



THE WILLIAM DAVIDSON INSTITUTE
AT THE UNIVERSITY OF MICHIGAN

Methodologies of Analyzing Inter-Regional Income Inequality and Their Applications to Russia

By: Konstantin Gluschenko

William Davidson Institute Working Paper Number 984
April 2010

Methodologies of Analyzing Inter-Regional Income Inequality and Their Applications to Russia^{*}

Konstantin Gluschenko

Institute of Economics and Industrial Engineering,
Siberian Branch of the Russian Academy of Sciences,
and

Department of Economics,
Novosibirsk State University

pr. Lavrentieva 17, 630090 Novosibirsk, Russia

E-mail: glu@nsu.ru

Home page: http://econom.nsu.ru/staff/chair_et/gluschenko/

Abstract

This paper provides an overview of methodologies used to analyze inter-regional income inequality, and a critical survey of empirical studies that deal with Russian regions. It discusses implications of the growth theory regarding dynamics of inter-economy income inequality. Methodologies for empirically analyzing income inequality are classified as the cross-section approach, time series approach, and distribution dynamics approach. Specific methodologies are described within the framework of this classification, touching upon the subject of their applicability fields. The survey of studies on income inequality among Russian regions summarizes more than 30 papers grouped according to main approaches used for the analyses.

JEL classification: C20, D31, O15, O18, O41, P25, R11, R15

Keywords: spatial inequality, convergence, economic growth, beta-convergence, distribution dynamics, income mobility, Russian regions.

* I am grateful to the Economics Education and Research Consortium (EERC) for financial support of attending its workshops (Odessa, Ukraine, June 27-29, 2008, Kiev, December 18-20, 2008, and June 27-29, 2009) as well as to participants to the workshops – in particular, Denis Nizalov, Irina Murtazashvili, and Alexander Shepotylo – for helpful comments. Special thanks to Evgeny Zhelobodko for fruitful discussions.

For the Russian version, see http://mpra.ub.uni-muenchen.de/18443/1/MPRA_paper_18443.pdf

1. INTRODUCTION

The problem of income inequality among regions within countries came to the attention of a great number of researchers. This problem is of particular importance for Russia because of great variability of its regions in their welfare. The literature provides a wealth of methodologies for analyzing inter-regional income inequality. (For the most part, they aim to answer the question of whether income convergence of regions takes place, i.e. whether the cross-region income distribution tends to become more equal.) However, this wealth not always is applied properly in empirical studies. Not infrequently, researches are uncritical in choice of a methodology of analysis, not taking account of its applicability field, which can lead to wrong conclusions. This paper aims at two purposes. The first is to provide an organic (although not exhaustive) overview of methodologies used to analyze inter-regional income inequality. The second purpose is to critically survey papers that deal with income inequality among Russian regions.

Methodologies being used to analyze income inequality among regions within counties are those originally developed for international studies. There are two opinions on this point. According to Barro and Sala-i-Martin (2004), regions are even more suitable matter of enquiry, as differences in technology, preferences, and institutions are likely to be smaller than those across countries. Magrini (2003) holds a different viewpoint that regions and countries are far from being interchangeable concepts because of much greater openness of regional economies. However, this implies not the negation of practicability to apply the same methodologies to international and inter-regional analyses but the necessity to take into account this feature of regions.

Specifically, this concerns indicators used to measure income. Depending on research objectives, a researcher can study inequality among countries or regions from either, so to speak, social or economic position. In the first case, researcher's interest is differentiation of the standard – or, more broadly, quality – of living across locations (here, studies on income inequality are adjacent to studies on other aspects of socio-economic inequality as well as poverty). In the second case, researcher's interest is differences in productivity (level of development) of economies of countries or regions. Here, the exploration of income inequality is adjacent to studies on economic growth (consisting in, e.g., checking consistency of predictions from one or another growth model with observed facts, that is, in testing growth models). In international studies, personal income per capita and GDP per capita are in both cases practically interchangeable empirical material. But this is not the case when one deals

with regions. Gross regional product (GRP) is far from being an adequate indicator of region's population welfare, if for no other reason than that a sizeable part of GRP can be consumed outside the region. On the other hand, personal incomes can inadequately measure productivity of the regional economy because of transfers from the national budget. For example, GRP per capita in the Chukchi Autonomous Okrug (a region of Russia) in 2003 was equal to 3.5 relative to the national average, while personal income per capita equaled 2.6, the relative GDP per capita having increased and personal income per capita decreased as compared to the previous year (author's calculations using data from Rosstat, 2008, pp. 157-158 and 352-353).

One more aspect concerning data is ensuring their comparability across space. This requirement is obvious for international analyses, and is rather simply realized by converting national indicators into a single currency with the use of official exchange rate or – what is more correct – purchasing power parity. But it is of no less importance for regions within a country despite a single currency, since the purchasing power of the national currency differs across regions. For example, the maximum to minimum ratio of the cost of a fixed basket of goods and services across Russian regions amounted to 2.6 as of the end of 2007, and that of an (unofficial) cost of living index across US cities in the fourth quarter of 2002 came to the same 2.6 (author's calculations using data from Rosstat, 2008, pp. 965-966, and ACCRA, 2002). It is thus apparent that while analyzing inter-regional income inequality, it is more correct to deal with real values, and not with nominal ones.¹

The term "real" with respect to some cost indicator means that it is denominated in monetary units with a uniform purchasing power. However, the sense of this term differs depending on whether comparability across time or across territories is meant. In the former case, the indicator is "cleaned" from inflation with the use of a cross-time price index, e.g., consumer price index. In the latter case, the indicator is "cleaned" from differences in prices between territories, which requires using spatial price indexes (the Geary-Khamis index, Eltetö-Köves-Szulc index, etc.; see, e.g., Diewert, 1999). In this paper, the term real is always used in reference to indicators that are comparable across territories. No one country enjoys official statistical data on regional costs of living (regional consumer price levels, spatial price indexes). Therefore researchers are forced to estimate real incomes, exploiting one or another available indicator (usually, the cost of some

¹ It should be noted that, generally speaking, analyzing nominal incomes per capita is not meaningless. It is thought that workers make an economically based decision to move to a different location, reasoning from comparisons of nominal, and not real, incomes across locations. However, the field of research in which nominal indicators matter is fairly small.

consumer basket) as a proxy of the regional cost of living. Lacking proper proxies, which is not a rare case in many countries, the researchers have no other way as to use nominal incomes under very strong assumption of approximately equal cost of living across regions.

In the case of GRP, the problem is even more severe. Being the added value produced in a region, GRP is seemingly comparable across regions of the country. But this is not the case. Suffice it to say that prices for building production with identical properties vary greatly across Russian regions. Thus, the nominal value of the relevant component of GRP gives no idea of the actual value added by construction in a region. Similar considerations can be applied to other components of GRP. Then the problem analogous to that appearing in international comparisons is arrived at: the problem of assessing “regional purchasing power parities.” This problem is much more difficult than that of estimating real incomes, as a single index that reflects inter-regional differences in prices is not sufficient for computing real GRP; each main constituent of GRP needs its own index. Granberg and Zaitseva (2002, 2003) put forward a methodology of assessing real GRP in Russia. The differences between nominal and real GRPs for 1999 proved to be very sizeable, varying across regions from -29% to $+33\%$ for the produced GRP, and from -52% to $+32\%$ for the used GRP (Granberg and Zaitseva, 2003, pp. 12 and 15). However, because of intricacy of such an assessment, researchers that use GRP in studying inter-regional inequality are forced, as a rule, to exploit nominal figures.

The openness of regional economies also requires explicitly taking into account interactions and interrelations of regions in theoretical and econometric models. This issue will be touched upon in the next two sections.

2. THE GROWTH THEORY AND INTER-ECONOMY INCOME INEQUALITY

The theoretical foundation of studies on dynamics of inter-regional income inequality is provided by the economic growth theory. As the literature on this theory is quite voluminous, no need would be to address it here if there were not one circumstance. Namely, a stereotype is fairly widespread among economists engaged in researches on spatial economic inequality that the growth theory predicts income convergence among economies. Apparently, it is due to Barro and Sala-i-Martin’s publications (in particular, those of 1992 and 2004) based on neoclassical growth models. However, the above implication does not follow even from these models.

Let us briefly discuss what these models suggest. Consider a closed one-sector economy characterized by a production function with exogenous labor-augmenting technical progress:

$Y(t) = F(K(t), L(t) \cdot A(t))$, where t is time, $Y(t)$ is total output (hereafter, income), $K(t)$ is physical capital, $L(t) = L(0)e^{vt}$ is labor (the number of workers), v is the rate of population growth, $A(t) = A(0)e^{\xi t}$ is the state of technology, and ξ is the rate of technical progress. The production function exhibits constant returns to scale and diminishing marginal productivity. All produced income $Y(t)$ is devoted to consumption and gross investment, so that $\partial K(t)/\partial t = sY(t) - \delta K(t)$, where δ is the depreciation rate of physical capital, and s is the saving rate. The latter is either exogenous like in the Solow-Swan model or is a solution of the social-welfare maximization problem like in the Ramsey model in its Cass-Koopmans version (then $s = s(t)$ in the general case, but the equilibrium value of s – which solely is of interest to us – is time-invariant). It is worth noting that the production function can also include human capital (as, e.g., in the formulation due to Mankiw, Romer, and Weil) that is omitted here for simplicity.² Assuming there is no unoccupied population (thus L represents the total population), income per capita is $y(t) = Y(t)/L(t)$. Since the production function is assumed homogeneous of degree one, income per “effective worker”, $\tilde{y}(t) = Y(t)/(L(t) \cdot A(t))$, is given by $\tilde{y}(t) = F(\tilde{k}(t), 1) \equiv \varphi(\tilde{k}(t))$, where \tilde{k} is the quantity of capital per effective worker.

The steady-state value of \tilde{y} (i.e. that in the long-run equilibrium, $\tilde{y}^* = \varphi(\tilde{k}^*)$), is defined by the condition

$$s\varphi(\tilde{k}^*) = (\xi + v + \delta)\tilde{k}^*. \quad (1)$$

For the Cobb-Douglas production function $Y = cK^a \cdot (L \cdot A)^{1-a}$, a log-linear approximation of Equation (1) around the steady state yields (see, e.g., Durlauf and Quah, 1999):

$$\ln \tilde{y}(t) - \ln \tilde{y}^* = (\ln \tilde{y}(0) - \ln \tilde{y}^*)e^{-\lambda t}, \quad (2)$$

where λ is the speed with which the growth path converges to the equilibrium path. Depending on whether accumulation is exogenous or endogenous, λ depends on the model parameters in different ways. But anyway $\lambda > 0$; whence it follows that $\tilde{y}(t) \rightarrow \tilde{y}^*$ as $t \rightarrow \infty$. Since $y(t) = \tilde{y}(t) \cdot A(0)e^{\xi t}$, Equation (2) can be rewritten as

$$\ln y(t) = \ln \tilde{y}^* + \ln A(0) + \xi t + (\ln y(0) - \ln A(0) - \ln \tilde{y}^*)e^{-\lambda t}, \quad (3)$$

where $\ln \tilde{y}^* + \ln A(0) + \xi t$ is the equilibrium growth path. Benefiting from Equation (1), \tilde{y}^* can be expressed as $\tilde{y}^* = g((\xi + v + \delta)/s)$, where $g(\cdot)$ is some function. To transit to econometric models,

² Barro and Sala-i-Martin (2004) and Durlauf and Quah (1999) describe and analyze the mentioned models.

Equation (3) is transformed into

$$\ln y(t) = \alpha + \beta_+ \ln y(0), \quad (4a)$$

or into an equivalent equation in terms of income per capita growth rates,

$$\ln(y(t)/y(0)) = \alpha + \beta \ln y(0), \quad (4b)$$

where $\alpha = (1 - e^{-\lambda t})(\ln A(0) + \ln \bar{y}^*) + \xi t$, $\beta_+ = e^{-\lambda t} < 1$, $\beta = \beta_+ - 1 = e^{-\lambda t} - 1 < 0$. A version of Equation (4b) with average growth rate over the time span $(0, t]$ is often used. To derive it, the left-hand and right-hand sides of Equation (4b) are divided by t . It is seen from (4b) that, since $\alpha > 0$ and $\beta < 0$, the lower the initial level of income per capita, $y(0)$, the higher the growth rate.

What implications can be drawn, if there is a set of economies $\{i\}$ governed by the above model? Let the economies be homogeneous in that they have the same production functions $\varphi(\cdot)$ and the values of parameters ξ , v , δ , s , and $A(0)$, differing in the initial income level $y_i(0)$ only due to different initial quantities of capital per capita $k_i(0) = K_i(0)/L_i(0)$. Then all economies have the same equilibrium growth path; their individual paths converge to it, incomes per capita in poor economies growing faster than those in rich economies. Such an evolution is referred to as unconditional (or absolute) convergence. Figure 1a provides an example. As it is seen, unconditional convergence will eventually lead to the equalization of incomes per capita across economies. However, if an econometric analysis based on Equation (4a) or (4b) confirms the unconditional convergence hypothesis (the estimate of β_+ or β is statistically significant with $\beta_+ < 1$ or $\beta < 0$), this does not necessarily imply income convergence. The reasons will be considered in Section 3.1.

Given heterogeneous economies, the growth path of each economy converges to its own equilibrium growth path. Such a situation is referred to as conditional convergence. In this case α cannot be considered as a constant that is identical for all economies. Then it is some function $\alpha_i = \alpha' + \alpha(x_{i1}, \dots, x_{in})$, arguments of which are observable parameters (e.g., the average rate of population growth v_i) and proxies characterizing – as the researcher supposes – unobservable parameters: $A_i(0)$, ξ_i , etc.³ Then Equation (4b) becomes:

$$\ln(y_i(t)/y_i(0)) = \alpha' + \alpha(x_{i1}, \dots, x_{im}) + \beta \ln y_i(0). \quad (5)$$

Conditional convergence suggests only that income per capita grows faster the further it is

³ Generally speaking, the value of β also cannot be assumed identical for all economies. For example, in the model with the exogenous saving rate, we have $\lambda = (1 - a)(\xi + v + \delta)$ in the neighborhood of the steady state (Sala-i-Martin, 2004, p. 57). Hence, if each economy has its own values of parameters involved in λ , then β differs across economies.

from its equilibrium value for a given economy. But it is uninformative as to dynamics of income inequality. Figure 1b provides an example of conditional convergence. There are two groups of homogeneous economies: the growth paths of economies 1 and 4 converge to the equilibrium path $\ln \bar{y}_{1,4}^* + \ln A_{1,4}(0) + \xi_{1,4}t$, those of economies 3 and 5 converge to $\ln \bar{y}_{3,5}^* + \ln A_{3,5}(0) + \xi_{3,5}t$, and economy 2 has its own equilibrium growth path $\ln \bar{y}_2^* + \ln A_2(0) + \xi_2 t$. By and large we observe income divergence of economies.

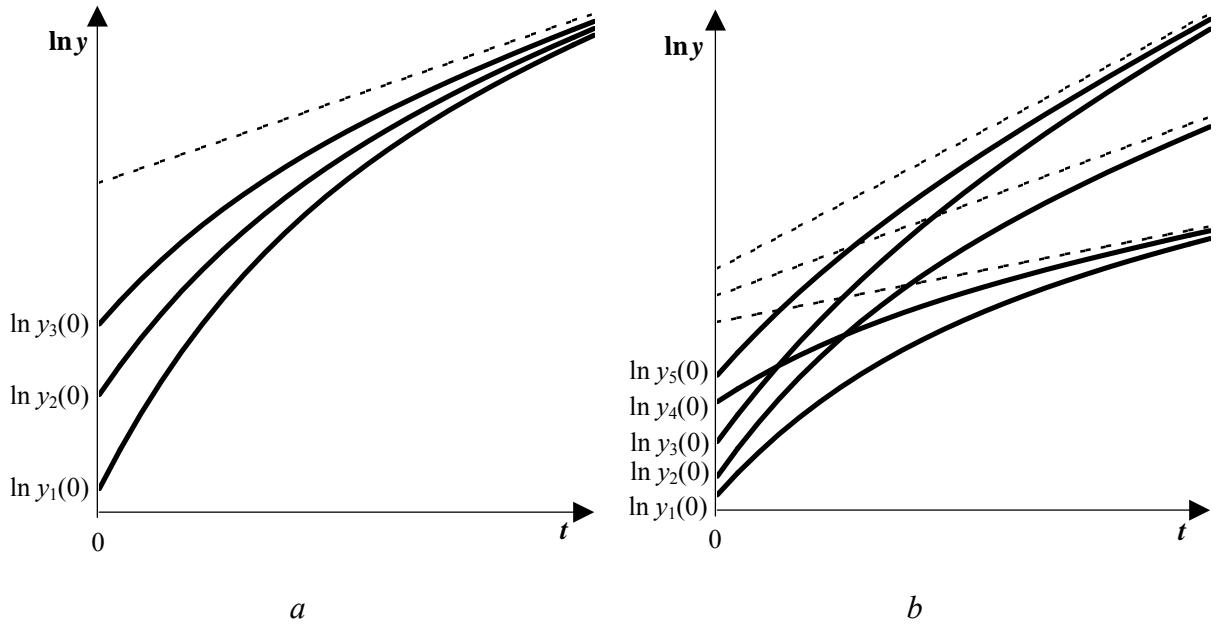


Figure 1. Unconditional (a) and conditional (b) convergence.
Dashed lines denote equilibrium growth paths.

Figure 1b demonstrates a particular case. Depending on the values of the structural parameters for specific economies (i.e. on the mutual arrangement of their equilibrium growth paths), and initial levels of per capita incomes, the following patterns of income dynamics in a set of economies are possible:

1. Global income convergence of economies, e.g., when the rate of technical progress is the same across all economies (that is, when the equilibrium growth paths are parallel) and the distance between extreme equilibrium paths is less than the spread of initial income levels. However, in contrast to unconditional convergence, income inequality does not vanish at the limit, stabilizing at some constant level.

2. Income divergence of economies, e.g., when each of them has its own equilibrium growth

path, and these paths diverge.

3. Local – or club – income convergence (polarization): convergence occurs within two or more groups of homogeneous economies, “convergence clubs.” It is this case that is depicted in Figure 1b. Given club convergence, an inequality indicator over all economies may either decrease or increase.

Thus, the neoclassical growth models predict income convergence for the only case: under the very strong condition of homogeneity of economies under consideration. If it is not valid, these models are not able to provide some defined information regarding income distribution dynamics.

It follows from the neoclassical models that economic growth of poor economies should be faster than that of richer economies even if the case in hand is conditional convergence. Let us write Equation (5) as $\ln(y_i(t)/y_i(0)) - \alpha(x_{i1}, \dots, x_{im}) = \alpha' + \beta \ln y_i(0)$. The left-hand side of this relationship represents the income per capita growth rate adjusted for the specificity of economy i (i.e. for the difference in economies’ equilibrium growth paths). Since $\beta < 0$, the less the $y_i(0)$, the higher the “adjusted” growth rate.

However, there are a number of growth models which do not imply conditional convergence. For example, if the saving rates differ across different kinds of income (the saving rate out of wage income equals s_1 , and the saving rate out of interest income equals s_2), then the economy may have two stable equilibrium states \tilde{y}_1^* and \tilde{y}_2^* , $\tilde{y}_1^* < \tilde{y}_2^*$. Albeit such a model is consistent with the neoclassical paradigm (constant returns to scale and decreasing marginal productivity), it predicts polarization even if economies are homogenous: growth of economies with low initial levels of income per capita converges to an equilibrium growth path with \tilde{y}_1^* , and that of economies with high initial income converges to a path with \tilde{y}_2^* , so implying a “poverty trap” (Galor, 1996).

A number of models abandon the neoclassical premises. Romer (1986) puts forward a model of endogenous growth in which knowledge is assumed to be an input in production that has increasing marginal productivity. The implications of this model is that “The level of per capita output in different countries need not converge; growth may be persistently slower in less developed countries and may even fail to take place at all.” (Romer, 1986, p. 1003). Another example is models with threshold effects. Azariadis and Drazen (1990) consider thresholds in accumulation of human and physical capital. They stem from spillovers between individual investments arising when aggregate capital is sufficiently high. As a result, the elasticity of output with respect to physical

capital has either of two values, depending on whether physical capital per worker is below or above a certain threshold, similarly for human capital per worker. Thus, one of four production functions is possible for a given economy, switching from the production functions being possible as human and physical capital accumulates. This model has multiple equilibria \bar{y}^* , which yields complex growth dynamics differing from conditional convergence.

Let us consider what happens once the openness of economies is introduced into a growth model. Borts and Stein (1964) developed a version of the Solow-Swan model with global labor and capital markets. It turned out, however, that under the assumption of perfect integration of factor markets the convergence rate would be infinite. Barro and Sala-i-Martin (1992) mention an extended neoclassical growth model that includes international trade and a global capital market. These features created a sharp distinction between domestic product (GDP or GRP) and personal income. In the case of homogenous economies, unconditional convergence of domestic products took place, whereas personal incomes did not converge. However, Barro and Sala-i-Martin's empirical analysis of gross state products and personal incomes across US states did not corroborate these implications. Therefore, Barro and Sala-i-Martin (2004) modify the model with some additional constraints. The difference between behavior of domestic product and personal income disappears in the new model; its analysis evidences that the possibility of capital to flow between economies accelerates convergence. Nonetheless, the model has some paradoxical properties. The authors conclude that at this stage there is no a fully satisfactory way to apply the Ramsey model to an open economy (Barro and Sala-i-Martin, 2004, p. 190). However, if the interaction between economies is confined to migration of workers, the properties of neoclassical models hold, the convergence speed increasing.

Lucas (1993) considers a different way of cross-economy interactions, namely human capital spillovers. As Durlauf and Quah (1999) note, whether unconditional or club convergence will happen in this model, depends on the nature of interactions, that is, whether all economies interact uniformly or there are coalitions of economies such that economies within a coalition interact more with one another than with those outside. Quah (1997) proposes a model which explicitly includes such coalitions. The coalitions form of economies with more or less similar stock of human capital, which gives then a possibility to increase current consumption through interaction with one another. At the same time, admitting an economy with human capital below the average over a coalition to this coalition would slow growth for all the economies already in the coalition, which is a reason for

stability of coalitions over time.

Concluding this section, let us address the case when economies under consideration are Russian regions. An application of neoclassical or other growth models to them for the period of 1992-1998 is nonsensical, as the economic recession occurred in that period rather than growth. This recession was of a specific nature, namely, the transformational one. Therefore, one or another special model describing such a process should be the theoretical foundation of analyses (there are a number of relevant models in the literature). But, regarding the period of 1999-2007, the growth theory is seemingly quite applicable. However, this theory is relevant over long time spans; serious empirical researches cover spans of several tens of years (in particular, to eliminate business cycle effects). It is doubtful that regularities of the economic growth process could manifest themselves during that short time span of eight years only.

3. METHODOLOGIES OF ANALYZING SPATIAL INCOME INEQUALITY

Methodologies for analyzing inter-economy income inequality may be broadly classified as the cross-section approach (including its panel version), time series approach, and distribution dynamics approach. For definiteness sake, we will tell about regions, albeit the presentation in this section pertains equally to international analyses. Regions will be denoted by index r (sometimes also by s); $r = 1, \dots, N$.

3.1. Cross-section approach

Very likely, the most popular method of studying inter-regional income inequality is β -convergence. A reason for widespread use of this method – in addition to the stereotype outlined in the previous section – is one more stereotype that β -convergence implies income convergence of regions. To argue this, Barro and Sala-i-Martin are cited again,⁴ although they constantly emphasize themselves that it is not true (e.g., in Barro and Sala-i-Martin, 2004, pp. 51-52 and 462-465).

To analyze β -convergence, econometric versions of Equations (4a), (4b), and (5) are used, considering regional incomes per capita $\{y_{rt}\}$ in two points in time, $t = 0$ and $t = T$. Taking account

⁴ They are sometimes referred to as pioneers of β -convergence as well as σ -convergence. This is not the case, however. Analyzing β -convergence was applied for different purposes well before their work (see, e.g., Friedman, 1992). All the more this holds true for so simple and obvious concept as σ -convergence, a decrease of some income inequality indicator (in particular, the standard deviation of log incomes) over time. Xavier Sala-i-Martin's credit is in that in his Ph. D. dissertation of 1990 (Robert Barro being its supervisor), he has given those, hitherto nameless, concepts the names, and demonstrated the way the β -convergence equation follows from the neoclassical growth models.

of random shocks ε_r , Equations (4a) and (4b) take, respectively, the forms:

$$\ln y_{rT} = \alpha + \beta_+ \ln y_{r0} + \varepsilon_r, \quad (6a)$$

$$\ln(y_{rT}/y_{r0}) = \alpha + \beta \ln y_{r0} + \varepsilon_r. \quad (6b)$$

The dependent variable in (6b) is frequently represented as the average annual growth rate:

$$\ln(y_{rT}/y_{r0})/T = \alpha + \beta \ln y_{r0} + \varepsilon_r. \quad (6c)$$

Hypothesis $\beta_+ < 1$ is tested in model (6a), and hypothesis $\beta < 0$ is tested in models (6b) and (6c). If the hypothesis is not rejected, then unconditional β -convergence holds.

Function $\alpha(x_{r1}, \dots, x_{rm})$ in Equation (5) is, as a rule, represented in the log-linear form, which yields the following econometric version of this equation:

$$\ln(y_{rT}/y_{r0})/T = \alpha + \beta \ln y_{r0} + (\alpha_0 + \alpha_1 \ln x_{r1} + \dots + \alpha_m \ln x_{rm}) + \varepsilon_r. \quad (7)$$

Recall that x_{r1}, \dots, x_{rm} are variables describing (as the researcher believes) parameters involved in the neoclassical growth model. If hypothesis $\beta < 0$ in model (7) is not rejected, then conditional β -convergence holds.

Let us turn first to unconditional β -convergence. A negative relationship between the income per capita growth rate and the initial income level (i.e. $\beta < 0$) suggests that income per capita in poorer regions tends to grow faster than that in richer ones. Seemingly (and as follows from Figure 1a), in doing so, inter-regional income inequality should diminish, e.g., the variance of log per capita income, $\sigma_t^2 = \sigma^2(\ln y_t)$, has to decrease. Such a deduction would be true, if the dynamics of regional growth strictly followed the theoretical paths depicted in Figure 1a. Then an initially poorer region would permanently keep poorer than a richer one, albeit the difference in incomes between them progressively diminishes. In other words, the income ranking of regions would keep invariable. In reality, however, the situation is different. Because of some contingencies unaccounted in the theoretical model (random shocks ε_r in its econometric version), some of regions may “outrun” their theoretical path, hence leaving other regions behind (instead of catching up), and some of regions may “lag behind” the theoretical paths.⁵ To put it differently, relative mobility of regions happens (it will be discussed in Section 3.3). In that instance β -convergence do not necessarily leads to income convergence, in particular, to decrease of σ_t^2 (i.e. to σ -convergence; we

⁵ The intuition can be captured by a transparent example of two countries with non-synchronous business cycles. If the poorer country experiences expansion in some spell, while the richer country experiences recession in the same spell, then the poorer country can take the lead over the richer one for some time.

will also return to it in Section 3.3). The inference that σ -convergence occurs, based on that an estimate of β_+ in Equation (6a) is less than 1, is an analogue of Galton's fallacy,⁶ which is pointed out by a number of economists, e.g., Friedman (1992), Quah (1993), Wodon and Yitzhaki (2006), to name a few.

These informal considerations can be expressed in a strict form, following Barro and Sala-i-Martin (2004), pp. 50-51. Assume that the evolution of income per capita is given by Equation (6a) with $\beta_+ < 1$ for points in time $t - 1$ and t instead of 0 and T . Let N be great, and regression residual ε_r has zero mean, the same variance σ_ε^2 for all regions, and is independent over time and across regions. Then the variance of log income σ_t^2 tends to $\sigma^2 = \sigma_\varepsilon^2 / (1 - \beta_+^2)$ as $t \rightarrow \infty$, and the evolution of σ_t^2 can be expressed as $\sigma_t^2 = \sigma^2 + (\sigma_0^2 - \sigma^2)\beta_+^{2t}$, where σ_0^2 is the value of σ_t^2 at $t = 0$. Hence, income inequality falls only if $\sigma_0^2 > \sigma^2$; otherwise it rises despite β -convergence holds. It also follows herefrom that σ -convergence implies β -convergence, while the reverse is not true.

Thus, to study trends of inter-regional income inequality, it is more reasonable to directly analyze the evolution of some inequality indicator. If income convergence happens, an analysis of β -convergence provides no additional information. If it does not happen, the results of the analysis of β -convergence can prove to be misleading.

Let us consider a striking example of such cases. Nominal personal incomes per capita per month in 2005 by Russian region from Rosstat (2008), pp. 157-158, are taken as y_{r0} . Composite subjects of the Russian Federation (which include autonomous *okrugs*) are considered as single regions; the Chechen Republic, for which income data are lacking, is omitted. So, we have 79 regions. The data for the final period, y_{rT} , are so constructed that the income distribution remains the same, thus assuring the absence of inter-regional income convergence. This is done by interchanging the positions of incomes in regions that are adjacent in the original table from Rosstat (2008).

The first line in Table 1 reports estimation results for regression (6b) with such variables (T is taken to be 1). Despite that income inequality does not change over the time span of 0 to T (the

⁶ Galton (1884) found that the adult children of tall parents tended to be shorter than their parents, and the children of short parents tended to be taller than parents (by the way, the term "regression" had its origin just in this). This fact seemingly implies that the heights of all adult people should even out over time. (Contrary to certain assertions in the literature, Galton himself not in the least makes such an inference. Quite the reverse, he explains why this does not happen – see p. 256 of his memoir. Thus the term "Galton's fallacy" or "Galton's paradox" is unjust.) Some details apart, the relationship found by Galton is formalized just as Equation (6a) (with no logarithms) with $\beta_+ = 2/3$.

standard deviation of log income $\sigma(\ln y_0) = \sigma(\ln y_T) = 0,407$, the Gini coefficient $G(y_0) = G(y_T) = 0,237$), the regression results suggest β -convergence ($\beta < 0$). Here it is, “Galton’s fallacy!”

Table 1. Paradoxes of β -convergence

Regression	Constant α		β	
	Estimate	p-value	Estimate	p-value
$\ln(y_{rT}/y_{r0}) = \alpha + \beta \ln y_{r0} + \varepsilon_r$	5.992 (0.945)	0.000	-0.687 (0.108)	0.000
$\ln(y_{r0}/y_{rT}) = \alpha + \beta \ln y_{rT} + \varepsilon_r$	5.992 (0.945)	0.000	-0.687 (0.108)	0.000
$\ln(y'_{rT}/y_{r0}) = \alpha + \beta \ln y_{r0} + \varepsilon_r$	11.985 (1.889)	0.000	-0.375 (0.217)	0.087
$\ln(y_{r0}/y'_{rT}) = \alpha + \beta \ln y'_{rT} + \varepsilon_r$	5.992 (0.945)	0.000	-0.844 (0.054)	0.000

Standard errors are in parentheses.

Since β -convergence occurs in the transition from $\{y_{r0}\}$ to $\{y_{rT}\}$, it is reasonable to expect that it should not occur backward in time, from $\{y_{rT}\}$ to $\{y_{r0}\}$. Nonetheless, the second line in Table 1 suggests that β -convergence is observed backward in time as well (regression results not changing at all). What is more, β -convergence (therewith bi-directional) can be observed even with rising income inequality. Let us square incomes for period T (constructed as indicated above): $y'_{rT} = y_{rT}^2$. Then the inequality indicators in the final period increase twofold and up as compared to the initial period 0: $\sigma(\ln y'_{rT}) = 0,815$, $G(y'_{rT}) = 0,503$. The third line in Table 1 reports estimation results for the respective regression. The estimate of β is negative again (being significant at the 10-percent level)! Backward β -convergence occurs as well (the fourth line in Table 1). This does not defy common sense here, however, as income inequality decreases in the transition from $\{y'_{rT}\}$ to $\{y_{r0}\}$.

It is not to be supposed that the paradoxes discussed above can be found in artificial examples only. For instance, bi-directional β -convergence is observed in the case of nominal incomes per capita across Russian regions in 1995 and 2001 with $\beta = -0,054$ forward in time and $\beta = -0,055$ backward; true enough, both estimates are significant at the 15-percent level (author’s estimates using data from Rosstat, 2008, pp. 157-158). Wodon and Yitzhaki (2006) find bi-directional β -convergence (with $\beta = -0,21$ forward in time and $\beta = -0,43$ backward) for Latin American countries in 1960 and 1998.

It is argued on occasion that unconditional β -convergence is, indeed, a questionable tool for analyzing trends of spatial income inequality, but this does not hold for conditional β -convergence. However, first, as has been demonstrated in the previous section, conditional β -convergence is not able to yield any information at all on the evolution of income inequality. Second, all of the

preceding regarding unconditional β -convergence is true for conditional β -convergence, since Equation (7) succeeds to all disadvantages of Equations (6a)-(6c). Equation (7) can be rewritten in an equivalent form similar to (6a):

$$\ln y_{rT} = \alpha + \beta_+ \ln y_{r0} + (\alpha_0 + \alpha_1 \ln x_{r1} + \dots + \alpha_m \ln x_{rm}) + \varepsilon_r. \quad (8)$$

The attribute “conditional” in the term means that convergence is observed if it is conditioned on differences between regional economies (i.e. between equilibrium growth paths), hypothetically described by variables x_{r1}, \dots, x_{rm} . In other words, as stated in the previous section, one can adjust per capita incomes y_{rT} (or growth rates) for regional differences: $\ln y^o_{rT} = \ln y_{rT} - (\alpha_0 + \alpha_1 \ln x_{r1} + \dots + \alpha_m \ln x_{rm})$, after which convergence becomes unconditional. Since parameters $\alpha_0, \dots, \alpha_m$ are unknown, the adjusted incomes are estimated as $\ln y^o_{rT} = \hat{v}_r$, where \hat{v}_r are estimates of the residuals in regression $\ln y_{rT} = \alpha_0 + \alpha_1 \ln x_{r1} + \dots + \alpha_m \ln x_{rm} + v_r$. Then Equation (8) transforms to $\ln y^o_{rT} = \alpha + \beta_+ \ln y_{r0} + \varepsilon_r$ or $\ln(y^o_{rT}/y_{r0})/T = \alpha + \beta \ln y_{r0} + \varepsilon_r$, i.e. a regression identical to Equation (6c). Regarding this regression, we may tell again about Galton’s fallacy and so on.

Conditional β -convergence poses a specific problem of linkage between the original economic model of growth and a regression of the form (8). Returning to theoretical relationship (3), regression (8) in fact estimates $\hat{y}_r^* = g((\xi_r + v_r + \delta_r)/s_r)$ – i.e. parameters of $g(\cdot)$ – and (possibly) $A_r(0)$ and ξ_r expressed as functions of other economic variables. Auxiliary variables x_{r1}, \dots, x_{rm} are chosen in an ad hoc way, the choice being substantiated by some heuristic considerations. Durlauf and Quah (1999), pp. 277-281, list several tens of variables used as x_{rj} by different authors in international studies. Among them there are, e.g., corruption, religion, and rule of law. It is far from clear, in what way such variables can enter into the economic model; and if they are added to it, whether this would give a regression of the form (8). No more or less strict grounds of representing function $\alpha(x_{r1}, \dots, x_{rm})$ as a log-linear one exist as well. One more problem is possible endogeneity of auxiliary variables, i.e. reverse causality, when it is not a change in the value of some variable x_{rj} that contributes to economic growth, but, vice versa, this change is caused by growth.

An outgrowth of cross-sectional analysis is the panel-data analysis, a panel being treated as a set of cross sections observed across a number of points in time. This method allows better taking into account heterogeneity of regions and characterizing time variation of parameters. However, all fundamental problems related to β -convergence still stand (and new problems caused

by features of the panel-data analysis arise in addition). I do not dwell on this issue; see Durlauf and Quah (1999) and Magrini (2004) for a detailed discussion.

So both unconditional and conditional β -convergences are useless in studying trends of the evolution of inter-regional income inequality. Widespread use of them in such studies is based on a mass fallacy caused by the two considered stereotypes. Does this imply that the very concept of β -convergence is fallacious? Not at all. It is not the concept itself that is guilty but the misinterpretation of the concept, the improper use of it. By means of the β -convergence analysis, researchers tackle a question which fundamentally cannot be answered by this method. Empirical analysis of β -convergence is able to shed light on the only issue: whether behavior of economies possesses properties emerging from one or other version of the neoclassical growth model. Its area of application is fairly narrow: verification of economic models of growth.⁷

Nor disclaims all the aforesaid reasonableness of the use of cross-sectional (as well as panel) analysis for studying inter-regional income inequality. Regressions of the form (7) and (8) are well applicable for this purpose, if one abandons the misinterpretation of β -convergence or dismisses this concept at all. In the latter case, a researcher is no more constrained by a growth model giving rise to this concept (and the sign of β or $\beta_+ - 1$ has no predetermined interpretational burden); initial incomes y_{r0} may be absent at all among explanatory variables. The type of regression itself is not a priori restricted: it may be additive, multiplicative (additive in logarithms) or some else. The objective of an analysis is now to reveal what determines income per capita or its growth rate in regions (hence, what determines inter-regional inequality) rather than to find whether income convergence of regions occurs. Such an analysis can be called “causal cross-sectional analysis.”

Nonetheless, the problem of the linkage with the economic theory does not disappear. It is almost always possible to find some set of regressors that are statistically significant in a relationship explaining a dependent variable. Then the questions arise: What are groundings in the theory for the relationship found? What is the mechanism of action on a phenomenon under study by factors that are characterized by chosen explanatory variables? The best answer to these questions would be an economic model (as the starting point of the analysis or as its result). However, authors of such papers often restrict themselves to intuitive considerations.

⁷ True enough, the possibilities of this method are limited in that respect as well, as pointed out by many economists, e.g., Durlauf and Quah (1999). (Add to this that it is incomprehensible how to interpret cases of bi-directional β -convergence.) However, this issue is beyond the scope of the paper.

As indicated in the Introduction, because of openness of regional economies, econometric models have to explicitly take into account cross-region interactions. An obvious way is to insert variables describing such interactions into a model. For example, analyzing β -convergence between regions in a number of countries, Barro and Sala-i-Martin (2004) include in regression (7) – among other variables – the annual rate of net migration into a region⁸. However, such a method of characterizing cross-region interactions is not exhaustive. Unaccounted (e.g., unobservable) ways of regions to impact on one another always remain. This gives rise to a phenomenon similar to autocorrelation in time series, namely to spatial autocorrelation. Its results are also similar: biases in standard errors of regression estimates (sometimes, inconsistency of estimates themselves is possible as well), which biases statistical inferences.

The time axis is one-dimensional; observations on it are naturally ordered; a lagged (preceding) observation is always unique. This does not hold for space; therefore autocorrelation in spatial series is much more complex than in time series. A special branch of econometrics, spatial econometrics, takes spatial autocorrelation up (Anselin, 2001, 2002; Arbia, 2006; LeSage and Pace, 2009). It suggests a few approaches to modeling spatial interactions. Let we have an original econometric model $z_r = \alpha_0 + \alpha_1 x_{r1} + \dots + \alpha_m x_{rm} + \varepsilon_r$. In matrix form, it looks like

$$\mathbf{z} = \mathbf{X}\boldsymbol{\alpha} + \boldsymbol{\varepsilon}, \quad (9)$$

where \mathbf{z} is an $N \times 1$ column vector of a dependent variable (e.g., incomes per capita or their growth rates by region), \mathbf{X} is an $N \times (m + 1)$ matrix of explanatory variables including the constant, $\boldsymbol{\alpha}$ is an $(m + 1) \times 1$ column vector of coefficients to be estimated, and $\boldsymbol{\varepsilon}$ is an $N \times 1$ column vector of regression residuals. The approaches differ in ways of modeling spatial effects that modify the original model (9). The choice of one or other approach is dictated by intuitive considerations discussed in the literature (Anselin, 2002; Magrini, 2004; and others).

One approach bases on the assumption that the dependent variable is autocorrelated. Insertion of the autoregressive term into Equation (9) yields the spatial autoregressive model

$$\mathbf{z} = \mathbf{X}\boldsymbol{\alpha} + \rho \mathbf{W}\mathbf{z} + \boldsymbol{\varepsilon}. \quad (10)$$

Here, ρ is the autoregressive coefficient, and $\mathbf{W}\mathbf{z}$ is called a spatial lag which is a counterpart of a preceding observation (lagged variable) in time series models. Denote $\mathbf{W}\mathbf{z} = \mathbf{z}_{(-1)} = (z_{r(-1)})$. For a

⁸ By the way, this can serve to illustrate the above-mentioned endogeneity problem. It may well be that a higher net migration rate in a region is not a reason for a higher per capita growth rate there but a result of it, as rapid rise in incomes likely stimulates inflow of migrants from other regions.

given r , the spatial lag is a weighted average of spatial observations (different from the r th one, since w_{rr} is always taken as zero): $z_{r(-1)} = \sum_{s=1}^N w_{rs} z_s$. The spatial weight matrix $\mathbf{W} = (w_{rs})$ is an $N \times N$ one. Its element w_{rs} characterizes the contribution of region s to the impact to the value of z_r in region r . The elements are typically standardized, such that their sum over each row equals one. Frequently, \mathbf{W} is a “neighborhood matrix:” $w_{rs} = 1/n_r$, if regions r and s are neighbors (share a common border), and $w_{rs} = 0$ otherwise; n_r is the total number of r ’s neighbors. Then the spatial lag is simply an average of z_s in neighboring regions. There is no common method of choosing weights, though. An alternative way of constructing the spatial weight matrix is to take regions with distance between them less than a prescribed value as neighbors. Decaying functions of physical or economic distance between regions are also used as weights w_{rs} . Some specific examples can be found in Section 4.

Another approach considers residuals in regression (9) as autocorrelated, which gives the spatial error model

$$\mathbf{z} = \mathbf{X}\boldsymbol{\alpha} + \mathbf{v}, \mathbf{v} = \rho\mathbf{W}\mathbf{v} + \boldsymbol{\varepsilon}. \quad (11)$$

Both dependent and explanatory variables are autocorrelated in this model. Combining both equations of model (11), we get $\mathbf{z} = \mathbf{X}\boldsymbol{\alpha} + \rho\mathbf{W}\mathbf{z} - \rho\mathbf{W}\mathbf{X}\boldsymbol{\alpha} + \boldsymbol{\varepsilon}$. Note that models (10) and (11) coincide in the case of “pure” spatial autocorrelation (when regressors \mathbf{X} are absent).

The presence of spatial autocorrelation in regressions (10) and (11) leads to endogeneity. Hence, the ordinary least squares is inapplicable to them. Special estimation methods are needed such as the maximum-likelihood method, generalized method of moments, method of instrumental variables, etc.

One more approach, like the previous one, bases on the assumption that residuals in regression (9) are autocorrelated. However, covariance matrix $\boldsymbol{\Omega} = (\omega_{rs}) = (\text{cov}(\varepsilon_r, \varepsilon_s))$ is modified rather than the regression itself; the generalized least squares is applied for estimation. This requires some assumption as to the structure of $\boldsymbol{\Omega}$. For example, it is reasonable to assume that inter-regional interactions decay with rising distance between them, L_{rs} , as, say, $\omega_{rs} = \kappa + \mu \cdot \exp(-\lambda L_{rs})$, $\lambda > 0$. Coefficients κ , μ , and λ are estimated with the use of regression $\hat{\varepsilon}_r \hat{\varepsilon}_s = \kappa + \mu \cdot \exp(-\lambda L_{rs}) + \xi_{rs}$, where $\{\hat{\varepsilon}_r\}$ are estimates of the residuals in regression (9). Thereafter regression (9) is reestimated, using the estimated covariance matrix.

There are a number of tests for spatial autocorrelation. The simplest and most popular in spatial econometrics test bases on Moran's I statistic, a spatial counterpart of the Durbin-Watson statistic (Anselin, 2001; Arbia, 2006). As applied to regression (9), this statistic is $I = \mathbf{\epsilon}'\mathbf{W}\mathbf{\epsilon}/\mathbf{\epsilon}'\mathbf{\epsilon}$ ($\mathbf{\epsilon}'$ denotes the transpose of $\mathbf{\epsilon}$). A statistically significant difference of I from zero suggests spatial autocorrelation of the regression residuals, hence, the need to modify the model specification – e.g., by use of a model of the form (10) or (11) instead (9). The test for spatial autocorrelation can be applied directly to raw data, in which case $I = \mathbf{z}'\mathbf{W}\mathbf{z}/\mathbf{z}'\mathbf{z} = \text{cov}(z, z_{(-1)})/\sigma^2(z)$.

Spatial autocorrelation admits fairly simple visualization. Let z_r be income per capita in region r , \mathbf{W} be a simple neighborhood matrix (then spatial lag $z_{r(-1)}$ is the average of incomes per capita in neighboring regions), and z^* be a border between high and low incomes (e.g., the average of incomes over all regions). In the Moran scatterplot, the axes of coordinates are z and $z_{(-1)}$. These produce four quadrants: quadrant HH contains rich regions surrounded by rich ones ($z_r > z^*$ and $z_{r(-1)} > z^*$), quadrant HL contains rich regions surrounded by poor ones ($z_r > z^*$ and $z_{r(-1)} < z^*$), quadrant LL contains poor regions surrounded by poor ones ($z_r < z^*$ and $z_{r(-1)} < z^*$), and quadrant LH contains poor regions surrounded by rich ones ($z_r < z^*$ and $z_{r(-1)} > z^*$). A concentration of observations in quadrants HH and LL evidences spatial autocorrelation (the regression line of z on $z_{(-1)}$ is also sometimes depicted in this plot). A map, in which a color indicates belonging of a region to one of the above four groups, can provide an interesting picture of inter-regional interactions.

It should be noted that, despite the importance of spatial interactions (and a very probable risk of misspecification of econometric model when they are neglected), spatial econometrics is rarely used in cross-sectional analyses.

3.2. Time series approach

It is by no means a fact that long-run dynamics of per capita incomes in regions really follow paths like depicted in Figure 1. It is not inconceivable that the actual path for each region is, e.g., a random walk with drift. Although no tendency for income convergence of regions takes place in such an event, the income paths can appear to be approaching one another during some time span. Herefrom the advisability follows to directly study income dynamics, i.e. to model time series of regional incomes.

A method due to Carlino and Mills (1996) proceeds from the neoclassical growth model.

Simplifying, its idea is as follows. Regional income per capita related to the national average y_{rt} is assumed to tend to a time-invariant equilibrium value, y_r^* (the difference in equilibria across regions implies conditional convergence). Then $\ln y_{rt} = \ln y_r^* + d_{rt}$, where d_{rt} are deviations from the equilibrium. These are modeled as consisting of a deterministic linear trend and a stochastic process: $d_{rt} = d_{r0} + \delta_r t + v_t$, where d_{r0} is the initial deviation from the equilibrium and δ_r is the rate of convergence to the equilibrium (unlike the β -convergence models, it differs across regions). Herefrom, $\ln y_{rt} = (\ln y_r^* + d_{r0}) + \delta_r t + v_t$. If a region is above its equilibrium initially, i.e. $d_{r0} > 0$, one should expect the trend to be negative ($\delta_r < 0$), and vice versa. Assuming v_t to be a first-order autoregressive process,⁹ $v_t = \rho v_{t-1} + \varepsilon_t$, we obtain $\ln y_{rt} = \alpha_r + \rho \ln y_{r,t-1} + \delta'_r t + \varepsilon_t$, where $\delta'_r = (1 - \rho)\delta_r$. Econometric analysis aims to find whether dynamics of incomes per capita in a region really satisfy this model. This holds if time series $\{\ln y_{rt}\}_{t=0,\dots,T}$ is stationary about the trend, i.e. if $\rho < 1$ (the unit root hypothesis $\rho = 1$ is rejected). Carlino and Mills designate such a case as “stochastic convergence.”

It is easily seen that the method considered is of interest only from viewpoint of verifying economic growth models; it is uninformative as to income convergence of regions. Indeed, assume that one finds stochastic convergence for all regions under consideration. This implies that the paths of log relative per capita incomes are described by linear trends with an individual rate for each region. With this, certain of rates are positive and certain of them are negative. As equilibrium values y_r^* are unobservable, there is no way of telling what are dynamics of income distribution across the entire set of regions and whether income convergence or divergence of regions occurs. Besides, characterizing convergence by linear trends is lame: having reached the equilibrium, the income path would be driven further by the linear trend and move away from the equilibrium.

Bernard and Durlauf (1995) propose a different way. According to their definition, income convergence between regions r and s occurs when

$$\lim_{t \rightarrow \infty} E(\ln y_{rt} - \ln y_{st}) = 0, \quad (12)$$

$E(\cdot)$ standing for the expectation (mean). This relationship follows from the neoclassical growth model – see Equation (3) – for a pair of homogeneous regions. To test time series for consistency with definition (12), a model $\ln y_{rst} = v_t$ is used, where $\ln y_{rst} = \ln y_{rt} - \ln y_{st}$. Assuming $v_t = \rho v_{t-1} + \varepsilon_t$,

⁹ This condition is accepted for ease in exposition only. In fact, Carlino and Mills (1996) (as well as Bernard and Durlauf, 1995, and Evans and Karras, 1996, see below) use more general assumptions regarding autocorrelation.

we get a standard regression of testing a time series for stationarity: $\ln y_{rst} = \rho \ln y_{rs,t-1} + \varepsilon_t$. Series $\{y_{rst}\}_{t=0,\dots,T}$ is stationary, hence satisfies relationship (12), if $\rho < 1$. Evans and Karras (1996) propose a similar method. Instead of Equation (12), they adopt a condition of the form $\lim_{t \rightarrow \infty} E(\ln y_{rt} - \ln \bar{y}_t) = \mu_r$, where \bar{y}_t is the average over all regions, and μ_r is an arbitrary constant.

This makes possible analyzing both unconditional ($\mu_r = 0$ for all r) and conditional ($\mu_r \neq 0$) convergence. The key distinction is that convergence of all regions with one another is analyzed rather than convergence in a pair of regions. To do this, a $T \times N$ panel is tested for stationarity, the panel being treated as a set of time series $\ln y_{rt} - \ln \bar{y}_t$ observed in N objects. Hobijn and Franses (2000) define convergence directly in terms of stationarity. Together with a test for stationarity, they introduce a procedure that distinguishes convergence clubs.

Relationship (12) directly suggests income convergence of regions r and s . It occurs in the entire set of regions, if condition (12) is true for $N - 1$ region pairs, where one of elements in the pairs is fixed as a benchmark. However, Bernard and Durlauf (1995) use a technique of testing time series for consistency with definition (12) that is much narrower than this definition.¹⁰ Stationarity of series $\{\ln y_{rst}\}_{t=0,\dots,T}$ means that incomes in regions r and s already coincide accurate to random shocks. That is, the process of convergence as such has already been completed and the equilibrium growth path determines income dynamics in both regions. Then what does stationarity suggest? The suggestion is that the effect of a single random shock is transient. A deviation from the path $\ln y_{rst} = 0$ (caused by the shock) decays with time (halving in time $\theta = \ln 0.5 / \ln \rho$); and income difference tends to return to this path. Thus convergence here characterizes the short-run properties of income dynamics, namely, the behavior of transient deviations from the predetermined long-run path $\ln y_{rst} = 0$.

If convergence as such happens, i.e. if a nonzero gap in incomes between r and s permanently decrease (accurate to random shocks), then series $\ln y_{rst}$, satisfying definition (12), turns out to be nonstationary. In this case, two kinds of convergence combines: long-run convergence, i.e. convergence of long-run trends of incomes in regions r and s to each other (convergence of $\ln y_{rst}$ to 0), and short-run convergence, as in the preceding case, i.e. convergence of actual paths affected by short-run random shocks to long-run trends (stationarity of $\ln y_{rst}$ about a long-run trend). Figure 2 illustrates the difference between long-run and short-run convergences. Figure 2a demonstrates a

¹⁰ Note that Equation (12) is not equivalent to a definition of stationarity (see, e.g., Greene, 2003, p. 612).

case corresponding to the hypothesis tested in Bernard and Durlauf (1995). Figure 2b demonstrates a process combining long-run and short-run convergences. The testing technique due to Bernard and Durlauf does not identify such processes (as they themselves note in Bernard and Durlauf, 1996).

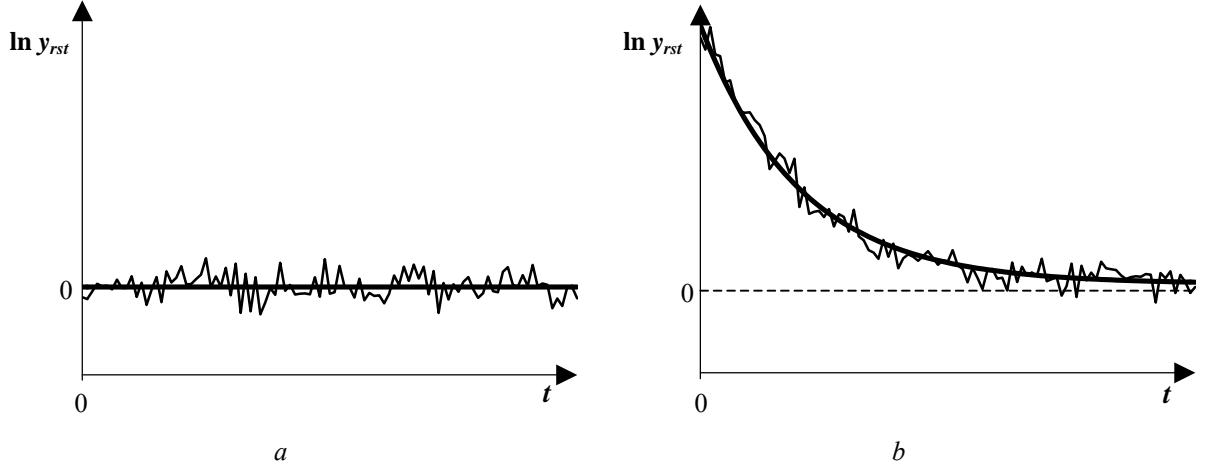


Figure 2. Short-run (a) and long-run (b) convergence.
Thick lines represent long-run trends; thin lines depict actual paths.

Nahar and Inder (2002), using definition (12), attempt to overcome this shortcoming in their test for long-run convergence. Its idea is as follows. The evolution of income disparities in a pair of regions is modeled as $(\ln y_{rst})^2 = h(t) + \varepsilon_t$, where $h(t)$ is a long-run trend, and ε_t are random shocks with standard properties. A polynomial of some degree k approximates function $h(t)$, so that $(\ln y_{rst})^2 = \alpha_0 + \alpha_1 t + \alpha_2 t^2 + \dots + \alpha_k t^k + \varepsilon_t$. If long-run convergence (income convergence of regions) happens, then $h(t)$ is decreasing with time; hence, $dh(t)/dt < 0$ must hold for all t . To test for convergence, one considers whether the time average of this derivative is negative:

$$\frac{1}{T} \sum_{t=0}^T \frac{dh(t)}{dt} = \sum_{i=1}^k \alpha_i \frac{i}{T} \sum_{t=0}^T t^{i-1} < 0. \text{ However, it is not equivalent to the negativity of } dh(t)/dt \text{ in all points in time. Therefore, the test is not adequate. It is easy to see, considering a continuous-time counterpart of the above relationship: } \frac{1}{T} \int_0^T \frac{dh(t)}{dt} dt = \frac{1}{T} (h(T) - h(0)) < 0. \text{ Thus, merely the fact that } h(T) < h(0) \text{ suffices to accept the convergence hypothesis. In the general case, this obviously does not evidence long-run convergence. For example, a U-shape path of income disparities may satisfy this test. Besides, the test does not take account of possible autocorrelation of regression residuals, which makes it impossible to discriminate between deterministic and stochastic trends.}$$

The first disadvantage of the Nahar-Inder method is due to too general representation of the long-run trend. A way out can be, e.g., to restrict the function class of $h(t)$ to asymptotically decaying trends. A way to get rid of the second disadvantage is obvious: one should discard the condition that the regression residuals are uncorrelated, representing them as an autocorrelated process.¹¹ Gluschenko (2006a) uses these two ways, although in a different context of analyzing price convergence of regions. He applies function $h(t) = \ln(1 + \gamma e^{-\delta t})$ as a long-run asymptotically decaying trend, where $\delta > 0$, and γ and δ are parameters to be estimated,. Phillips and Sul (2007) put forward a more general method. They represent the evolution of incomes in a set of economies by a panel model $\ln y_{rt} = \delta_{rt}\mu_t$, where μ_t is a single common component and δ_{rt} is a time varying idiosyncratic component. The latter, in turn, is modeled as $\delta_{rt} = \delta_r + \sigma_r L(t)^{-1} t^{-\alpha} \xi_{rt}$, where $L(t)$ is a slowly varying function (i.e. such that $L(at)/L(t) \rightarrow 1$ and $L(t) \rightarrow \infty$ with $t \rightarrow \infty$) and ξ_{rt} are random shocks. Convergence occurs when $\delta_r = \delta$ and $\alpha \geq 0$. The authors propose a test for global convergence as well as a method of clustering panels into club convergence groups.

It is worth noting that, while analyzing a time series, the necessity may occur to take account of a structural break, an instantaneous change in the path of a process under consideration caused by some externals. In the case of Russia, such breaks owe to the 1998 financial crisis and, possibly, the beginning of the global economic crisis in the last months of 2008. Structural breaks have to be explicitly present in time series models; otherwise standard tests would reject stationarity because of model misspecification. (Among the above papers, only Carlino and Mills, 1996, take structural breaks into account.)

Analyzing the evolution of pairwise income differences between regions yields a detailed pattern of income convergence. In particular, it immediately reveals convergence clubs. However, such an analysis can prove to be too cumbersome. For example, 79 Russia's regions produce $79 \times 78/2 = 3081$ pairs. A palliative may be a comparison of the income evolution in each region with the evolution of the national average income rather than pairwise comparison of regions. Another way is to fix one region as a comparison benchmark. Theoretically, this way is proper, as $\ln y_{rst} = \ln y_{rqt} - \ln y_{sqt}$. That is, one can obtain all possible region pairs from $N - 1$ pairs (r, q) , where q is fixed and r runs remaining regions. However, care must be exercised here. For example, the

¹¹ Here is a problem, however, associated with that one needs to estimate a non-standard distribution of a unit root test statistic (even with a polynomial approximation of the trend, as such distributions are tabulated for $k \leq 2$ only).

absence of convergence between regions r and q as well as between s and q does not imply that there is no convergence between regions r and s . As is shown in a number of papers, other results of an analysis may depend on the choice of benchmark in practice.

3.3. Distribution dynamics approach

While an observation in the two preceding approaches is income per capita in an individual region in a point in time, this approach takes the entire cross-section distribution in a given point in time as the observation. In what follows, y_t denotes income per capita at time t . Depending on the context, y_t is interpreted as a random variable or as a set of its realizations $y_t = (y_{1t}, \dots, y_{rt}, \dots, y_{Nt})$, i.e. values of income across regions. The probability density of y_t at time t is denoted $f_t(y_t)$; for brevity, it will be hereafter called simply distribution.

All informal methods of analyzing income inequality dynamics pertain to the approach under consideration. They consist in examining changes of some measure of income dispersion $\Phi(y_t)$ (which characterizes income inequality) over time. A decreasing trend suggests income convergence of regions (σ -convergence). Typically, the presence or absence of such a trend is directly seen from data or a graph of $\Phi(y_t)$ over a certain period, which allows dispensing with formal tests. Widespread measures of inequality are the standard deviation of the logarithm of per capita income ($\Phi(y_t) = \sigma(\ln(y_t))$, hence the name), Gini coefficient $G(y_t)$, Theil index, and coefficient of variation; sometimes, other measures are exploited. An advantage of the Theil index is its decomposability. For example, the total value of income inequality in a country can be decomposed to a part due to intra-regional inequality and a part due to inter-regional inequality.

Analyzing σ -convergence involves only one parameter of income distribution. Therefore it gives but a fairly poor idea of inequality dynamics. In particular, such an analysis is unable to reveal club convergence which may be accompanied by either increase or decrease of income inequality. It is more informative to examine changes in the shape of income distribution over time. The most popular method of estimating distribution is kernel density estimation described by Silverman (1986):

$$\hat{f}_t(y_t) = \frac{1}{Nh} \sum_{r=1}^N K(z_{rt}), \text{ where } z_{rt} = (y_t - y_{rt})/h, h \text{ is a smoothing bandwidth, and } K(\cdot) \text{ is a}$$

kernel function (Gaussian kernel $K(z_{rt}) = (2\pi)^{-1/2} \exp(-z_{rt}^2/2)$ is often applied). Two or more modes in the distribution may evidence income polarization of regions; their emergence in a series of $f_t(y_t)$ may suggest club convergence.

Figure 3 demonstrates an artificial example of a bimodal cross-region income distribution. It suggests polarization as a result of income convergence among poor regions and convergence among rich regions. However, one ought not to rely on graphical analysis here; it is necessary to test the distribution for multimodality. In doing so, hypothesis to be tested is $H_0: f(y_t)$ has k modes versus $H_1: f(y_t)$ has more than k modes. With $k = 1$, it is a test for the presence of convergence clubs among regions; with $k > 1$, the number of clubs is found. Tests of this kind are described in the literature, e.g., by Silverman (1986) and Herzfeld (2008).

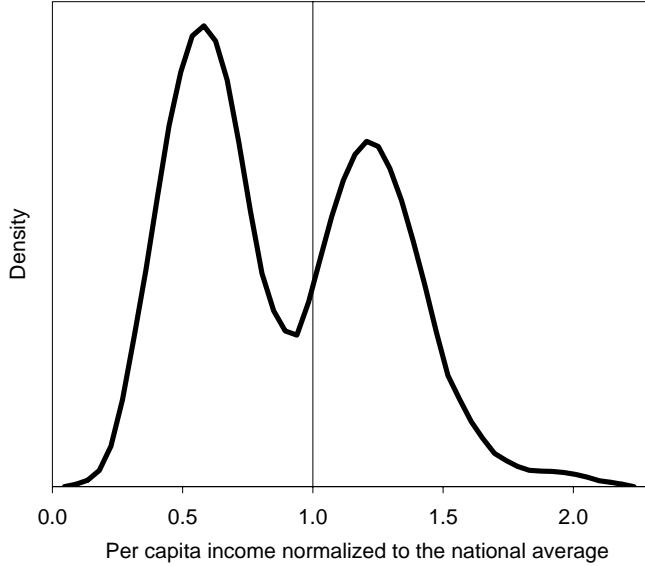


Figure 3. Club convergence (polarization)

Albeit the method considered makes it possible to detect polarization, it does not provide its quantitative characterization. There are methodologies focused just on measuring the degree of polarization (Esteban and Ray, 2007). Intuitively, polarization means that a society is divided into groups, with substantial intra-group homogeneity and inter-group heterogeneity (some researchers refer to polarization as the division into two groups). Such a state is considered as a source of tensions and social conflicts. Polarization rises when intra-group inequality decreases and inter-group inequality increases; nonetheless, total inequality in the society may diminish. Esteban and Ray (1994) propose the following index to measure polarization:

$$P_t^{(\text{ER})} = \frac{1}{\bar{y}_t} \sum_{r=1}^N \sum_{s=1}^N n_{rt}^{1+\alpha} n_{st} |y_{rt} - y_{st}|, \quad (13)$$

where n_{rt} and n_{st} are population shares of regions r and s at time t , and $0 < \alpha \leq 1,6$ is the degree of “polarization sensitivity” of the index (the index can be normalized to the median income $m(y_t)$ rather than to the mean one). It is easily seen that if $\alpha = 0$, polarization index (13) turns into the Gini coefficient weighted with population shares of regions (thus, α defines how far is the departure of the polarization index from inequality measurement). A researcher chooses the value of α a priori; sometimes, the degree of polarization is computed with a few values of α (e.g. $\alpha = 0.5, 1, 1.5$). Note that the Esteban-Ray index does not contain the number of region groups (clubs).

The Wolfson (1994) polarization index predetermines the presence of two groups. In essence, it measures the distance between a distribution under consideration and the “perfectly bimodal” one. It has the form

$$P_t^{(W)} = (0.5 - L_t(0.5) - 0.5G(y_t)) \cdot \bar{y}_t / m(y_t); \quad (14)$$

$L_t(0.5)$ stands for the value of the ordinate of the Lorenz curve at the median income.

The index due to Zhang and Kanbur (2001) assumes the composition of region groups to be predetermined (the number of groups may be arbitrary). It exploits decomposition of the Theil index into inter-group and intra-group inequality, being the ratio of these. Grouping of regions bases on features assigned by the researcher. These can be, e.g. poor and rich regions; extracting, manufacturing, and agricultural regions; the groups may represent federal districts or economic zones, etc.

One more aspect of studying income distribution dynamics is analyses of regions’ movement *within* the distribution – i.e. income mobility of regions – and its impact on changes in inequality. There are two concepts of mobility: relative (or rank) mobility and absolute (or quantity, or level) mobility. Relative mobility concerns changes in regions’ per capita incomes that results in switches of their position relative to one another on the income axis, i.e. reranking of regions or shifts of regions to other quantiles. The values themselves of the income changes are of no importance; the concern here is, e.g., whether poor regions can pass into the class of rich regions. Absolute mobility concerns changes in regions’ income levels themselves. That is, the interest here is with shifts of regions along the income scale irrespective of their relative positions; we do not care here whether some regions outrun or lag behind others. Note that the concept of absolute mobility also relates to the case when absolute changes of relative (to, say, the country average) incomes are dealt with. A different standpoint exists though. Some authors –

Bogomolova *et al.* (2001) and Formby *et al.* (2004) among them – regard this case as incorporating elements of both concepts.

A widespread tool of analyzing income mobility is transition matrix (or transition probability matrix) $\mathbf{P} = (p_{ij})$. To construct it, regions are divided into M income classes. A matrix element p_{ij} represents the fraction of regions having transferred from class i to class j over the time span under consideration, $[t, t + \tau]$. To put it otherwise, it is an estimate of probability of being a region in income class j at time $t + \tau$ on condition that it has been in class i at time t . A diagonal element p_{ii} shows the fraction of immobile regions, i.e. those remained in the same income class i . There are a number of mobility measures, most of which can be represented as a function of transition matrix: the total fraction of regions leaving their initial class (the normalized distance of \mathbf{P} away from the identity matrix), the distance between \mathbf{P} and perfect mobility, etc.; see Bogomolova *et al.* (2001) and Formby *et al.* (2004).

The difference between relative and absolute mobility lies in the way of constructing income classes. Analyzing relative mobility, these classes represent income quantiles. Thus, the number of regions in each class is always constant and equal to N/M . However, the boundaries of classes at times t and $t + \tau$ in income terms will be different in the general case. Considering absolute mobility, the entire possible range of incomes, e.g. $[\min(\min(y_{rt}), \min(y_{r,t+\tau})), \max(\max(y_{rt}), \max(y_{r,t+\tau}))]$ or $[0, \infty)$, is divided into M equal or unequal intervals representing the income classes. In this case, the boundaries of income classes are time-invariant, but the numbers of regions in classes vary over time; in particular, some income classes may turn out to be empty in the beginning or end of the time span.

The aggregation of regions into income classes results in the loss of information on intra-class mobility. To get rid of this in the case of relative mobility, Yitzhaki and Wodon (2004) propose to construct income classes so that each quantile contains one object (region), i.e. $M = N$. Then the number of quantile is the rank of region in ascending order of income; and the transition matrix contains in each row and each column exactly one unity, the rest of other elements equaling zero. This leads to a measure of relative mobility over time $[t, t + \tau]$ that Yitzhaki and Wodon refer to as the Gini symmetric index of mobility:

$$S_{t,t+\tau} = \frac{1}{2} \frac{G(y_t) \cdot (1 - \Gamma_{t,t+\tau}) + G(y_{t+\tau}) \cdot (1 - \Gamma_{t+\tau,t})}{G(y_t) + G(y_{t+\tau})},$$

where $\Gamma_{t,t+\tau} = \text{cov}(y_t, R(y_{t+\tau})/\text{cov}(y_t, R(y_t))$ and $\Gamma_{t+\tau,t} = \text{cov}(y_{t+\tau}, R(y_t))/\text{cov}(y_{t+\tau}, R(y_{t+\tau}))$ are the Gini correlation coefficients, $R(y_t)$ and $R(y_{t+\tau})$ are ranks, and $G(y_t)$ and $G(y_{t+\tau})$ are Gini coefficients. The attribute “symmetric” is due to the fact that $S_{t,t+\tau} = S_{t+\tau,t}$. The value of $S_{t,t+\tau}$ may vary from 0 to 1; the greater the value, the higher the relative mobility. Coefficients $\Gamma_{t,t+\tau}$ and $\Gamma_{t+\tau,t}$ are bounded within the band of -1 to 1, lesser values corresponding to higher mobility. In general, $\Gamma_{t,t+\tau}$ and $\Gamma_{t+\tau,t}$ are not equal. The equality holds if the shapes of distributions $f_t(y_t)$ and $f_{t+\tau}(y_{t+\tau})$ are similar. It is worth noting that these coefficients, hence the mobility index $S_{t,t+\tau}$, are not sensitive to monotonic transformations of the distributions. Such a transformation can be decreases or increases of inter-regional income gaps, suggesting income convergence or divergence of regions. Changes of this kind are captured by absolute mobility.

A methodology of analysing absolute mobility suggested by Quah (1993, 1996) provides a way of obtaining rich information on the evolution of income distribution that characterizes both its changing external form and intra-distribution dynamics. Some characterization of absolute mobility Λ enables one to derive the law of motion of the income distribution $f_{t+\tau}(y_{t+\tau}) = \Lambda \cdot f_t(y_t)$, since Λ is an operator mapping income distribution at time t into distribution at time $t + \tau$. With such a law in hand, one can ask about the long-run behavior of the distribution: whether it tends towards income equality or some constant inequality or polarization, etc. Assuming Λ to be time-invariant (i.e. considering the evolution of the distribution as a Markov process), the application of the above transformation n times yields a distribution at time $t + n\tau$, that is, $f_{t+n\tau}(y_{t+n\tau}) = \Lambda^n f_t(y_t)$. Taking $n \rightarrow \infty$ yields the ergodic distribution, $f_\infty(Y)$, such that $f_\infty(Y) = \Lambda_\infty f_\infty(Y)$, where Λ_∞ is the limit of Λ^n with $n \rightarrow \infty$. The ergodic distribution is the long-run limit of the distribution of incomes (under the assumption the underlying transition mechanism is time-invariant). Judging from unimodality or multimodality of the ergodic distribution, one can conclude whether there is a trend towards club convergence (polarization).

In the discrete version of the above framework in Quah (1993), transition matrix $\Lambda = \mathbf{P}'$ (\mathbf{P}' is the transpose of \mathbf{P}) characterizes absolute mobility; vector $\mathbf{f} = (f_i)$ represents income distribution (f_i is the fraction of regions being in income class i). To estimate Λ_∞ , relationship $\Lambda_\infty \approx \Lambda^n = (\mathbf{P}^n)'$ with a sufficiently great n is exploited (practically, n has the order of 10^1 , sometimes 10^2 , depending on the number of income classes and the criterion of convergence of Λ^n to Λ_∞). All rows of matrix Λ_∞ are equal to one another and to transposed vector of the ergodic distribution \mathbf{f}_∞ . A switch to the

continuous representation of incomes as in Quah (1996) makes it possible to avoid the loss of information because of discretization. In this case, absolute mobility is characterized by a stochastic kernel which can be considered as the probability matrix with infinite number of rows and columns. It is a probability density of incomes at $t + \tau$ conditional on incomes at t : $\Lambda = f(y_{t+\tau} | y_t)$.

Then $f_{t+\tau}(y_{t+\tau}) = \int_{-\infty}^{\infty} f(y_{t+\tau} | y_t) f_t(y_t) dy_t$. The stochastic kernel is estimated in a manner like the

univariate distributions above, taking into account the fact that $f(y_{t+\tau} | y_t) = f(y_{t+\tau}, y_t) / f(y_t)$. It shows where different parts of the initial distribution move over time. The line $y_{t+\tau} = y_t$ is the immobility line, peaks (local maximums of Λ) along which meaning convergence clubs.

The methodology discussed above yields a pattern of income distribution dynamics but does not provide an explanation for it. To address this issue, Quah (1997) proposes a way of incorporating additional factors. Its essence is as follows. Conditional incomes per capita in regions are estimated as $y_{rt}^o = y_{rt}/u_{rt}$, $u_{rt} = \sum_{s \in C(r,t)} \omega_{rst} y_{s,t-\theta(r,t)}$, where $C(r, t)$ is the collection of regions

functionally associated with region r (e.g., through trade or neighborhood), ω_{rst} is the relative strength of region s in affecting the evolution of region r , and $\theta(r, t)$ is a lag representing the delay with which region r is affected by the development of regions in $C(r, t)$. In words, the conditional income in the region is income per capita in it relative to the weighted average of lagged incomes per capita in functionally related regions. Thereafter the stochastic kernel mapping the unconditional distribution to the conditional one, $\Lambda^o = f(y_{rt}^o | y_t)$, is estimated. Its structure reveals some reasons behind features of income distribution. Conditioned income distributions can also give information on dynamics, if one analyzes absolute conditional-income mobility, estimating stochastic kernel $\Lambda^{oo} = f(y_{t+\tau}^o | y_t^o)$.

As in the case of cross-sectional regression analysis, spatial effects can affect estimates of the distribution evolution. Rey (2001) modifies the Quah methodology so that it includes an explicit description of these effects. His central idea is a spatial transition matrix that conditions a region's transition probabilities on the income class of the region's neighbors (in the continuous version, $\Lambda = f(y_t | y_{t(-1)})$). Magrini (2004) proposes one more way, directly conditioning incomes on spatial effects.

4. STUDIES ON INCOME INEQUALITY AMONG RUSSIAN REGIONS

This section reviews key papers on the issue of inter-regional income inequality in Russia. The review is not exhaustive, as some studies certainly remain unknown to me. Besides, the review does not cover a fairly great number of publications, where “analysis” consists in rendering and “systematization” of statistical data. The studies considered below are grouped by technique of analysis. However, this grouping is rather relative, since most of papers apply a few techniques that not infrequently are associated with different approaches (as they are defined in the previous section). By region is meant a subject of the Russian Federation. Some papers treat the composite subjects of the Russian Federation as single regions, i.e. do not consider the autonomous *okrugs* separately. Throughout this section income indicators (gross regional product – GRP, personal incomes, etc.) are per capita ones. If a measure of income inequality used to analyze σ -convergence is not indicated, the standard deviation of log incomes is meant.

4.1. Papers based on the cross-section approach

Papers exploiting cross-section analysis make up the most plentiful group. Most of them focus on the issue of β -convergence.

Mikheeva (1999, 2000) employs data on GRPs and regional personal incomes in 1990-1996. To estimate GRP, she applies an original methodology. Taking GRPs for 1994 from the official statistical sources as a base, GRPs for other years are computed with the use of sectoral volume indexes in respective regions. Such a methodology provides temporal comparability of GRP, but not cross-regional one (as the base GRPs are not spatially comparable). Thus, GRPs obtained are in fact nominal. Personal incomes are transformed to real terms with the use of regional consumer price indexes (CPI). The author fails in estimating parameters of the Solow model, concluding herefrom that the neoclassical growth theory does not explain dynamics of the transitional process. Nonetheless, her further analysis bases on the concept of β -convergence.

Dividing regions into two income classes of those with incomes above the Russian average (“rich”) and those with incomes below the average (“poor”), their transition between classes over 1990-1996 is considered. This, in fact, is equivalent to constructing a 2×2 transition matrix. Collections of regions relating to its elements form four groups. Measuring income by GRP, 21% of rich regions transited by 1996 into the class of poor regions; no one region transited into the class of rich regions. The regression analysis suggests unconditional β -convergence neither in the entire

cross-section nor in any of the four groups. No σ -convergence is found as well. Analyzing personal incomes yields the same results, except for the group of regions having been poor both in 1990 and 1996 that exhibits statistically significant β -convergence. (In the case of personal incomes, 64% of initially rich regions turned out to be poor in 1996, while 10% of initially poor regions transited into the class of rich regions).

The above analysis is supplemented with analyzing determinants of income inequality. Regression models of the form (9) are applied, in which the variable to be explained, $\mathbf{z} = (z_r)$, is a deviation of income from the national average: $z_r = \ln(y_{r,1996} / \bar{y}_{1996})$. In the case of GRP, this analysis suggests a positive impact of the initial (in 1990) level of GRP, investment, and expenditures of the regional budget. A negative impact pertains to the proportions of agriculture and the service sector in GRP, and belonging a region to one of the following economic areas: Central Black-Soil, Volga-Region, Northern Caucasus, and Far Eastern. In the case of personal income, a positive impact have the initial level of income, the proportions of industry and the service sector in GRP, investment, expenditures of the regional budget, inflation rate, and belonging to the Northwestern Economic Area. The impact of the proportion of agriculture in GRP, belonging to either the Volga-Region and Far Eastern economic areas, and (surprisingly) volume of exports per capita is negative.

Carluer and Sharipova (2004) employ three indicators in nominal terms: personal income in 1985-1999, GRP in 1994-1999, and industrial output in 1995-2000 by region. In the case of the first and last indicators, no unconditional β -convergence is found, while it takes place in the case of GRP. Analyzing σ -convergence yields similar results (dispersion of per capita GRP over 1994-1999 turning out to be relatively stable). At the same time, GRP and industrial output exhibit conditional β -convergence. Among conditioning variables (i.e. those responsible for differences between regional paths of equilibrium growth), education efforts, public expenditures, and financial aid prove to be statistically significant in the regression of personal income. In the regression of GRP, those are investment, education efforts, financial aid, and health level. Investment, financial aid, and social policy are statistically significant in the regression of industrial output. However, a reader has no choice but to guess what indicators quantify the above variables (and even what some of them mean), as the article is silent in this respect.

The authors also analyze conditional β -convergence from an economic geography standpoint,

assuming spatial differences in economic growth to be due to geographical locations of regions. The distance to Moscow characterizes location in the regression of personal income. This regression suggests no conditional β -convergence. In the regressions of GRP and industrial output, geographical factors are characterized by dummies of being a frontier region and location in North, or South, or West, or East (according to authors' classification). GRP does not exhibit conditional β -convergence, while industrial output does. Curiously, the relationship between the growth rate and the location in South or West proves to be negative. Albeit the results reported in the paper are discrepant, the authors conclude that regional divergence is a real phenomenon in Russia.

Solanko (2003, 2006) examines the evolution of personal incomes in 1992-2005 (the first publication covers 1992-2003). Real incomes are estimated with the use of regional CPIs like in Mikheeva (1999, 2000). Regions are divided into poor and rich ones. A region was classified as poor if income per capita there divided by the regional cost of a 19-staples basket was less than one third of the national average in 1992. This class contains 28% of regions. No σ -convergence is found during the period under consideration, except for 1998. Income inequality fell dramatically in that year; however, it reached pre-crisis values after a few years and then exceeded them. Such a pattern holds for the entire cross-section and for both poor and rich regions. While Solanko's results conform to those obtained by Mikheeva (1999, 2000) in this respect, they fundamentally differ regarding unconditional β -convergence, suggesting it for the whole 1992-2005 period and for different sub-periods. Analyzing panel data suggests β -convergence as well. It holds also for classes of poor and rich regions, being found in 1995-2005 among poor regions, and in 1992-2005, 1992-2002, 1995-2005, and 1999-2005 among rich regions. Analyzing conditional β -convergence, the following variables serve for conditioning: distance to Moscow (found insignificant), university graduates per 1,000 inhabitants (as a proxy of investment in human capital; it is also found insignificant), share of extractive industries in the total industrial output, share of workforce employed in agriculture (its sign is found negative), the number of small and medium-sized enterprises per capita (as a proxy of regional policies). Conditional β -convergence over 1992-2005 is found among all regions as well as among both poor and rich regions.

Ledyayeva and Linden (2008) analyze conditional β -convergence with the use of panel data on nominal GRP across Russian regions in 1996-2005. They use domestic investment per capita, foreign direct investment per capita, export per capita, "resource index," and a dummy for the year

1998. The “resource index” is the arithmetic mean of production of electricity, black metals, oil, gas, and coal relative to the cross-sectional averages. The analysis suggests conditional β -convergence. The most important factor of economic growth proves to be domestic investment, while foreign direct investment seems not to have an effect. There is some small evidence that export is positively related to growth; however, this result is not stable. The variable for the 1998 financial crisis, as should be expected, is negatively related to economic growth. A fairly unexpected result is very weak effect and even statistical insignificance of the resource index. Considering sub-periods of 1996-1999 and 2000-2005, the results for the former are qualitatively similar to those for the whole period, while in the latter sub-period, not only no β -convergence takes place, but all variables turn out to be insignificant. No β -convergence is found among both poor and rich regions (those with GRP below and above the national average).

Melnikov (2005) considers the evolution of GRPs over 1995-2003. In more recent papers, Melnikov (2007, 2008) extends this period to 2004 and includes the evolution of regional personal incomes over 1997-2005 into consideration. He affirms that the nominal values are adjusted to real ones by dividing by the cost of the fixed basket of goods and services for inter-regional comparison of population’s purchasing capacity (hereafter, the PC basket) across regions. However, as the Russian Statistical Agency, Rosstat, publishes data on the costs of the PC basket only since 2002, the way of computing real incomes for earlier years remains an enigma. The Theil index serves as a measure of income inequality. Taking GRP, inequality among regions rose during 1995-2000 (except in 1998), then decreased since 2001, although an increase was observed again in 2004. The decomposition of inter-regional income inequality to inequality within the federal districts and inequality between the districts suggests that the rise in the total inequality is mainly due to increasing inequality within the districts. The author proposes his own economic zoning of Russia, dividing its territory into 11 “macroregions.” Then the pattern for 2000-2004 changes dramatically. Inequality among macroregions yields the most part of the overall inter-regional inequality, inequality within macroregions remaining practically time-invariant. Prior to analyzing β -convergence, spatial autocorrelation is tested. Albeit the test suggests its presence, this fact is disregarded in the further analysis, not incorporating spatial effects into econometric models.

Regarding GRP, the author analyzes conditional β -convergence over 2000-2004, using dummies for belonging a region to federal district or macroregion as conditioning variables. Convergence is found in both cases; however, all conditioning variables are jointly insignificant in

the case of federal districts and jointly significant in the case of macroregions. Regarding personal incomes, their evolution drastically differs from that of GRPs and fundamentally contradicts to that reported by Solanko (2008). Throughout 1997-2005, σ -convergence is found (with small deviations from this trend in 2003 and 2004). The same holds for intra-territory and inter-territory inequalities, be the territories federal districts or Melnikov's macroregions. Conditional β -convergence over 1997-2005 manifests itself with conditioning by federal districts as well as by macroregions. However, all conditioning variables are jointly insignificant in both cases, which suggests unconditional convergence (and it is found, indeed).

A voluminous study by Lugovoy *et al.* (2007) deals with real GRPs over 1998-2004, computing these on the basis of estimates for 1999 reported by Granberg and Zaitseva (2002). The spatial sample lacks the Chukchi Autonomous Okrug and the Republic of Ingushetia thought of as outliers. While analyzing σ -convergence (which covers a longer time span, 1996-2004), four indicators measure income inequality: the coefficient of variation, Gini coefficient, inter-quartile range, and inequality range (the difference between maximal and minimal values of the log per capita GRP). This paper is the only one which tests for statistical significance of changes in inequality (both annual and over longer time spans, in particular, 1996-2003). The test does not reject the hypothesis of the equality of the coefficient of variation across time. Thus, the data exhibit neither σ -convergence nor σ -divergence. Like Melnikov (2005, 2007, 2008), the authors reveal spatial autocorrelation, but, contrastingly, this finding motivates the application of spatial econometrics techniques.

Two spatial weight matrices are used to construct spatial lags. The first looks like $\mathbf{W}_{(1)} = (\omega_r/T_{rs})^2$, where T_{rs} is the minimal traveling time required to cover the distance between capitals of regions r and s by highways, and ω_r is a normalizing factor. The second matrix is computed as $\mathbf{W}_{(2)} = (\omega_r Y_{rs}/T_{rs})^2$, where Y_{rs} is GRP in region s if it is a neighbor of r , and $Y_{rs} = 0$ otherwise. Analyzing unconditional β -convergence¹² with the use of a model of the form (10) with spatial weights $\mathbf{W}_{(2)}$, the estimate of β has the level of statistical significance of 7%. The authors deem such a level insufficient to evidence β -convergence. The estimate of the coefficient on spatial lag equals 0.46, suggesting strong spatial interaction of regions. For example, a 1-percent increase in the growth rate in all neighboring

¹² Strictly speaking, it cannot be called unconditional, since conditioning does take place here. The growth rate of a region is conditioned by the spatial lag, i.e. a somehow weighted sum of the GRP growth rates in neighboring regions.

regions will result in an almost 0.5-percent increase in the growth rate in the given region.

A model of the form (11) with spatial weights $\mathbf{W}_{(1)}$ is applied for analyzing conditional β -convergence. The estimate of β is significant at the 1-percent level. The conditioning variables are the share of fuel industry in region's industrial output (which is positively related with the GRP growth rate), per capita financial aid from the federal government, and a dummy for depressed regions (the latter two are negatively related with the GRP growth rate). Conditional convergence is also analyzed without taking account of spatial autocorrelation, but with additional inclusion of variables characterizing inter-regional links and human capital. These are availability of a non-freezing seaport, per capita volume of railway passenger traffic, and postgraduate students per 10,000 inhabitants. All these variables prove to be positively related with the GRP growth rate. Their inclusion eliminates spatial autocorrelation. The study under consideration also includes constructing and estimating a model of growth in Russian regions with the use of a rich variety of factors, and application of a sophisticated technique of analysis. Although this subject is closely adjacent to our topic, I am forced to content myself with merely mentioning it, as discussing it would take too much room.

Kholodilin, Oshchepkov, and Siliverstovs (2009), like the above authors, consider the evolution of real GPRs (computed in a similar way), however, over a longer period, 1998-2006. (Along with this, they use nominal GRPs deflated to the 1998 prices. Although they place emphasis on results obtained with these data, we will not discuss them.) In addition to the Chukchi Autonomous Okrug and the Republic of Ingushetia, the Republic of Kalmykia is also excluded from the sample. Like in Lugovoy *et al.* (2007), the regression analysis bases on spatial econometrics technique, using a spatial weight matrix of the form $\mathbf{W}_{(i)} = (\omega_r \cdot q_i(L_{rs})/L_{rs}^2)$, where ω_r is a normalizing factor, L_{rs} is the great circle distance between capital cities of regions r and s , and $q_i(L_{rs}) = 1$ if $L_{rs} \leq l_i$, and $q_i(L_{rs}) = 0$ otherwise, l_i being the i th quartile of the distance distribution, $i = 2, 3, 4$. As follows from the text of the paper, only one of these matrices is actually employed, but it is unclear which one. A sufficient dissimilarity from Lugovoy *et al.* (2007) is in that β -convergence is analyzed not only over all regions, but also over each of four groups corresponding to the quadrants of the Moran scatterplot (*HH*, *HL*, *LL*, *LH*). These groups are fairly stable over time. In 1998, group *HH* (rich regions surrounded by rich ones) and group *LL* (poor regions surrounded by poor ones) contained 62% of regions. Thus, regions with high (low) per capita GRP are located, for the most part, adjacent to like ones.

Unconditional β -convergence is analyzed with the use of models of the forms (10) and (11). Among all regions, no convergence occurs (the estimate of β has the expected sign, but is statistically insignificant), as well as among regions in groups *HL* and *LH* (in the latter, β has the wrong sign). However, β -convergence is found in groups *HH* and *LL*, its rate in group *HH* being almost twice as high as in *LL*. Based on the logic of the β -convergence concept, this would imply that the income gap between regions of these two clusters is going to rise. The elasticity of growth rate vis-à-vis its spatial lag is equal to 0.32-0.45, which is close to the estimate obtained by Lugovoy *et al.* (2007). The authors do not formally analyze conditional β -convergence, but they qualitatively examine what differentiates regions across groups and, in particular, what makes regions from group *HH* to stand out from the rest of regions. For this purpose, they compare investment, GRP structure, population, labor, human capital, foreign trade, and natural conditions. Regions from group *HH* take a leading position in all investment and foreign trade characteristics; they are also characterized by a relatively high share of industrial production in GRP as well as by the lowest share of agriculture. Group *LL* is the only one that has a positive migration balance; its regions also possess more favorable natural conditions. Curiously, the average temperature in group *HH* is lower than in group *LL* by almost 7°C in January and by 3°C in July. Comparing all rich regions with all poor ones, the temperatures in the former are lower by 5°C and 3°C, respectively. Thus, the worse the climate in a region, the more efficient the region's economy, which does not support the main assertion in much-talked-of book by Hill and Gaddy (2003).

Buccellato (2007) also applies spatial econometrics, examining the evolution of nominal GRPs over 1999-2004. Like in the above paper, the estimation of models of the forms (10) and (11) with a neighborhood matrix suggests the absence of unconditional β -convergence. Models of the same forms are applied to analyzing conditional β -convergence. Based on different econometric characteristics, the spatial lag model (10) is found more suitable. Both models suggest conditional β -convergence under conditioning with the share of oil and gas production in GRP, ratio of exports and imports to GRP (proxying openness to trade), share of population employed in research and development, and per capita foreign direct investment.

Among studies applying cross-section analysis, there are a number of those not based on the β -convergence concept. Their purpose is to explain reasons for inter-regional differences in efficiency of economic growth, in other words, to find determinants of income inequality of regions.

Berkowitz and DeJong (2002, 2003, 2005) evaluate links between growth of personal incomes and initial conditions, reform policies, and entrepreneurial activity across Russian regions. The first two papers cover 48 regions and the period of 1993-1997, somewhat differing in explanatory variables and econometric techniques. In the latter paper, the number of regions increases to 70; and the period under consideration expands to 2000 (to verify results of the first two papers, the analysis is also performed for 1993-1997). To adjust growth of nominal incomes to that of real ones, regional CPIs are used (while the initial levels of real incomes are assessed with the use of the cost of a staples basket). The number of small private enterprises in place in each region as of December 1995 (relative to the regional population) quantifies entrepreneurial activity. To avoid endogeneity (a reverse causality between entrepreneurial activity and economic growth), the paper of 2005 applies the method of instrumental variables. In the first stage, a regression of entrepreneurial activity on a number of variables is estimated. Of these variables, initial income, human capital (measured as the share of adults that received at least some postsecondary training), and the reformist orientation of the population (measured as the share of the population that voted for pro-reformist candidates in the December 1993 parliamentary elections) exhibit a positive and statistically significant correspondence with entrepreneurial activity. In the second stage, income growth is regressed on the instrumented entrepreneurial activity and a number of other variables (of which only region's initial production potential proves to be statistically significant).

The main finding of all the three papers resides in that regional variations in the adoption of economic reforms appear to be an important factor in accounting for regional variations in new-enterprise formation, which in turn exhibits a strong positive correspondence with personal income growth. In a subsequent study, Berkowitz and DeJong (2010) apply a different way of assessing real incomes across regions. They normalize nominal incomes using the relative cost of the PC basket across regions. For the end of 2007, these are actual costs; for earlier periods, they are deflated with the use of regional CPI data. Besides, “atypical” Moscow and the Tyumen Oblast are excluded from the region sample. Nonetheless, the results for 1993-2000 differ only slightly from those obtained in Berkowitz and DeJong (2005). However, the pattern for 2000-2007 proves to be quite different. Since 2000, bank-issued became an important engine of growth (its regional variations being a source of inter-regional differences in growth rate), and not entrepreneurial activity.

Berkowitz and Jackson (2006) examine the impact of small enterprises on changes in income distributions, covering 66 Russian regions (Moscow and St. Petersburg are omitted from the

sample). The variable to be explained is the change in the income share of the bottom two quintiles (40% of the population) within regions over 1995-2001. The variable measuring small enterprise formation is represented by three versions: the employment share of small enterprises in 2001, the number of de novo private firms per thousand people in 1995, and the change in the de novo firm share of private employment from 1995 to 2001. To address the potential endogeneity problem, large and small scale privatizations in 1993 are used as instruments for this variable. Regressors in the second-stage regression include, along with instrumented small enterprise development (in one or another version), the initial income share of the bottom 40% of the population, region's population, and education. The results obtained suggest that the development of small enterprises causes a decrease in income inequality, increasing the share of income earned by the two lower quintiles (by 1.25% as the share of the workforce in small enterprises increases by one standard deviation). A similar analysis for Poland yields a close estimate. However, the comparison scores against Russia. While the small private enterprise sector dynamically developed in Poland, it actually decreased between 1995 and 2001 in Russia. This is a major reason why the income distributions in Poland and Russia diverged so dramatically during the 1990s.

The aim of Ahrend's (2005) study is to investigate the main causes behind large differences in regional growth performance in Russia for the period from the start of transition to the 1998 crisis. In doing so, a panel of 77 regions is used, measuring regional performance by GRP and personal income, as well as by industrial production in order to cover the whole period under consideration (since official data on GRP and real income growth are available only since 1994). GRPs are transformed into the 1994 constant prices by using the national sectoral deflators. Data on real income and industrial production growth are taken from official statistical publications, where regional CPIs are used to adjust nominal incomes, while annual growth of industrial production is computed with the use of prices in a base year.¹³ A great number of variables, divided into three groups, serve as explanatory ones. The first group characterizes politico-institutional features of regions, among them governors' political orientations, measures of institutional efficiency, and proxies for the risk of violent conflict. The second group includes indicators of economic reform, such as the degree of privatization and price liberalization, or the level of subsidies. The third group describes initial conditions, including economic, geographical and structural features of regions.

¹³ Goskomstat (1998), pp. 162-169, does not report whether national or regional prices are used to compute the industrial-production growth in regions. Anyway, cross-region comparability of this index is vague.

The econometric analysis is organized in the following way. A hypothesis regarding the impact of some factors on economic growth is formulated. This hypothesis is tested, applying a panel regression that includes a relevant variable subgroup from the above groups as well as “core variables” used in all regressions (human capital, initial income or production, and the Moscow dummy). In total, ten hypotheses are tested. The results suggest that the initial competitiveness of a region’s industry, as measured by the share of exports in regional production, is the most significant and robust explanation for differing regional performance. Industrial structure, resource endowment, as well as natural resources and human capital, also played an important role. Urbanization had a positive impact on growth performance, too. An interesting part of the results is what did not matter. Political variables (such as the governor’s political orientation) prove to be among such factors. Institutional features and the degrees to which regions implemented economic reforms played a small role in explaining differences in performance of regional economies.

Mitsek and Mitsek (2009) consider factors that influence nominal personal incomes in Russian regions. They analyze a panel of three cross-sections (for 2004, 2005, and 2006) across 68 regions. The authors exclude the Northern Caucasian republics, the republics of Kalmykia, Tuva, Khakasia, and Altai, as well the Sakhalin Oblast from the sample, motivating this by “features of their economies.” Income is regressed on capital-labor ratio, labor productivity, capital productivity, proportion of employed people in region’s population, share of exports to non-CIS countries in GRP, proportion of workers with university education, and relative contributions of eight sectors to GRP. According to the results obtained, capital-labor ratio, labor productivity, and capital productivity are responsible for three quarters of inter-regional income gaps (capital-labor ratio being responsible for almost a half). It should be noted, however, that the adequacy of computing the contribution of variable x_i as $\hat{\alpha}_i(x_{i\max} - x_{i\min})/(y_{\max} - y_{\min})$, where $\hat{\alpha}_i$ is the estimated coefficient on this variable, casts doubts. Of variables characterizing structure of region’s economy, estimates are reported only for three, namely, contributions of agriculture, extractive and manufacturing industries to GRP. All the three estimates prove to be negative. A negative relationship between share of extractive industries in regional economy and personal income seems strange and contradicts to results of the above-discussed studies.

4.2. Papers based on the time series approach

Time series analysis does not find wide application in examining spatial income inequality. This holds as well for studying cross-regional income inequality in Russia. I have found only two relevant papers that use this approach for analyses.

Babetski and Maurel (2002) analyze monthly time series of personal incomes over February 1995 through November 1999. Although the incomes are referred to as real, no mention is made of how they are estimated. A deviation of income growth over a 12-month period from the cross-sectional average, $d_{rt} = \ln(y'_{rt}/y'_{r,t-12})$, serves as the dependent variable, where $y'_{rt} = y_{rt} / \bar{y}_\tau$, and \bar{y}_τ is the geometric average over a territorial unit $\{r\}$ included in a given panel. The authors use 11 panels corresponding to economic areas (and consisting of time series across respective regions) and a panel consisting of time series aggregated over economic areas. The authors suppose dynamics of d_{rt} to be an autoregressive process $d_{rt} = \rho d_{r,t-1} + \varepsilon_{rt}$, and test each panel $\{d_{rt}\}$ for stationarity. Thus, although the proclaimed aim is analyzing convergence, actually only short-run properties of the income evolution are analyzed (see Section 3.2). In essence, the technique of the analysis is similar to that proposed by Evans and Karras (1996). Stationarity is found in all cases. Hence, in the period under consideration, spatial differences in per capita income growth rates were permanent (accurate to random disturbances) both within the economic areas and between them. An interesting feature of this paper is an analysis of the role played by institutional factors. In doing so, autoregressive coefficient ρ in the regression of d_{rt} is represented as a function of static variables x_{ri} characterizing such factors: $\rho = \rho_0 + \alpha_1 x_{r1} + \dots + \alpha_m x_{rm}$. These variables are price regulations, small-scale privatization, and subsidizing producers. In this case, the panel of economic areas and that of all regions are analyzed. The results obtained suggest that price regulations and subsidizing producers decreases the speed of short-run convergence, while small-scale privatization increases it.

A study due to Kwon and Spilimbergo (2005) is quite non-typical. Analyzing time series over 1993-2002, they consider the issue of a dependence of changes in regional incomes on migration and fiscal policy rather than that of income convergence among regions. Unfortunately, the description of the raw data is very vague. It seems that GRP per capita is used to measure income (for the early years of the transition, according to IMF estimates), and real incomes are assessed with the use of some regional price indexes. To analyze a relationship between income and geographical mobility of labor, the authors use a panel vector autoregression which includes income

and regional population with three-year lags as well as regional and year dummies. An impulse response function obtained suggests that Russia's regional populations weakly respond to income shocks, which is consistent with other evidence of low labor mobility in Russia.

A relationship between regional fiscal policy and income is also analyzed with the use of a panel vector autoregression comprising income, fiscal surplus, and "oil shock" (oil share in regional income in the previous year multiplied by oil price in the current year). The findings demonstrate that regional governments have used procyclical fiscal policy, expanding regional expenditures in booms and contracting them in recessions. The authors ascribe this to very limited discretion of regions in the formulation and conduct of fiscal policy. Given this, the federal government should take the responsibility for mitigation of regional income shocks. To check whether this is the case, a cross-sectional analysis is applied. The authors estimate a regression of the ratio of net federal transfers to income on per capita income, oil shock, and net revenues per capita for each year in the period under consideration. Judging from these estimates, federal transfers do not seem to play a significant role in absorbing regional shocks. However, since about 2000, their effect was becoming notably stronger.

4.3. Papers based on the distribution dynamics approach

This group of studies is relatively large and varied. A number of them dispense with formal tests, drawing inferences from plots or tables of inequality indicators.

Bradshaw and Vartapetov (2003) consider the problem of regional inequality in Russia in a wider context than that of only income disparities, dealing with a number of consumption variables, social and health indicators, etc. As to income variables, they use GRPs (those for 1990-1993 are drawn from Mikheeva, 1999) and personal incomes (adjusting them to real terms with the use of the cost of the staples baskets or the subsistence minimum). Considering 1990-2001, the authors compute a number of inequality indicators (the average absolute deviation from the mean, unweighted and weighted standard deviation, Lorenz curve-based coefficient of concentration, and Theil index) and trace their change over time (although, without a formal test). In general, all indicators suggest divergence that is substantially weaker for GRPs than for personal incomes. However, excluding Moscow, the Tymen Oblast, and the Republic of Ingushetia from the sample, the pattern changes dramatically. Personal income convergence occurred among the remaining regions during 1990-1997, following which weak divergence started.

A regression of growth rates of real personal income over 1994-1999 on the 1994 income level suggests that the higher the initial income, the higher the growth rate (which rejects the β -convergence hypothesis). The authors also address the issue examined by Kwon and Spilimbergo (2005): Do federal fiscal transfers in Russia facilitate mitigating inter-regional income inequality? During 1995-1999, the shares of the federal transfers in regional revenues had a strong negative correlation with regional real per capita tax revenue. This is evidence for transfers of federal financial support to the poorest regional economies. A comparison of inequality in regional per capita tax revenues before and after transfers suggests that transfers lower it by a factor almost equal to 1.5. At the same time, federal transfers proved to be incapable of preventing income divergence of regions.

Galbraith, Krytynskaia, and Wang (2004) exploit data on nominal payroll per employee across 14 economic sectors and all regions in Russia, covering 1990-2000. They measure inequality by the Theil index. Its evolution reveal income (payroll) divergence among regions as well as among sectors, the path of inter-regional inequality being qualitatively similar to that obtained by Bradshaw and Vartapetov (2003). An interesting feature of this paper is that both inter-regional and inter-sectoral contributions to the overall payroll inequality are estimated. For example, the Tymen Oblast provided the most contribution to inter-regional inequality since 1992; beginning in 1994, Moscow became the second most important in this respect. Comparing contributions of regions and within-region sectors to inequality in 1990 and 2000, striking changes are seen. First, payroll inequalities among the same sectors across regions rose dramatically over the decade. Second, sector ranks changed. Most notably, whereas in 1990 agricultural incomes were in the middle of the Soviet income distribution, a decade later they were at the bottom. Meanwhile, the finance sector had moved up, surpassing science and management. Third, high relative incomes in industrial production and construction in modern Russia are due to extraordinary relative gains in just a handful (five to six) of regions.

A similar analysis is performed for China. Despite radically different ways of transition to the market economy in Russia and China, the evolution of inequality manifests some similar features. In both countries, income divergence occurred, incomes diverging more sharply on a regional than on a sectoral basis; relative income rose most sharply in the financial and political centers and in the regions providing hard currency export earnings. Like in Russia, relative gains in the financial sector rose greatly in China. At the same time, fall in relative incomes in Chinese agriculture was

not as great as in Russia. Protection of education, health care, and science in China enabled them to keep or even improve their positions, while in Russia these sectors suffered absolute and relative losses.

Litvintseva, Voronkova, and Stukalenko (2007) analyze inequality by 20-percent income classes across 78 regions over 2000-2004. They estimate real personal incomes in the quintiles with the use of the regional costs of the PC basket for 2004 and apply regional deflators (it is not clear which ones) to adjust incomes to the 2004 prices. Inequalities within income classes (measured by the Gini coefficient) and, as a rule, the income growth rates over 2000-2004 are found to increase from lower to upper classes¹⁴ (inequality falling in the two lower classes and in the uppermost class and rising in the third and fourth classes). In 2004, the Gini coefficient ranged from 36.6% in the Far-Eastern federal district up to 50.3% in the Central district. When social transfers in kind are taken into account, inter-regional income inequality turns out to be lower (in 2004, the Gini coefficient for Russia as a whole equaled 42.9% before social transfers and 38.5% after them). Nonetheless, inequality rises in this case, too.

Fedorov (2002) focuses on polarization of Russian regions in 1990-1999. He uses data on personal incomes and expenditures converted to real terms with the use of regional CPIs. The Gini coefficient and the Theil index weighted by proportions of regions' population measure inequality. Both incomes and expenditures diverged during 1991-1996. In 1998-1999, the increases of income inequality leveled off and those of expenditure inequality even reversed. Trends of the Esteban-Ray and Wolfson polarization indexes, (13) and (14), are similar to those of inequalities, which suggests increasing polarization of Russian regions. To establish along which dimension polarization occurs, the author applies a modified Kanbur-Zhang index, defining it as a ratio of between-group inequality to the total inequality (recall that the Theil index quantifies inequality in the Kanbur-Zhang index). Four dimensions of polarization are analyzed. These dimensions include West vs. East (regions to the west of Urals vs. regions of Urals, Siberia, and the Russian Far East), national republics vs. ethnically Russian regions, regions with large capital cities (of more than 800,000 inhabitants) vs. regions with smaller capitals, and exporting regions vs. other regions (the former are defined as regions with export shares in the total country's export greater than 1% in 1997). Polarization in both income and expenditures along "West–East" and "national status" dimensions

¹⁴ In Russia as a whole, real income increased by 39% in the first quintile and by 62% in the fifth quintile over 2000-2004.

remained fairly constant over 1990-1999, while polarization along “capital city size” and “export share” dimensions increased sharply during 1992–1995.

A work of Balatskii and Saakyants (2006) stands somewhat apart. However, by convention it can be classed with studies on income polarization of regions. They propose a method of estimating a relationship between income inequality among region groups and country’s economic growth. GDP of the country, Y , is assumed to be equal to the sum of GRP. Then

$$Y = N \cdot ((1 - n_H)y_L + n_H y_H) = Ny_L \cdot (1 + n_H(\kappa - 1)), \quad (15)$$

where y_H and y_L are GRP per capita in the group of rich and poor regions, respectively, N is the total population, n_H is a proportion of population in rich regions, and $\kappa = y_H/y_L$ is the coefficient of income differentiation. Assuming N , n_H , and y_L to be time-invariant, growth rate and income differentiation are related by the following equation:

$$\frac{d \ln Y}{dt} = \frac{n_H \kappa}{1 + n_H(\kappa - 1)} \cdot \frac{d \ln \kappa}{dt} = \Omega \frac{d \ln \kappa}{dt}.$$

Parameter Ω , designated as “the coefficient of income erosion,” is the elasticity of GDP vis-à-vis income differentiation. It shows how many percent of the GDP growth are “absorbed” by a one-percent increase of income differentiation. Such an interpretation bases on the fact that, given that κ is fixed, the produced GDP may be used, according to Equation (15), to increase income in poor regions (y_L), or to enlarge the group of rich regions (n_H), or to support the increased country’s population (N). If κ rises, these three channels receive less funds (income “seized” due to rise in κ comes to rich regions), the value of Ω quantifying whither less.

This model is estimated for 1998-2001 and 2001-2004 with data on nominal GRPs. Moscow and the Tyumen Oblast are classed with rich regions; the rest of regions are deemed poor. The Russian GDP increased in 2001 by 23.9% relative to 1998. The coefficient of income differentiation equaling 34.2% gives $\Omega_{1998} = 0.24$. Herefrom GDP “absorbed” by income divergence over 1998-2001 amounted to 5.7%; then the GDP growth fallen to share of the bulk of Russia’s population came to 18.2%. Over 2001-2004, GDP rose by 22.0%, while the income gap increased by 2.5%, because of which the GDP growth for the greater part of the population was 21.2%.

Dolinskaya (2002) considers changes in the shape of cross-region income distribution and absolute income mobility of regions in 1970-1991 and 1991-1997 (Moscow and St. Petersburg are omitted from the sample). The data are personal incomes per capita normalized to the national average. For 1992-1997, real income by region is obtained by deflating the nominal figure by the

regional CPI. The estimated income distributions became progressively wider (so suggesting income divergence of regions) and skewed toward low incomes during 1991-1999. The evolution of income distribution during the pre-reform times, 1970-1990, was fundamentally different, being characterized by narrowing the distribution over time (albeit very slow) with a stable median. To analyze income mobility, Quah's (1993) transition matrix approach is applied. These matrices are constructed by dividing regions into five classes with relative incomes confined in the bands of $[0, 0.7)$, $[0.7, 0.9)$, $[0.9, 1.1)$, $[1.1, 1.3)$, and $[1.3, \infty)$. The transition matrix for the pre-reform period demonstrates that income mobility of regions in that time was fairly low; the probability of remaining in the same income class equals 0.68. In the ergodic (long-run) distribution, the middle income class contains 52% of regions, the class of poor (with relative income below 0.7) being empty. Income mobility in 1991-1999 is much higher than in 1970-1990; the extreme classes – the uppermost and lowermost ones – are persistent (the probabilities of remaining in the first and fifth classes are equal to 0.69 to 0.82 and 0.75 to 0.87, respectively). Such a structure of income mobility results in a highly uneven long-term income distribution, in which a bit more than one fifth of regions fall into the middle income class, while a half of regions fall into the two lower classes (where relative income is below 0.9).

The income mobility pattern obtained is used for examining impact of industrial structure and regional fiscal policies to inter-regional income inequality. The shares of processing and extractive industries in regional industrial output in 1995 characterize the industrial structure; federal transfers in total expenditures of a region and the share of subsidies to producers in total region's expenditures characterize fiscal policy. Two methods are applied here. The first one consists in constructing and analyzing matrices, an element of which is the average of the above indicators over regions contained in the relevant cell of the transition matrix. The second method uses a regression of region's transition between income classes on the above indicators (excluding transfers), change in the share of subsidies to producers in total region's expenditures over 1992-1997, initial income class of the region, and the number of new enterprises (created during 1992-1997) relative to the regional total. By and large, the obtained results suggest that rise in incomes in more successful regions was due to natural resource endowments, while less successful regions were trapped at relative low income levels due to uncompetitive industries and the lack of resources for their restructuring. Income dynamics during 1991-1997 were largely unrelated to the development of the new private sector, but were instead determined by operations of enterprises existed prior to the

transition.

Carluer (2005) also applies a technique due to Quah (1993), but he studies club convergence, interpreting income classes used to construct transition matrices as convergence clubs. The data used are nominal personal incomes over 1985-1999 across Russian regions; incomes for 1992-1999 are deflated to the 1991 prices with the use of the national CPI. Four income classes are defined so that they correspond to quartiles of the income distribution for 1985, whence the boundaries of the classes in income terms (in prices of those times) are computed. The transition matrix obtained has some qualitative features similar to those of Dolinskaya's (2002) matrices. Namely, the probability of remaining in the same income class is in general less than 0.5; however, the extreme classes are fairly persistent (the probabilities of remaining in the first and fourth classes equal 0.82 and 0.86, respectively). But, the ergodic distribution differs greatly. Almost two thirds of regions fall into the lowermost income class, the uppermost class contains less than one fifth of regions, and the two middle classes contain less than one tenth of regions each. The pattern found, indeed, suggests polarization of regions: the most part of them move to either "the poorest" or "the richest" class (as they are defined in this paper). Such a considerable disagreement with Dolinskaya's results seems to be caused by different ways of defining income classes rather than by the difference in data (nominal vs. real incomes).

Yemtsov (2005) studies the evolution of both inter- and intra-regional income inequality as well as poverty in 1994-2000. He uses personal per capita incomes drawn from household budget surveys across regions. Incomes are converted into the real terms with the use of regional subsistence minimums. Over 1994-2000, the overall income inequality in Russia, measured by the Theil index, rose from 0.297 to 0.339 (a 14% increase), inter-regional inequality rising by 47%, and intra-regional inequality rising by 3%. While the inter-regional inequality contributed 25% to the overall inequality in 1994, this contribution increased to 32% in 2000. Thus, inter-regional differences are playing an increasingly important role in determining overall inequality (they account for up to 85% of the total increase in inequality in Russia between 1994 and 2000). This pattern changes dramatically if one removes the richest regions, Moscow, St. Petersburg, and the Tyumen Oblast. The overall inequality in the group of the rest regions turns out to be sufficiently lower (0.192 in 1994 and 0.225 in 2000), but its rise over 1994-2000 increases to 17%. Along with this, inter-regional inequality rises by only 5%, while intra-regional one rises by 22%. The contribution of inter-regional inequality to the overall inequality remains fairly stable, decreasing

from 26% to 23%. Thus, the increased gap between the high-income regions and the rest of the country, and increasingly unequal income distributions within poorer regions feature the period of 1994-2000. Measuring inter-regional inequality by the standard deviation of log incomes, it remained relatively stable over 1992-1997, and increased afterwards. Such a path of σ_t considerably differs from that obtained by Mikheeva (1999, 2000).

To analyze income mobility of regions, transition matrices are applied. The division of regions into income classes is the same as in Dolinskaya (2002). The results differ, however. For example, probability of remaining in the uppermost income class is much lower, equaling 0.38. In the long-run income distribution, about three quarters of regions fall into three lower classes with circa equal share of regions in each class. At the same time, the pattern found is qualitatively similar to that obtained by Dolinskaya (2002). The author also uses a transition matrix for obtaining poverty projections under different scenarios of the evolution of income inequality. Besides, he gives much attention to analyzing intra-regional income inequality.

Lavrovsky and Shiltsin (2009) consider dynamics of nominal GRP over 2000-2005. The evolution of income inequality (measured by the coefficient of variation of GRP per capita) manifests divergence. However, when Moscow, the Tymen Oblast and Chukchi Autonomous Okrug (AO) are excluded, weak convergence and stabilization of inequality take place. A contribution of these three regions explains a major part of inter-regional inequality. A half of the coefficient of variation was due to Moscow and the Tymen Oblast in 2000, and to the Tymen Oblast and Chukchi AO in 2005. To estimate a transition matrix, five income classes are constructed, with GRP relative to the cross-sectional average confined in the bands of [0, 0.25), [0.25, 0.5), [0.5, 1), [1, 2), and [2, ∞). Income mobility of regions over 2000-2005 proves to be very low. Owing to this, the long-run income distribution differs only slightly from the actual distribution in 2005. No tendency to polarization of regions is found; 1.3% of regions fall into the lowermost income class, and 3% of them fall into the uppermost income class in the long-run distribution. At the same time, the long-run distribution is skewed toward low incomes, namely, three fourth of regions concentrate in the first three income classes.

Herzfeld (2008) studies the evolution of nominal GRP (relative to the cross-sectional average) over 1994-2004 and performs a similar analysis for China. Dynamics of two inequality measures, the Gini coefficient and the coefficient of variation, suggest evident divergence (their values catching up with these in China by 2002-2004). Omitting Moscow, the Tymen Oblast, Chukchi AO,

and Republic of Yakutia, income inequality of regions becomes on average stable over time, albeit with fluctuations. By and large, such a pattern of σ -convergence agrees well with that obtained by Lavrovsky and Shiltsin (2009). To analyze club convergence, kernel estimates of the cross-region GRP distribution are tested for multimodality. The results indicate that the hypothesis of bimodality cannot be rejected at the 10% significance level in five out of 11 years. However, an additional mode (sometimes, two) in the distributions is due to only Moscow, the Tymen Oblast, and Chukchi AO (and it seems that this group of regions can be hardly deemed a convergence club). In China, contrastingly, no convergence clubs emerge after 1982, although very rich regions also are there. Hence, a gap between them and the rest of China is not as great as in Russia.

Since the multimodality test cannot reveal whether regions move to two distinct convergence clubs, mobility of regions within the distribution is analyzed with the use of mobility matrix in order to find a trend to polarization. A division into six income classes is chosen. Denoting the standard deviation of the relative GRP by σ , the income band is $(0, 1 - 0.9\sigma]$ for the lowermost class and $(1 + 1.5\sigma, \infty)$ for the uppermost class, the rest classes being 0.6σ wide. Persistence turns out to be highest in the second lower and the uppermost income classes in both Russia and China. The estimated long-run GRP distribution for Russia is left-skewed (87% of regions fall into the three lower classes, a half of regions falling into the second lower class), but does not evidence polarization. For China, the pattern is quite different. The long-run distribution is bimodal; the modes correspond to the second lower and the uppermost income classes. Thus, Chinese regions tend toward polarization, a significant proportion of regions moving to the uppermost income class.

5. COMPARISON OF THE STUDIES

Some results of the reviewed papers agree with one another, while some others contradict. However, it is difficult to directly compare these results, as different papers vary in raw data, adjustment methods (transforming from nominal into real terms), time spans, and sometimes in region samples. Therefore it makes sense to consider these differences and reliability of the results, and then, leaning upon this, to attempt obtaining some general pattern of the evolution of inter-regional inequality in Russia. Table 2 summarizes main features of the reviewed papers: time spans under consideration, income indicators used, indicators applied for estimating real incomes, and basic methods of analyses.

Table 2. Summary of studies on inter-regional income inequality in Russia

Paper	Time span	Income indicator	Indicator of regional price level	Main method of analysis
Mikheeva (1999, 2000)	1990-1996	GRP, personal incomes	— CPI	β -convergence
Carluer and Shapirova (2004)	1985-1999 1994-1999 1995-2000	Personal incomes GRP Industrial production	— — —	β -convergence
Solanko (2006, 2008)	1992-2005	Personal incomes	CPI	β -convergence
Ledyanova and Linden (2008)	1996-2005	GRP	—	β -convergence
Melnikov (2005, 2007, 2008)	1995-2004 1997-2005	GRP Personal incomes	Cost of the PC basket Cost of the PC basket	β -convergence
Lugovoy et al. (2007)	1998-2004	GRP	From Granberg and Zaitseva (2003)	β -convergence
Kholodilin et al. (2009)	1998-2006	GRP	From Granberg and Zaitseva (2003)	β -convergence
Buccellato (2007)	1999-2004	GRP	—	β -convergence
Berkowitz and DeJong (2002, 2003, 2005)	1993-1997, 1993-2000	Personal incomes	CPI	Causal cross-sectional analysis
Berkowitz and DeJong (2010)	1993-2007	Personal incomes	Cost of the PC basket + CPI	Causal cross-sectional analysis
Berkowitz and Jackson (2006)	1995-2001	Personal incomes (by quintile)	—	Causal cross-sectional analysis
Ahrend (2005)	1995-1997 1994-1998	GRP Personal incomes	— CPI	Causal cross-sectional analysis
Mitsek and Mitsek (2009)	2004-2006	Industrial production Personal incomes	— —	Causal cross-sectional analysis
Babetski and Maurel (2002)	1995-1999	Personal incomes	CPI(?)	Time series analysis
Kwon and Spilimbergo (2005)	1993-2002	GRP(?)	CPI(?)	Time series analysis
[65] Bradshaw and Varapetov (2003)	1990-2001	Personal incomes, GRP	Cost of staples basket, subsistence minimum	σ -convergence
Galbraith et al. (2004)	1990-2000	Payroll	—	σ -convergence
Litvintseva et al. (2007)	2000-2004	Personal incomes (by quintile)	Cost of the PC basket + CPI(?)	σ -convergence
Fedorov (2002)	1990-1999	Personal incomes, personal expenditures	CPI	Polarization
Balatskii and Saakyants (2006)	1998-2004	GRP	—	Polarization (?)
Dolinskaya (2002)	1991-1997	Personal incomes	CPI	Transition probability matrices
Carluer (2005)	1985-1999	Personal incomes	—	Transition probability matrices
Yemtsov (2005)	1994-2000	Personal incomes	Subsistence minimum	σ -convergence, transition probability matrices
Lavrovsky and Shiltsin (2009)	2000-2005	GRP	—	Transition probability matrices
Hertzfeld (2008)	1999-2004	GRP	—	Convergence clubs, transition probability matrices

Of the 31 papers, 21 deal with personal income and 16 deal with GRP, seven papers dealing with both indicators. The dual meaning of the term “real” regarding monetary indicators, when the case at hands is their comparability both over time and across regions, manifests itself in many papers, especially in those analyzing GRP. Some of them refer to personal incomes or GRP as real when these are deflated to the average Russian prices of a base year. But from the standpoint of inter-regional comparisons they still remain nominal, not taking into account different purchasing power of ruble across regions. Almost all papers that use personal incomes make attempts to provide cross-region comparability of this indicator, except for Carluer (2005), Carluer and Sharipova (2004), and Mitsek and Mitsek (2009), and Berkowitz and Jackson (2006). Note however, that dealing with nominal incomes in the latter is warranted by the essence of this study (as the development of small entrepreneurship in a region should influence just on nominal incomes in this region).

Two methods of estimating real incomes can be distinguished in the reviewed papers. The first (which is also applied by the Russian official statistics) is to deflate nominal incomes with the use of regional CPIs. Seemingly, such a method should provide comparability of incomes across both time and space. However, this is not the case. Regional CPIs are not comparable with one another, as its own structure of the CPI commodity basket is used in each region (that is, commodity weights involved in computing regional CPI are region-specific). As Gluschenko (2006b) finds, this method severely distorts inter-regional differences, even so that nominal incomes turn out to be a more exact “estimate” of real incomes than that obtained with the use of regional CPIs. Therefore, results in papers where real incomes are estimated through regional CPIs should be treated very guardedly. Gluschenko (2006b) does not consider the accuracy of estimating the growth rate of so estimated real. The aforesaid, nonetheless, implies that results in papers where such growth rates are dealt with rather than incomes themselves should be also treated warily.

The second method consists in applying the cost of some commodity basket by region as an indicator of the regional costs of living. Bradshaw and Vartapetov (2003) and Yemtsov (2005) use the cost of a staples basket and the subsistence minimum. Prior to 2000, both indicators were equivalent. The cost of the staples basket served as a basis for computing the subsistence minimum, expenditures for manufactured goods and services being added as 40% of the cost of the basket. In 2000, the methodology changed, directly taking account of prices for manufactured goods and services. In spite of smallish representativeness of the above indicators (involving 19 to 33 food

items), they fairly well worked during the economic recession, 1992-1998. But they yield progressively less exact estimates of real incomes in subsequent years, at least, since 2002 (Gluschenko, 2006b). Therefore results due to Bradshaw and Vartapetov (2003) and Yemtsov (2005) can be deemed reasonably reliable. Since 2002, the Russian Statistical Agency (Rosstat) publishes the cost of the PC basket (recall, the fixed basket of goods and services for inter-regional comparison of population's purchasing capacity) by region. This indicator is fairly representative, involving 83 goods and services. Melnikov (2007, 2008) applies it to estimate real personal incomes. Litvintseva, Voronkova, and Stukalenko (2007) and Berkowitz and DeJong (2010) use the cost of the PC basket for some point in time, computing its values for other points in time through deflating by regional CPIs. It is not clear how great are biases caused here by such a deflating; however, because of the possibility itself of such biases, the results of these papers need reservations.

Most of papers using GRP deal with nominal values, which is quite explicable by difficulties of estimating real GRP, mentioned in the Introduction. Melnikov (2005, 2007, 2008) and Bradshaw and Vartapetov (2003) approximate real GRP, applying the cost of the PC basket and the subsistence minimum (or the cost of the staples basket), respectively. These indicators capture prices for consumer commodities only, while inter-regional relationships between prices for other goods and services may be quite different. However, the degree of inexactness caused by this is unknown. Lugovoy *et al.* (2007) and Kholodilin, Oshchepkov, and Siliverstovs (2009) use indexes computed by Granberg and Zaitseva (2002, 2003) that make possible to estimate real GRP. But, as these indexes are available for the only year, these authors extend them to other years, applying some deflators. Here, it is also unknown how great the inaccuracy of such a combined method of estimating real GRP is.

One methodological mistake regarding region sample in some papers should be noted (surprisingly, not only foreign, but also Russian authors make it). In the Russian statistical publications, data for a subject of the Russian Federation that includes autonomous *okrugs* (which are, in turn, also the subjects of the Federation) cover such a composite region entirely. Therefore, if a region sample involves autonomous *okrugs*, then double counting occurs; the autonomous *okrugs* are present both explicitly, as such, and implicitly, as parts of the inclusive regions. To avoid double counting in such events, it would be well to recalculate data for a composite region, cleaning them from data for the autonomous *okrugs* (e.g., to compute personal incomes per capita for the south of

the Tymen Oblast, i.e. for this *oblast* without the Khanty-Mansi and Yamalo-Nenets Autonomous Okrugs).

By and large, based on the results of analyzing σ -convergence with the use of different inequality indicators¹⁵, the following pattern of the evolution of inter-regional income inequality in Russia seems to be the most plausible one. During 1992-2005, divergence of real personal incomes among regions occurred. In 1998, the inequality fell dramatically, and then started increasing again. The “effect of 1998” is probably due to a more severe fall of incomes in rich regions because of the financial crisis in Russia, especially in Moscow. Regarding GRP per capita, the pattern is vaguer. Most likely, the GRP inequality also rose during 1992-2005, but more slowly and with deviations from this trend in some years (in particular, in 1998).

The distributions of both personal income and GRP skewed to the left over time, especially in the early years of the market reforms, but thereafter this process slowed down or even stopped. At the same time, there were no convergence clubs. A secondary mode (or modes) in the right-hand tail of the distribution has to be interpreted as outliers rather than convergence clubs, since these are due to two or three regions differing drastically from the rest ones in the level of personal incomes and GRP (Moscow, the Tymen Oblast, and, sometimes, one more region). Carluer’s (2005) results disagree with such a conclusion, suggesting sharp polarization of Russian regions with breaking down into two clubs. However, this owes to infelicitous, very non-uniform discretization of the income space. The width of the second and third income classes are as small as 6.2 and 8.5 rubles (in the 1991 prices), while the widths of the first and fourth classes are 83 and 310 rubles, respectively. Obviously, the probability of falling into so narrow middle classes is fairly small and therefore most of regions fall into outermost income classes, seemingly evidencing polarization.

For the reasons discussed in Sections 2 and 3.1, results of analyzing β -convergence hardly merit consideration (note that this method is also present, although not indicated in Table 2, in a few papers more, namely, in Berkowitz and DeJong, 2010, Lavrovsky and Shiltsin, 2009, and Yemtsov, 2005, having a secondary importance there). However, if one abandons interpretation based on the β -convergence concept, some useful information can be obtained from the results of regression analyses. The unconditional β -convergence regression in fact suggests how the initial level of

¹⁵ Such an analysis is performed in many papers, for which this method is not indicated as a main one in Table 2, namely, in Carluer (2005), Carluer and Sharipova (2004), Fedorov (2002), Herzfeld (2008), Lavrovsky and Shiltsin (2009), Lugovoy *et al.* (2007), Melnikov (2005, 2007, 2008), Solanko (2006, 2008), and Yemtsov (2005).

income per capita influences on the growth rate of income per capita; e.g., whether the growth rate will be higher in regions with lower initial incomes. No clear-cut conclusion can be drawn here, since the discussed papers report discrepant results. Some authors find a negative relationship, some find a negative one, and some authors find no statistically significant relationship at all. It can be surmised that the direction of such a relationship was diverse during different periods of time. For example, this relationship was for sure positive for some time as rich regions enjoyed more rapid growth as well. Because of this, they broke farther and farther away from the main body of regions.

Analysis of conditional β -convergence, treated as the causal cross-sectional analysis, yields richer information, revealing determinants of inter-regional income inequality. Not considering all many and varied explanatory variables that appear in the relevant regressions (including those in papers focused directly on causal analysis), it can be concluded that the most sufficient influence on income inequality is exerted by the share of extractive industries in region's GRP (which has a positive relationship with income per capita and its growth rate), and by the share of agriculture (which has a negative relationship).

One more conclusion that seems to be sufficiently reliable is the presence of interdependence of GRPs (maybe, personal incomes as well) of neighboring regions, which not always can be characterized by certain economic indicators. Such interdependence manifests itself in spatial autocorrelation found by Melnikov (2005, 2007, 2008), Buccellato (2007), Lugovoy *et al.* (2007), and Kholodilin, Oshchepkov, and Siliverstovs (2009). It follows herefrom that, while performing cross-sectional analysis of GRP or personal income, spatial econometrics methods ought to be applied in order to avoid biased estimations.

6. FROM ANALYSIS TO POLICY IMPLICATIONS

Results of analyzing inter-regional income inequality are of practical interest by themselves, providing a general pattern of inequality, its determinants and trends (true enough, the value of the results is determined by their validity which depends on proper use of data and methods of analysis). However, this pattern is a mere “intermediate product”, an initial material for formulation of economic policy. And so, from the standpoint of practice, policy implications based on the analysis performed would be of even greater interest. Out of the 31 discussed papers, only six contain such implications (although sometimes they obviously follow from the results of analysis, as, e.g., a need to develop small entrepreneurship in order to increase incomes in low-income classes in Berkowitz

and Jackson, 2006). Considering publications that are not covered by this survey because of their poor analysis, implications in them generally boil down to a trivial point that it is necessary to mitigate (or even to level off) inter-regional differences in well-being or economic development.

However, the new economic geography leads to the conclusion that inequality in territorial levels of development is fundamentally irremovable in a market economy, being caused by such strong factors as agglomeration effect, resource endowment, or favorable economic-geographical position. Certain of measures for mitigation of inter-regional disparities, suggested in the literature, such as prior industrial development of backward regions or faster increase of investment in them, are, in fact, aimed at counteraction to the above factors. A consequence might be a slowdown of economic growth in the country as a whole.

For soundness of policy implications, one should have an idea of what ought to be achieved. In other words, at least an idea of an acceptable (if not rational) level of inter-regional inequality should be on hand. It is deviation from this level that would make it possible to judge how much insistent is the task of smoothing disparities in population's welfare or regional economic development. However, this problem is not even properly formulated so far. It is obscure, what means the concept itself of "acceptable" or "rational" (from what standpoint?) level of inequality, how to approach the choice of the acceptability (or rationality) criteria, and how to evaluate the evolution of inter-regional inequality from this point of view.

Certainly, it might be relied upon the experience of the EU, where a per capita GRP below 75% of the EU average is accepted as the ill-being line of a region. However, this figure is an a priori one there, too; besides, Russia is far from being the EU. If big countries of a conformable type of economic growth ("catching-up development") are considered, such as Brazil, China, and India, both the degree of income inequality and trends of its change are similar to those in Russia (Zubarevich, 2009). Lacking a benchmark, advices regarding reduction of income inequality among Russian regions rest only upon speculative ideas of what degree of inequality is excessive.

It should be pointed out that even when policy implications are not mere appeals for leveling, measures put forward by researchers for mitigation of inter-regional income disparities in Russia are often fairly superficial. It is quite explainable. To find such measures, much more subtle and deeper analysis of reasons for inequality and mechanisms of its formation and a model evaluation of consequences of some or other measures are needed. Let us consider this issue by a few examples.

One of factors that interfere with decrease of inter-regional income inequality in Russia is low

labor mobility. Its main cause is the housing problem. At first glance massive development of rental housing (instead of the current housing policy that focuses on purchases of dwellings by families for owner-occupation, which further binds them to their regions of residence) would facilitate the reduction of the inter-regional income. It stands to reason that this is necessary in any case, since there is no other way to solve the housing problem in Russia. However, consequences for inter-regional income inequality might turn out to be opposite to expectations which base on that migration to high-income regions would increase labor supply there, causing a decrease in wages, and outflow of labor from low-incomes would cause rise in price for labor because of lower supply. In the context of the new economic geography, a different logic is plausible. Inflow of labor would stimulate attraction of capital, resulting in further concentration of economic activity in high-income regions and further rise in incomes there, while a consequence of decreasing labor supply (and, all the more, that of accelerated diminution of human capital caused by more rapid flight of the most skilled manpower) in low-income regions would be a progressive curtailment of economic activity. Consequences of development of transport infrastructure (proposed as one of measures of regional policy, e.g., by Lugovoy *et al.*, 2007) could prove to be similar.

Measures for mitigation of inter-regional income disparities are far from being always a subject of regional policy. It is known that there is a negative relationship between per capita income in a region and the share of agriculture in GRP (see, e.g., Mikheeva 1999, 2000). This looks quite natural, as agriculture uses less skilled labor than (on average) other sectors. But, comparing the gap in incomes between urban and rural population in Russia and, say, Belarus, it is seen that this gap is much wider in Russia. The reason is activity of multi-regional (somewhere, regional) monopsonies that reallocate to their own advantage incomes of primary producers of agricultural products. And this problem is a subject of agricultural policy, not regional one.

The overall income inequality in a country is determined to a considerable degree by intra-regional inequality rather than inter-regional one. According to Yemtsov (2005), the latter accounts for a quarter to a third of the overall inequality in Russia. Solving the vital task of reducing poverty in the country will bring to decrease in intra-regional income inequality. (It is interesting to note in this respect that at present China moves from the principle “Let the rich will be more in number” to the principle “Let the poor will be less in number”.) As Litvintseva, Voronkova, and Stukalenko (2007) find, a moderate redistribution of incomes is sufficient to overcome poverty in Russia (namely, increasing income tax from the current rate of 13% to 18% for less than 30 millions of

better-offs). Reduction, and all the more overcoming, of poverty in this or other way will at the same time result in reduction of income gaps between regions. But this is, again, a subject of social policy rather than regional one.

REFERENCES

- ACCRA (2002). *ACCRA Cost of Living Index*, 35 (4).
- Ahrend, R. (2005). Speed of reform, initial conditions or political orientation? Explaining Russian regions' economic performance. *Post-Communist Economies*, 17(3), pp. 289-317.
- Anselin, L. (2001). Spatial econometrics. In B. Baltagi (ed.), *A Companion to Theoretical Econometrics*. Oxford: Blackwell, pp. 310-330.
- Anselin, L. (2002). Under the hood. Issues in the specification and interpretation of spatial regression models. *Agricultural Economics*, 27 (3), pp. 247-267.
- Arbia, G. (2006). *Spatial Econometrics: Statistical Foundations and Applications to Regional Convergence*. Berlin, Heidelberg, New York: Springer.
- Azariadis, C., and A. Drazen (1990). Threshold externalities in economic development. *Quarterly Journal of Economics*, 105 (2), pp. 501-526.
- Babetski, J., and M. Maurel (2002). Regional convergence and institutional development of Russia. In Samson, I., X. Greffe and E. Brunat (eds.), *Russia's Opening and the Common European Economic Space*, Moscow: RECEP.
- Balatskii, E.V., and K.M. Saakyants (2006). Income's divergence and economic growth. In Korovkin, A.G. (ed.), *Scientific Articles – Institute of Economic Forecasting, Russian Academy of Sciences*, Moscow: MAKS Press, pp. 583-601. [In Russian.]
- Barro, R.J., and X. Sala-i-Martin (1992). Convergence. *Journal of Political Economy*, 100 (2), pp. 223-251.
- Barro, R.J., and X. Sala-i-Martin (2004). *Economic Growth*. 2nd ed. Cambridge, MA, London: MIT Press.
- Berkowitz, D., and D. N. DeJong (2002). Accounting for growth in post-Soviet Russia. *Regional Science and Urban Economics*, 32 (2), pp. 221–239.
- Berkowitz, D., and D. N. DeJong (2003). Policy reform and growth in post-Soviet Russia. *European Economic Review*, 47 (2), pp. 337–352.
- Berkowitz, D., and D. N. DeJong (2005). Entrepreneurship and post-socialist growth. *Oxford Bulletin of Economics and Statistics*, 67 (1), pp. 25–46.
- Berkowitz, D., and D. N. DeJong (2010). Growth in post-Soviet Russia: a tale of two transitions. *Journal of Economic Behavior and Organization*, 73 (forthcoming).
- Berkowitz, D., and J. E. Jackson (2006). Entrepreneurship and the evolution of income distributions in Poland and Russia. *Journal of Comparative Economics*, 34 (2), pp. 338–356.
- Bernard, A.B., and S.N. Durlauf (1995). Convergence in international output. *Journal of Applied Econometrics*, 10 (2), pp. 97-108.
- Bernard, A.B., and S.N. Durlauf (1996). Interpreting tests of the convergence hypothesis. *Journal of Econometrics*, 71 (1), pp. 161-173.

- Bogomolova, T.Yu., V. S. Tapilina, and P.S. Rostovtsev (2001). *The Impact of Income Mobility on the Change of Income Inequality*. Novosibirsk: Novosibirsk State University. [In Russian.]
- Borts, G.H., and J.L. Stein (1964). *Economic Growth in a Free Market*. New York: Columbia University Press.
- Bradshaw, M. J., and K. Vartapetov K. (2003). A new perspective on regional inequalities in Russia. *Eurasian Geography and Economics*, 44 (6), pp. 403-429.
- Buccellato, T. (2007). Convergence across Russian regions: a spatial econometrics approach. *CSESCE Working Paper* No. 72.
- Carlino, G.A., and L. Mills (1996). Testing neoclassical convergence in regional incomes and earnings. *Regional Science and Urban Economics*, 26 (6), pp. 565-590.
- Carluer, F. (2005). The dynamics of Russian regional clubs: the time of divergence. *Regional Studies*, 39 (6), pp. 713–726.
- Carluer, F., and E. Sharipova (2004). The unbalanced dynamics of Russian regions: towards a real divergence process. *East-West Journal of Economics and Business*, 2004, 7 (1), pp. 11-37.
- Diewert, W.E. (1999). Axiomatic and economic approaches to international comparisons. In Heston, A., and R. E. Lipsey (eds.), *International and Interarea Comparisons of Income, Output, and Prices*. Chicago: University of Chicago Press, pp. 13-87.
- Dolinskaya, I. (2002). Transition and regional inequality in Russia: Reorganization or procrastination? *IMF Working Paper* No. WP/02/169.
- Durlauf, S. N., and D. Quah (1999). The new empirics of economic growth. In: Taylor, J. and M. Woodford (eds.), *Handbook of Macroeconomics*, Oxford: Elsevier, Vol. 1A, pp. 235–308.
- Esteban, J., and D. Ray (1994). On the measurement of polarization. *Econometrica*, 62 (4), 819-851.
- Estban, J., and D. Ray (2007). A comparison of polarization measures. *UFAE and IAE Working Paper* No. 700.07.
- Evans, P., and G. Karras (1996). Do economies converge? Evidence from a panel of U.S. states. *Review of Economics and Statistics*, 78 (3), 384-388.
- Fedorov, L. (2002). Regional inequality and regional polarization in Russia, 1990-99. *World Development*, 30 (3), pp. 443-456.
- Formby, J.P., W.J. Smith, and B. Zheng (2004). Mobility measurement, transition matrices and statistical inference. *Journal of Econometrics*, 120 (1), pp. 181-205.
- Friedman, M. (1992). Do old fallacies ever die? *Journal of Economic Literature*, 30 (4), pp. 2129-2132.
- Galbraith, J. K., L. Krytynskaia, and Q. Wang (2004). The experience of rising inequality in Russia and China during the transition. *European Journal of Comparative Economics*, 1 (1), pp. 87-106.
- Galor, O. (1996). Convergence? Inferences from theoretical models. *Economic Journal*, 106 (437), pp. 1056-1069.
- Galton, F. (1886). Regression towards mediocrity in hereditary stature. *Journal of the Anthropological Institute of Great Britain and Ireland*, 15, pp. 246-263.

- Gluschenko, K. (2006a). Russia's common market takes shape: price convergence and market integration among Russian regions. *BOFIT Discussion Paper* No. 7.
- Gluschenko, K. (2006b). Biases in cross-space comparisons through cross-time price indexes: the case of Russia. *BOFIT Discussion Paper* No. 9.
- Goskomstat of Russia (1998). *Methodological Regulations on Statistics. Issue 2*. Moscow. [In Russian.]
- Granberg, A.G., and Yu. S. Zaitseva (2002). Production and use of gross regional product: inter-regional comparisons. Article 2. Adjustments of GRP according to spatial differences in purchasing power of money. *Rossiyskiy Ekonomicheskiy Zhurnal [Russian Economic Journal]*, No. 11-12, pp. 48-70. [In Russian.]
- Granberg, A.G., and Yu. S. Zaitseva (2003). Inter-regional comparisons of gross regional product in the Russian Federation. *Voprosy Statistiki [Problems of Statistics]*, No. 2, pp. 3-17. [In Russian.]
- Greene, W.H. (2003). *Econometric Analysis*. 5th ed. Upper Saddle River, NJ: Pearson Education.
- Herzfeld, T. (2008). Interregional income distribution: a comparison of Russian and Chinese experience. *Post-Communist Economies*, 20 (4), pp. 431-447.
- Hill, F., and C. Gaddy (2003). *The Siberian Curse: How Communist Planners Left Russia in the Cold*. Washington, DC: Brooking Institution Press.
- Hobijn, B., and P.H. Franses (2000). Asymptotically perfect and relative convergence of productivity. *Journal of Applied Econometrics*, 15 (1), pp. 59-81.
- Kholodilin, K.A., A. Oshchepkov, and B. Siliverstovs (2009). The Russian regional convergence process: where does it go? *DIW Berlin Discussion Paper* No. 861.
- Kwon, G., and A. Spilimbergo (2005). Russia's regions: income volatility, labor mobility, and fiscal policy. *IMF Working Paper* No. WP/05/185.
- Lavrovsky, B.L., and Ye. A. Shiltzin (2009). Russian regions: leveling or stratification? *Ekonomika i Matematicheskie Metody [Economics and Mathematical Methods]*, 45 (2), pp. 31-36. [In Russian.]
- Ledyayeva, S., and M. Linden (2008). Determinants of economic growth: empirical evidence from Russian regions. *European Journal of Comparative Economics*, 5 (1), pp. 87-105.
- LeSage, J., and R.K. Pace (2009). *Introduction to Spatial Econometrics*. Boca Raton, FL: CRC Press.
- Litvintseva, G.P., O.V. Voronkova, and E.A. Stukalenko (2007). Regional income inequality and poverty level in Russia: an analysis adjusted for purchasing power of the ruble. *Studies on Russian Economic Development*, 18 (6), pp. 641-649.
- Lucas, R. E. (1993). Making a miracle. *Econometrica*, 61 (2), pp. 251-272.
- Lugovoy, O., V. Dashkeyev, I. Mazayev, D. Fomchenko, E. Polyakov, and A. Hecht (2007). *Analysis of Economic Growth in Regions: Geographical and Institutional Aspect*. Moscow: IET.
- Magrini, S. (2004). Regional (di)convergence. In Henderson, V., and J.-F. Thisse (eds.), *Handbook of Regional and Urban Economics*, Vol. 4. Amsterdam, New York, and Oxford: Elsevier Science, pp. 2741-2796.

- Melnikov, R.M. (2005). Analysis of the dynamics of inter-regional economic inequality: foreign approaches and the Russian practice. *Region: Ekonomika i Sotsiologiya* [Region: Economics and Sociology], No. 4, pp. 3-18. [In Russian.]
- Melnikov, R.M. (2007). Inter-regional economic inequality in the Russian economy: trends and prospects. *Regionalnaya Ekonomika: Teoria i Praktika* [Regional Economics: Theory and Practice], No. 8, pp. 26-33. [In Russian.]
- Melnikov, R.M. (2008). Inter-regional economic inequality in the Russian economy: trends and prospects. *Regionalnaya Ekonomika: Teoria i Praktika* [Regional Economics: Theory and Practice], No. 3, pp. 7-14. [In Russian.]
- Mikheeva, N. (1999). Analysis of interregional inequality in Russia. *Studies on Russian Economic Development*, 10 (5), pp. 514-521.
- Mikheeva, N. (2000). Differentiation of social and economic situation in the Russian regions and problems of regional policy. *EERC Working Paper* No. 99/09.
- Mitsek, S.A., and E.B. Mitsek (2009). Investments, productivity, and personal incomes in regions of the Russian Federation. *Regionalnaya Ekonomika: Teoria i Praktika* [Regional Economics: Theory and Practice], No. 36, pp. 25-28. [In Russian.]
- Nahar, S., and B. Inder (2002). Testing convergence in economic growth for OECD countries. *Applied Economics*, 34 (16), pp. 2011-2022.
- Phillips, P.C.B., and D. Sul (2007). Transition modeling and econometric convergence tests. *Econometrica*, 75 (6), pp. 1771-1855.
- Quah, D. (1993a). Empirical cross-section dynamics in economic growth. *European Economic Review*, 37 (2-3), pp. 426-434.
- Quah, D. (1993b). Galton's fallacy and tests of the convergence hypothesis. *Scandinavian Journal of Economics*, 95 (4), pp. 427-443.
- Quah, D. (1996). Convergence empirics across economies with (some) capital mobility. *Journal of Economic Growth*, 1 (1), pp. 95-124.
- Quah, D. (1997). Empirics for growth and distribution: stratification, polarization, and convergence clubs. *Journal of Economic Growth*, 2 (1), pp. 27-59.
- Rey, S. J. (2001). Spatial empirics for economic growth and convergence. *Geographical Analysis*, 33 (3), pp. 194-214.
- Romer, P.M. (1986). Increasing returns and long-run growth. *Journal of Political Economy*, 94 (5), pp. 1002-1037.
- Rosstat (2008). *Regions of Russia. Socio-Economic Indicators. 2008*. Moscow. [In Russian.]
- Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*. London: Chapman and Hall.
- Solanko, L. (2006). *Essays on Russia's Economic Transition*. Helsinki: Bank of Finland.
- Solanko, L. (2008). Unequal fortunes: a note on income convergence across Russian regions. *Post-Communist Economies*, 20 (3), pp. 287-301.

- Wodon, Q., and S. Yitzhaki (2006). Convergence forward and backward? *Economics Letters*, 92 (1), pp. 47-51.
- Wolfson, M.C. (1994). When inequalities diverge. *American Economic Review*, 84 (2), 353-358.
- Yemtsov, R. (2005). Quo vadis? Inequality and poverty dynamics across Russian regions. In: Kanbur, R., and A.J. Venables (eds.), *Spatial Inequality and Development*. Oxford: Oxford University Press, pp. 348-397.
- Yitzhaki, S., and Q. Wodon (2004). Mobility, inequality, and horizontal equity. In Amiel, Y., and J. A. Bishop (eds.), *Research on Economic Inequality*, Vol. 12. Oxford: Elsevier, pp. 179-199.
- Zhang, X., and R. Kanbur (2001). What difference do polarization measures make? An application to China. *Journal of Development Studies*, 37 (3), 85-98.
- Zubarevich, N.V. (2009). Myths and realities of spatial inequality. *Obshchestvennye Nauki i Sovremennost [Social Sciences and the Present]*, No. 1, pp. 38-53. [In Russian.]

DAVIDSON INSTITUTE WORKING PAPER SERIES - Most Recent Papers

The entire Working Paper Series may be downloaded free of charge at: www.wdi.umich.edu

CURRENT AS OF 5/10/10

Publication	Authors	Date
<i>No. 984: Methodologies of Analyzing Inter-Regional Income Inequality and Their Applications to Russia</i>	Konstantin Gluschenko	April 2010
<i>No. 983: Foreign News and Spillovers in Emerging European Stock Markets</i>	Evzen Kocenda and Jan Hanousek	May 2010
<i>No. 982: The Economic and Monetary Union's effect on (international) trade: the case of Slovenia before euro adoption</i>	Aleksander Aristovnik and Matevz Meze	April 2010
<i>No. 981: In Time of Troubles: Challenges and Prospects in the Middle East and North Africa</i>	Imed Drine	April 2010
<i>No. 980: A Composite Leading Indicator of Tunisian Inflation</i>	Mohamed Daly Sfia	March 2010
<i>No. 979: Financial stability, monetary autonomy and fiscal interference: Bulgaria in search of its way, 1879-1913</i>	Kalina Dimitrova and Luca Fantacci	Feb 2010
<i>No. 978: The Supply Side of Innovation: H-1B Visa Reforms and US Ethnic Invention</i>	William R. Kerr and William F. Lincoln	Feb 2010
<i>No. 977: Data Collection Procedures Equivalence in International Business Research</i>	Agnieszka Chidlow, Anna Morgan-Thomas and Pervez N. Ghauri	Feb 2010
<i>No. 976: Direct and Indirect Effects of FDI in Emerging European Markets: A Survey and Meta-analysis</i>	Jan Hanousek, Evzen Kocenda and Mathilde Maurel	March 2010
<i>No. 975: MODELING INSTITUTIONS, START-UPS AND PRODUCTIVITY DURING TRANSITION</i>	Zuzana Brixiová and Balázs Égert	Feb 2010
<i>No. 974: Pegging the future West African single currency in regard to internal/external competitiveness: a counterfactual analysis</i>	Gilles Dufrénot and Kimiko Sugimoto	Dec 2009
<i>No. 973: Unemployment and finance: how do financial and labour market factors interact?</i>	Donatella Gatti, Christophe Rault & Anne-Gaël Vaubourg	January 2010
<i>No. 972: Identifying Social Entrepreneurs Serving the Poor at the BoP</i>	Abraham M. George	December 2009
<i>No. 971: Inflation dynamics and the New Keynesian Phillips curve in the EU-4</i>	Bořek Vašíček	October 2009
<i>No. 970: International Financial Integration And Real Exchange Rate Long-Run Dynamics In Emerging Countries: Some Panel Evidence</i>	Guglielmo Maria CAPORALE, Thouraya HADJ AMOR and Christophe RAULT	Sept 2009
<i>No. 969: Once Bitten, Twice Shy: Experiences Of A Banking Crisis And Expectations Of Future Crises</i>	Shannon Mudd & Neven Valev	Sept 2009
<i>No. 968: Monetary policy rules and inflation process in open emerging economies: evidence for 12 new EU members</i>	Bořek Vašíček	Sept 2009
<i>No. 967: Extending Likages Between Organizational Analysis And Social Structure: A Case Study Of The Celebrity-Construction Of A Chinese Marketplace</i>	Mark Jacobs	Aug 2009
<i>No. 966: Chinese state's economic cooperation related investment: An investigation of its direction and some implications for outward investment</i>	Sumon Bhaumik and Catherine Yap Co	Aug 2009
<i>No. 965: Capital inflows, household debt and the boom-bust cycle in Estonia</i>	Zuzana Brixiova, Laura Vartia and Andreas Worgötter	July 2009
<i>No. 964: Labour Market Flexibility in Estonia: What More Can be Done?</i>	Zuzana Brixiova	July 2009
<i>No. 963: Market Globalization by Firms from Emerging Markets & Small Countries: An Application of the Neoclassical Trade Model</i>	Tamir Agmon	July 2009