or given city status after 1971 are generally richer and extract more resources per capita than other cities, but their wage and resource advantages are decreasing with time.

Our main interest is in testing the dynamic hypothesis. The interaction variables are of our concern. The latter are time-varying, so that we can obtain consistent estimates of their effect by running fixed effects specifications. The results of the consistent estimation are presented in Table 11. In almost all the specifications the interaction terms are highly significant and have signs and relative values fully in line with the dynamic hypothesis. In all cases the youngest cities display the most downward dynamics in both their relative wage and their per capita resource and energy resource extraction; the second youngest cities show downward dynamics for the same relative variables too, but in this case more modestly. Fixed effects specifications rule out that the established statistical regularity is due to some time-constant differences across the cities, and inclusion of a number of social, demographic, and economic variables allows us, at least in part, to control for important time-varying characteristics that could be responsible for the observed relative dynamics. The results are robust to changing the age variable underlying the age dummies construction. They are also robust to varying the time period used for estimating the regressions. As to relative wage, the dynamics predicted by the hypothesis are observed over the whole period 1991-2011; for relative per capita energy resource output the predicted dynamics are observed for 1991-2004; and for the relative per capita extractive output the same relative dynamics are observed for 2005-2011.

Table 11. Dynamics of relative wage and per capita extractive variables by age groups, fixed effects regressions

	1991-2011	1991-2004	2005-2011
	Rel. wage	Rel. per capita energy resource output	Rel. per capita extractive output
	(1)	(2)	(3)
	Panel A		
Year*Age1: foundation year>= 1972	-0.0242***	-0.4351***	-1.9746***
	[0.0025]	[0.0911]	[0.1022]
Year*Age2: foundation year >=1957 & <1972	-0.0118***	-0.0925	-0.4744***
	[0.0017]	[0.0688]	[0.0757]
Obs.	20,098	13,444	6,403
Ids.	1,090	1,078	1,004
Time FE	Yes	Yes	Yes
R-sq within	0.0194	0.0894	0.0728
R-sq between	0.131	0.111	0.123
R-sq overall	0.100	0.0306	0.0959
	Panel B		
Year*Age1: city status>= 1985	-0.0162***	-0.3974***	-0.9934***
	[0.0018]	[0.0720]	[0.0696]
Year*Age2: city status >=1970 & <1985	-0.0070***	-0.1434**	-0.2375***
	[0.0017]	[0.0650]	[0.0759]
Obs.	20,098	13,444	6,403
Ids.	1,090	1,078	1,004
Time FE	Yes	Yes	Yes
R-sq within	0.0171	0.0902	0.0405
R-sq between	0.0976	0.0691	0.0504
R-sq overall	0.0682	0.0211	0.0451

Standard errors in brackets. *** p<0.01, ** p<0.05, * p<0.1

8. Discussion

8.1. The increasing returns hypothesis

Using dummies for various ages, trend year variable and their interactions provided the differences-in-differences estimator to firmly establish the impact of year-to-year aging on the relative wage and per capita extractive output. The interest variables here are the interaction terms, and the estimation results presented in Tables 10 and 11 strongly support the dynamic hypothesis. Likewise, the respective results presented in Tables 8 and 9 unambiguously support the static hypothesis. Average wages are higher in younger cities and the same is true for per capita extractive outputs. At the same time, younger cities display downward dynamics in terms of both their relative wages and per capita extractive outputs. Our results suggest that younger cities are relatively richer but their wage differential shrinks more rapidly than for older cities. And the same is true for relative per capita extractive outputs. Thus, the younger a city, the higher its average wage and its per capita extractive output, and the more downward its relative dynamics of wage and per capita extractive output. These results are robust to changing the definition of city age, the time period considered, the inclusion of a full set of regional dummies, the dependent variable spatial lag, time fixed effects, many urban time-varying and time-invariant characteristics, and, in the case of using within estimator, city fixed effects. That the respective extractive measures follow closely the static and dynamic relationship between age and wage is a firmly established fact. We observe strong positive links between the extractive measures and wage, and strong negative links between age and wage, and between age and the extractive measures. And the relative dynamics of wage and the extractive measures by age groups are fully the same.

When beginning to explore the negative link between age and wage we treated it as a puzzle in context of the New Economic Geography. The latter emphasizes the importance of the agglomerative effects for spatial income distribution meaning that more densely populated areas, other things being equal, perform better and thereby earn more. At the same time, it is acknowledged that older locations are usually more populated because population concentration is a time-consuming process. We were prompted to ask why in Russia older cities are substantially poorer.

One of the potential answers is that for a long time Russia performed as a planned economy owing to which increasing returns effects are irrelevant for its spatial economic patterns. Spatial input allocation and income distribution were determined by the decisions of the planning officials rather than market forces, so that regularities typical for developed economies as to demographic measures, city age and wage might not work. We ran a number of regressions to estimate the link between age and various demographic measures such as log population size, population density, net migration, as well as the composition of the population with respect to percentages of labor, firms, and students. Among these six measures five are strongly positively linked with age. Older cities are bigger in terms of population size and density. This fact is just in line with the quotation from Davis and Weinstein (2002). Older cities have more time to accumulate population which in turn results in higher density as well. Further, older cities face more inflows of immigrants, so that their relatively large sizes are reinforced.

It could also be supposed that the positive relationship between age and population is explained by very old cities in the sample which have become very populous for their long history. However, the results for population are basically the same for the subsample restricted to cities not older than 80, i.e., those founded after the Soviet industrialization. For any subsample older cities are larger in terms of population size, population density and the number of immigrants.

For the composition of the population, older cities have more relative physical and human capital when they are measured by relative quantities of firms and college students. Their population composition is more advantageous from the standpoint of the share of the owners of physical and human capitals; at the same time, they are to maintain relatively more pensioners. Expectedly, labor share is the only exception where there is a negative link with city age. The labor share is a reflection of the fraction of retired people. The latter are more numerous in older cities. It is natural given the time during which cities could accumulate aging population. A smaller share of aged people in younger cities means a larger share of the remaining age groups including able-bodied people. Finally, as much evidence suggests (e.g., Henderson 2005; Gautier and Teulings 2009; Mulligan et al. 2012), more populated cities are featured by lower unemployment which is also positively correlated with average wage. Russian cities do not depart from this general regularity, and there is a strong negative association between city age and unemployment rate.

It could also be supposed that because of the Soviet legacy the demographic measures become poor predictors of the relative well-being of Russian cities. Or, one might suppose that agglomerative forces were negatively linked with city age. To check these links, we estimated short regressions of log wage on all the demographic measures we used previously as dependent variables. The results are unambiguously in favor of the agglomerative effects. There are strong positive relationships between log wage and a measure of city size such as log population size or density, net migration, and the measures of relative quantities of production factors, viz., able-bodied population, firms, and students. These results reveal simple positive associations. There are also estimation results from long regressions of log wage presented in Table 4 where the controls include all these demographic measures. What variables among them are highly significantly positively linked with log wage in all the specifications are log population, labor percentage and firm percentage. Unemployment rate is significantly negatively associated with log wage in all the specifications. In two specifications net migration is a significantly positive predictor of log wage. Some demographic measures, viz., population density and the share of students are insignificant in all the specifications, which is explained by these variables being close covariates of log population.

Generally, these results confirm the increasing returns hypothesis when it comes to the city age effect (Davis and Weinstein 2002, p. 1270) and agglomeration effect, i.e., older cities strongly tend to be more populated, to attract immigrants and to have higher shares of owners of physical and human capitals, or just employed persons; at the same time, all these measures are closely associated with average wage. Older cities are more populated and thereby, other things being equal, provide people with more employment and business opportunities, a better match between skills and jobs, and higher earnings. All these relationships mean that, despite the ongoing distorting effect of the Soviet past, the relationships between city age and the demographic measures as well as between the latter and per capita income are the same as those in developed market economies. The implications from these patterns in Russian data are twofold. First, the increasing returns hypothesis does fit the data on Russian cities, and second, it is this fact that rules out that the negative link between age and wage results from the regularities predicted by the increasing returns hypothesis. The very fact that city age is simultaneously positively linked with positive agglomerative predictors of wage and negatively linked with wage means that the agglomerative effects are at work in Russia, but when it comes to the age effect there is a force that heavily outweighs the agglomerative effects. In other words, though the increasing returns work in Russian cities, a force behind the link of our concern is different.

8.2. The institutions hypothesis

As a number of empirical works of Acemoglu et al. (2001; 2002; 2010; and 2014) suggest, spatial income differentials both across and within countries result from the respective institutional differences. In large countries such as those in Americas there are substantial cross-region differences in quality of their institutions (2010). It is fair to suppose that Russia is also heterogeneous with respect to its institutions. The dataset at our disposal does not contain institutional variables similar to those used in the analysis of Acemoglu et al. (2002), such as protection against expropriation or constraints on the executive, and as proxies for the quality of institutions we used variables based on the available data. Specifically, we used the crime rate; expenditures for legal services per capita (available only for 1997); the proportion of firms among all organizations; and the fraction of doctors in the population. The first is treated as a proxy for rule of law; the second is as a proxy for the comparative share of the transaction sector measuring the quality of institutions (Wallis and North 1986); the third is as a proxy for the availability of various organizational environments for the private sector (Wittfogel 1957; North et al. 2009); and the fourth is a proxy for the quality of public goods provision (Enikolopov et al. 2011).

The pair correlations between our institutional variables and wage are in line with the institutions hypothesis. All the institutional variables are positively correlated with log wage. The estimation results of the long regressions for the fraction of doctors and relative quantity of firms presented in Table 4 lead to the same conclusions. The proportion of firms among all organizations is highly significantly positively linked with log wage in all the specifications, and so is the fraction of doctors in population, but this is not robust to restricting the sample to cities with population size above 12,000. The results of estimating long regressions for crime rate and the expenditures for legal services still suggest that these variables are positively linked with log wage, though they are robust to the inclusion of neither regional dummies, nor geographical characteristics, nor other important controls. When it comes to the associations of these institutional variables with city age, crime rate is negatively linked, while the others are positively linked with city age.

The most robust estimates are for the positive links between the proportion of firms and wage, and between the firms proportion and age. These results suggest that older cities usually feature a bigger proportion of firms, while the latter strongly positively predicts wage. For the fraction of doctors the results are similar, though they are not robust to the exclusion of small towns. As for the other proxies of institutions, here the results are mixed and unreliable. Public goods provision and organizational opportunities for the private sector are better established in older cities that is positively correlated with their wellbeing. These regularities can be explained by the institutions hypothesis – better public goods provision and more organizational opportunities are conducive to better economic performance. The same regularities can be explained by the increasing returns hypothesis as well, in which case one would argue that public goods provision and the relative number of firms are just positive covariates of population size that in turn is strongly correlated with agglomerative effects. Whatever force underlies these regularities, the key fact here is that older cities are featured by more advantageous characteristics. And institutional differences, like those related to agglomeration effects, seem to play a role in making income differentials as the institutions hypothesis implies. However, what implication by no means follows from this hypothesis coupled with our evidence is that older cities in Russia have worse institutional features. Hence, the negative link between age and wage is not based on institutional differences. Again we can argue that older cities have more advantageous characteristics in many respects, including those implied by the increasing returns and institutions hypotheses, but there is a force that outweighs them.

8.3. The geography hypothesis

The established regularities for city age, wage, and other urban characteristics suggest systematic change of spatial income distribution. The geography hypothesis in its simple version relies on locational fundamentals. These time-constant factors in making spatial income distribution lead to persistent spatial patterns, like those in Japan throughout its long history

(Davis and Weinstein 2002). Its main prediction does not fit the facts of our case. However, the locational fundamentals or other, less fundamental geographic characteristics coupled with human activities are already time-varying characteristics. And, as follows from the preceding analysis, they underlie these regularities. This means that what works in this case is the geography hypothesis in its sophisticated version. To identify the force behind the link of our interest, a quasi-experiment was used. The results highlight resource depletion. New cities follow yet new resources deposits, and their depletion makes these cities poorer as time goes by. As already existing cities are getting relatively poorer due to their resource depletion, new cities begin in the sites of new resources that provide these cities with temporary economic advantages.

From a more general perspective, the process of city creation and evolution fits the long-standing pattern of territory development in Russia. As Kluchevsky (1911; see also Pipes 1993) put it, moving to yet undeveloped territories has always been most typical strategy of the Russian population and state as a response to various challenges. A near analogy here is slash-and-burn agriculture on the part of most peasants in medieval Russia. This type of agriculture in the Russian woods admitted rich harvests from new-plowed soil and their rapid downward annual dynamics. When the harvests became low enough, the respective peasant community moved to another place. Naturally, this pattern of using territory was possible, given the large empty spaces.[13] Apart from developing virgin soils, people were induced by the search for other resources, such as valuable furs called "soft gold"; striving for more freedom or; as was the case with numerous religious enthusiasts, for solitude. In all the cases the state followed its subjects, resulting in the expansion of controlled territories or developing once captured ones. Over the Soviet era the state used the vast space to retreat during the war or to deal with economic difficulties in peace time.

As Schumpeterian creative destruction suggests, in the long run, technological progress leads to a change in the relative importance of various resources and thereby the relative economic value of the respective territories. In the short run, resource deposit depletion in some territories and their development in others can also lead to a change in the spatial income distribution. In Russia examples of the former are cities founded near rich salt deposits such as Usolye-Sibirskoye, Solvychegodsk, Solikamsk, Sol-Iletsk, which, though some of them still contain abundant stocks of salt, no longer benefit from them because of the conditions of salt trade. Examples of the latter are territories that declined and in some cases became deserted because of depletion of their coal, oil, or other resource deposits.

Our results, supporting the sophisticated geography hypothesis, also echo some predictions from the field of international trade, in particular, the "new new" trade hypothesis. The underlying theory highlights an additional gain from free trade, which is related to fiercer competition between firms differing from each other in their productivity. It implies that higher productivity and thereby higher wages are a feature of exporting firms (Melitz 2003) and even more so of those investing abroad (Helpman et. al 2004). There is rich empirical evidence that exporting firms and those investing abroad are more productive compared to those failing to export and/or invest abroad. However, exporting firms may pay higher wages even under a given productivity (Helpman et al. 2010). There is also evidence that exporting and foreign-controlled firms pay more because of other characteristics (Breau and Brown 2011). The export activities of former socialistic countries tend to be driven by their comparative advantages (e.g., those related to factor endowments as the Heckscher-Ohlin theory suggests) to a greater extent than by the home market effect or market potential implied by the increasing returns and new trade theories (Simone 2008).

The position of an exporting and/or investing abroad firm may correlate with some features other than productivity-related ones which lead to their wage premium. These features may be related to local factor endowments. With respect to the differences across regions, the implication is that regions with higher shares of exporting firms and firms investing abroad have higher positive wage differentials. It can be imagined that this is the case in Russia, given that the bulk of its income comes from exports and there are substantial cross-regions differences in terms of the export share in regional output. However, the substantial part of Russian exports is comprised of extractive produce, and in this case exporting does not positively correlate with productivity (Davis and Cordano 2013). Thus, though younger cities are featured by wage premiums correlated with their relatively high per capita extractive output, their potentially greater involvement in

[13] According to Braudel (1992), the same pattern of developing space was typical for other nations/tribes given abundant unoccupied territories.

exporting and/or investing abroad is not a result of their advantages in productivity. What prediction of the "new new" trade theory is realized in case of Russian cities is that exporting activities are correlated with relatively high mark-ups. Higher mark-ups are present in many extractive industries, but they are scarcity-related rents rather than the result of productivity differentials. Temporary resource abundance in newly developed locations provides them with temporary income advantages. The latter are translated into wage premiums and higher mark-ups that in turn enable firms located there to be involved in exporting activities.

9. Conclusion

Among Russian cities there is a strong statistical regularity that seems to be at odds with the established theory in spatial economics. Younger cities are substantially richer, despite their smaller population sizes and thereby weaker agglomeration forces. The Soviet legacy does not impede agglomeration forces, which is seen in the strong positive association between city age and population size and between the latter and wages across Russian cities. Hence, the negative association between city age and wage is a result of a mechanism more powerful than that related to agglomeration forces. To determine this mechanism, we discussed three well-known hypotheses for spatial income distribution: the increasing returns hypothesis, the institutions hypothesis, and the sophisticated geography hypothesis. To check whether a hypothesis explained the link between city age and wage, we analyzed associations between city age and other urban characteristics whereon the hypothesis places its main emphasis. In particular, we used a quasi-experiment, in which time was to affect various age groups of cities differently.

The results turned out in favor of the sophisticated geography hypothesis. This was used here in its exhaustible resource-related version. To go into details of the respective mechanism, a model was developed relating the dynamics of income in several locations with their population growth and the resource use. What was of the key importance, the latter were related to age which varied across the locations. The main predictions from the model were that relative per capita income and per capita resource output decreased with age; and that stronger downward dynamics of relative per capita income and per capita extractive output were to be observed in younger locations. These two groups of predictions comprised the static and dynamic hypotheses.

The key variables of our version of the sophisticated geography hypothesis very closely replicated the link with city age both in its static and dynamic variants. The reproduction of the static link was shown by the extractive output measures being a highly significant positive predictor of wage in Russian cities, and their being significantly negatively associated with city age. As for the dynamic link, various age groups of cities showed the same relative dynamics both for wage and the extractive output measures. In particular, younger groups of Russian cities were relatively richer in terms of their average wages and they displayed relatively higher per capita extractive output; for all that, younger groups of the cities displayed downward dynamics in terms of both their average wage and per capita extraction. Of the key importance here is that the relationships between the age groups in terms of their relative wages and per capita extraction as well as their relative downward dynamics were exactly the same. Among the three age groups we distinguished, the youngest one displayed highest values of both average wage and per capita extractive output, and the most rapid downward dynamics of its relative wage and per capita extraction; the second age group displayed similar regularities, but more moderately; for the remaining oldest group, these regularities meant a lack of any advantages in both their wage and the extractive output measures, and decrease in their relative dynamics in wage or per capita extraction. Thus, the differences-in-differences specification gives reliable support to the tested pattern, which suggests that the relative extractive production in younger cities is a source of their income advantage, but it is due to the depletion that it declines with time which in turn gradually undercuts their wage bonuses.

These results reveal a particular pattern of Russian urban development. Specifically, new cities occur in sites that have rich deposits of valuable resources. It provides new cities with temporary advantages with respect to their resource endowments and ensuing incomes. As time goes by, the resources become depleted and the respective advantages vanish.

From a broader perspective, this pattern corresponds to the way the Russian population and state used their vast territory throughout the history, moving on from resource depleted areas to new richer areas. While this pattern is at work in Russia, it can be helpful when examining the spatial development of other resource-rich countries and regions. With the account of this pattern there is the potentially important role of exhaustible resources in changing regional differences in economic activities and incomes. And over longer historical intervals this pattern suggests a potential effect of technological progress on the dynamics of spatial income distribution via changing relative values of various natural resources.

References

[1] Acemoglu, Daron (2009) Introduction to Modern Economic Growth. Princeton and Oxford: Princeton University Press.

[2] Acemoglu, Daron, Melissa Dell (2010) Productivity Differences Between and Within Countries. American Economic Journal: Macroeconomics, 2(1): 169-188.

[3] Acemoglu, Daron, Simon Johnson, James A. Robinson (2001) The Colonial Origins of Comparative Development: An Empirical Investigation. American Economic Review, 91(5): 1369-1401.

[4] Acemoglu, Daron, Simon Johnson, James A. Robinson (2002) Reversal of Fortune: Geography and Institutions in the Making of the Modern World Income Distribution. Quarterly Journal of Economics, 117(4): 1231-1294.

[5] Acemoglu, Daron, Simon Johnson, James A. Robinson (2005) The Rise of Europe: Atlantic Trade, Institutional Change, and Economic Growth. American Economic Review, 95(3): 546-579.

[6] Acemoglu, Daron, Suresh Naidu, Pascual Restrepo, James A. Robinson (2014) Democracy does Cause Growth. NBER Working Paper 20004.

[7] SSSR. Administrativno-territorialnoye delenie soyuznych respublik (1987) Moscow: Izvestiya Sovetov Narodnych Deputatov.

[8] Alexeev, Michael, and Robert Conrad (2009) The Elusive Curse of Oil. Review of Economics and Statistics, 91(3): 586-598.

[9] Anselin, Luc (1988) Spatial Econometrics: Methods and Models. Dordrecht: Kluwer Academic.

[10] Ayuda, Maria Isabel, Fernando Collantes, Vincente Pinilla (2010) Long-run Regional Population Disparities in Europe during Modern Economic Growth: a Case Study of Spain. Annual Regional Science, 44(2): 273-295.

[11] Bariev, I. K. (2007) Koordinaty naselennych punktov, chasovye poyasa i ismenenia ischisleniya vremeni. Spravochnik. Moscow: Konek.

[12] Breau, Sebastien, and W. Mark Brown (2011) Exporting, Foreign Direct Investment, and Wages: Evidence from the Canadian Manufacturing Sector. Growth and Change, 42(3): 261-286.

[13] Braudel, Fernand (1992) Civilization & Capitalism 15th-18th Century: Volume I. The Structures of Everyday Life. Oakland: University of California Press.

[14] Breinlich, Holger, Gianmarco I. P. Ottaviano, and Jonathan R.W. Temple (2014) Regional Growth and Regional Decline. Handbook of Economic Growth, 2(B): 683-779.

[15] Buse, A. (1973) Goodness of Fit in Generalized Least Squares Estimation. The American Statistician, 27(3): 106-108.

[16] Cleveland, W. S. (1979) Robust Locally Weighted Regression and Smoothing Scatterplots. Journal of the American Statistical Association, 74(368): 829-836.

[17] Combes, Pierre-Philippe, Thierry Mayer, and Jacques-Francois Thisse (2008) Economic Geography. The Integration of Regions and Nations. Princeton: Princeton University Press.

[18] Dasgupta, Partha, and Geoffrey Heal (1974) The Optimal Depletion of Exhaustible Resources. Review of Economic Studies, 41(Symposium): 3-28.

[19] Davis, Graham A., and Arturo L. V´asquez Cordano (2013) International Trade In Mining Products. Journal of Economic Surveys, 27(1): 74-97.

[20] Davis, Donald R., and David E. Weinstein (2002) Bones, Bombs, and Break Points: the Geography of Economic Activity. American Economic Review, 92(5): 1269-1289.

[21] Dell, Melissa, Benjamin F. Jones, and Benjamin A. Olken (2012) Temperature Shocks and Economic Growth: Evidence from the Last Half Century. American Economic Journal: Macroeconomics, 4(3): 66-95.

[22] Dixit, Avinash K., and Joseph E. Stiglitz (1977) Monopolistic Competition and Optimum Product Diversity. American Economic Review, 67(3): 297-308.

[23] Dobkins L. Harris, Yannis M. Ioannides (2001) Spatial Interactions among U.S. Cities: 1900-1990. Regional Science and Urban Economics, 31(6): 701-731.

[24] Dobrynskaya, Victoria, and Edouard Turkisch (2009) Is Russia Sick with the Dutch Disease? CEPII No. 20.

[25] Duranton, Gilles (2007) Urban Evolutions: the Fast, the Slow, and the Still. The American Economic Review, 97(1): 197-221.

[26] Duranton, Gilles and Diego Puga (2014) The Growth of Cities. Handbook of Economic Growth, 2(B): 781-853.

[27] Economika Gorodov Rossii (2013). GMZ Rosstata. http://multistat.ru/?menu_id=9310004 (accessed April 1, 2013).

[28] Egorov, Georgy, Sergei Guriev, and Konstantin Sonin (2009) Why Resource-Poor Dictators Allow Freer Media: a Theory and Evidence from Panel Data. American Political Science Review, 103(4): 645-68.

[29] Enikolopov, Ruben, Maria Petrova, and Ekaterina Zhuravskaya (2011) Media and Political Persuasion: Evidence from Russia. American Economic Review, 101(4): 3253-3285

[30] Florida, Richard (1996) Regional Creative Destruction: Production Organization, Globalization, and the Economic Transformation of the Midwest. Economic Geography, 72(3): 314-334.

[31] Gabaix, Xavier, and Yannis M. Ioannides (2004) The Evolution of City Size Distribution. Handbook of Regional and Urban Economics, 4: 2341-2378.

[32] Gautier, Pieter A., and Coen N. Teulings (2009) Search and the City. Regional Science and Urban Economics, 39(3), 251-265.

[33] Giesen, Kristian, and Jens Suedekum (2012) The Size Distribution Across All "Cities": A Unifying Approach. SERC Discussion Paper No. 122.

[34] Glaeser, Edward L. (2011) Triumph of the City. How Our Greatest Invention Makes Us Richer, Smarter, Greener, Healthier, and Happier. New York: Penguin Press.

[35] Glaeser, Edward L., and Matthew E. Kahn (2001) Decentralized Employment and the Transformation of the American City. NBER Working Paper 8117.

[36] Glaeser, Edward L., Stuart S. Rosenthal, and William C. Strange (2010) Urban Economics and Entrepreneurship. Journal of Urban Economics, 67(1), 1-14.

[37] Hartwick, John M. (1977) Intergenerational Equity and the Investing of Rents from Exhaustible Resources. American Economic Review, 67(5): 972-974.

[38] Hartwick, John M. (1978) Increasing Returns, Exhaustible Resources, and Optimal Growth. Economics Letters, 1(3): 231-235.

[39] Helpman, Elhanan, Oleg Itskhoki, and Stephen Redding (2010) Inequality and Unemployment in a Global Economy. Econometrica, 78(4): 1239-1283.

[40] Helpman, Elhanan, Marc J. Melitz, and Stephen R. Yeaple (2004) Export versus FDI with Heterogeneous Firms. American Economic Review, 94(1): 300-316

[41] Henderson, J. Vernon (2005) Urbanization and Growth. Handbook of Economic Growth, 1: 1544-1591.

[42] Hornbeck, Richard (2012) The Enduring Impact of the American Dust Bowl: Short- and Long-Run Adjustments to Environmental Catastrophe. American Economic Review, 102(4): 1477-1507.

[43] Hotelling, Harold (1931) The Economics of Exhaustible Resources. Journal of Political Economy, 39(2): 137-175.

[44] Hounshell, David (1984) From American System to Mass Production, 1800-1932. Baltimore: Johns Hopkins University Press.

[45] James, Alex, and David Aadland (2011) The Curse of Natural Resources: an Empirical Investigation of U.S. Counties. Resource and Energy Economics, 33(2): 440-453.

[46] Kapoor, Mudit, Harry H. Kelejian, and Ingmar R. Prucha (2007) Panel Data Models with Spatially Correlated Error Components. Journal of Econometrics, 140(1): 97-130.

[47] Kelejian, Harry H., and Ingmar R. Prucha (1998) A Generalized Spatial Two-Stage Least Squares Procedure for Estimating a Spatial Autoregressive Model with Autoregressive Disturbances. Journal of Real Estate Finance and Economics, 17(1), 99-121.

[48] Kluchevsky, Vasily. O. 1911. History of Russia. Vol. I. London: J. M. Dent & Sons, Ltd.

[49] Krugman, Paul R. (1991a) Increasing Returns and Economic Geography. Journal of Political Economy, 99(3): 483-99.

[50] Krugman, Paul R. (1991b) History versus Expectations. Quarterly Journal of Economics, 106(2): 651-67.

[51] Krugman, Paul R. (1993) First Nature, Second Nature, and Metropolitan Location. Journal of Regional Science, 33(2): 129-144.

[52] Kuznets, Simon (1955) Economic Growth and Income Inequality. American Economic Review, 45(1): 3-28.

[53] Lappo, Georgy M. (ed.) (1998) Goroda Rossii. Enziklopediya. Moscow: Bolshaya Rossiyskaya Enziklopediya.

[54] Lappo, Georgy M. (2006) Vozrast Gorodov. Geographia, 7-8.

[55] Lozada, Gabriel A. (1995) Resource Depletion, National Income Accounting, and the Value of Optimal Dynamic Programs. Resource and Energy Economics, 17(2): 137-154.

[56] Malthus, Thomas R. (1798) An Essay on the Principle of Population. London: J. Johnson.

[57] Marshall, Alfred (1890) Principles of Economics. London: Macmillan and Co., Ltd.

[58] Melitz, Marc J. (2003) The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity. Econometrica, 71(6): 1695-1725.

[59] Michaels Guy, Ferdinand Rauch, and Stephen J. Redding (2012) Urbanization and structural transformation. Quarterly Journal of Economics, 127(2): 535-586.

[60] Miller, Merton, and Charles W. Upton (1985) A Test of the Hotelling Valuation Principle. Journal of Political Economy, 93(1): 1-25.

[61] Mulligan Gordon F., Mark D. Partridge, and John I. Carruthers (2012) Central Place Theory and its Reemergence in Regional Science. Annales of Regional Science, 48: 405-431.

[62] Neary, J. Peter (2001) Of Hype and Hyperbolas: Introducing the New Economic Geography. Journal of Economic Literature, 39(2): 536-561.

[63] North, Douglas C., John Joseph Wallis, and Barry R. Weingast (2009) Violence and Social Orders: A Conceptual Framework for Interpreting Recorded Human History. Cambridge: Cambridge University Press.

[64] Partridge, Mark D. (2010) The Duelling Models: NEG vs Amenity Migration in Explaining US Engines of Growth. Papers in Regional Science, 89(3): 513-536.

[65] Pipes, Richard (1973) Russia under the Old Regime. New York: Charles Scribner's Sons.

[66] Rappaport Jordan, and Jeffrey D. Sachs (2003) The United States as a Coastal Nation. Journal of Economic Growth, 8(1): 5-46.

[67] Regiony Russii: Sozialno-ekonomicheskiye pokazateli (2012) Moscow: Rosstat.

[68] Ricardo, David (1817) Principles of Political Economy and Taxation. London: John Murray.

[69] Ross, Michael (2001) Does Oil Hinder Democracy? World Politics, 53(3): 325-61.

[70] Rozenfeld, Hernan D., Diego Rybski, Xavier Gabaix, and Hernan A. Makse (2011) The Area and Population of Cities: New Insights from a Different Perspective on Cities. American Economic Review, 101(5): 2205-2225.

[71] Sachs, Jeffrey D., and Andrew M. Warner (2001) The Curse of Natural Resources. European Economic Review, 45(4-6): 827-838.

[72] Schumpeter Joseph A. (1942) Capitalism, Socialism, and Democracy. New York: Harper & Row.

[73] Simone, Gianfranco De (2008) Trade in Parts and Components and the Industrial Geography of Central and Eastern European Countries. Review of World Economics, 144(3): 428-457.

[74] Solow, Robert M. (1974) Intergenerational Equity and Exhaustible Resources. Review of Economic Studies, 47(symposium): 29-45.

[75] Strobl, Eric (2011) The Economic Growth Impact of Hurricanes: Evidence from U.S. Coastal Counties. Review of Economics and Statistics, 93(2): 575-589.

[76] Vollrath, Dietrich (2011) The Agricultural Basis of Comparative Development. Journal of Economic Growth, 16(4): 343-370.

[77] Wallis, John Joseph, and Douglas C. North (1986) Measuring the Transactional Sector in American Economy, 1870-1970. Long-term factors in American Economic Growth. Vol. 51 of The Income and Wealth Series. Ed. by S. Engerman and R. Gallman. Chicago: University of Chicago Press.

[78] Wang, Chunhua, and JunJie Wu (2011) Natural Amenities, Increasing Returns and Urban Development. Journal of Economic Geography, 11(4): 687-707.

[79] Williamson, Oliver E. (1985) The Economic Institutions of Capitalism. Firms, Markets, Relational Contracting. New York: Free Press.

[80] Wittfogel, Karl A. (1957) Oriental Despotism: A Comparative Study of Total Power. New Haven and London: Yale University Press.

[81] World Development Indicators (2014) The World Bank. http://data.worldbank.org/products/wdi

Alexander S. Skorobogatov
National Research University Higher School of Economics
e-mail: skorobogat@mail.ru