Alma Mater Studiorum - Università di Bologna
DEPARTMENT OF ECONOMICS

Is There Convergence of Russia's Regions? Exploring the Empirical Evidence: 1995 – 2010

Hartmut Lehmann
Maria Giulia Silvagni

Is There Convergence of Russia's Regions?
Exploring the Empirical Evidence: 1995 – 2010*

Hartmut Lehmann (Department of Economics, University of Bologna and IZA, Bonn)

Maria Giulia Silvagni (Department of Economics, University of Bologna)

Abstract

This paper analyzes convergence in per capita gross regional product of Russia's regions during the period 1995-2010, when regional data are available. Using a panel regression framework we find no evidence for beta-convergence. Instead we find divergence, which is, however, attenuated over time. Robustness checks that use regional real income instead of gross regional product confirm this outcome as do non-parametric estimates of convergence, namely estimates using Markov transition probability matrices and stochastic kernel plots of regional relative income. Decompositions of regional income and gross regional product also find no sigma-convergence of Russian regions. These decompositions point to the geographical concentration of extractive activities in the Urals and of business services and of the public administration in the Moscow area as the main culprit for this lack of convergence. They also establish that despite reforms to equalize provisions of public goods across Russia, the social services sector of the public administration, education and health still do not have the expected equalizing impact on regional income.

Keywords: Russian regions, convergence, regional inequality decomposition, regional distribution dynamics

June 2013

JEL Codes: O47, P25, R11, R12

Is There Convergence of Russia's Regions?
Exploring the Empirical Evidence: 1995 - 2010

1   Introduction

The Russian economy experienced dramatic changes soon after transition from the centrally planned economy started in the early nineties, when the country faced a precipitous decline in gross domestic product per head (GDP per head) and the appearance of open unemployment. OECD data show a slowing down of the fall of GDP per head from the mid-nineties and acceleration of the rise in unemployment since 1993: between 1995 and 1998, GDP per capita declined from 7,840 US $ to slightly less than 7,300 US $, whereas the unemployment rate increased from 5.9 percent in 1993 to 13.2 in 1998. This very prolonged transition recession ended in 1999 and since then, up to the economic crisis of 2008, Russian GDP per capita rose by 5-6 per cent per annum, while the unemployment rate fell back to around 6 percent in 2008.

Figure 1 Russian Gross Domestic Product per capita at constant 2005 prices, US $, 1995-2009
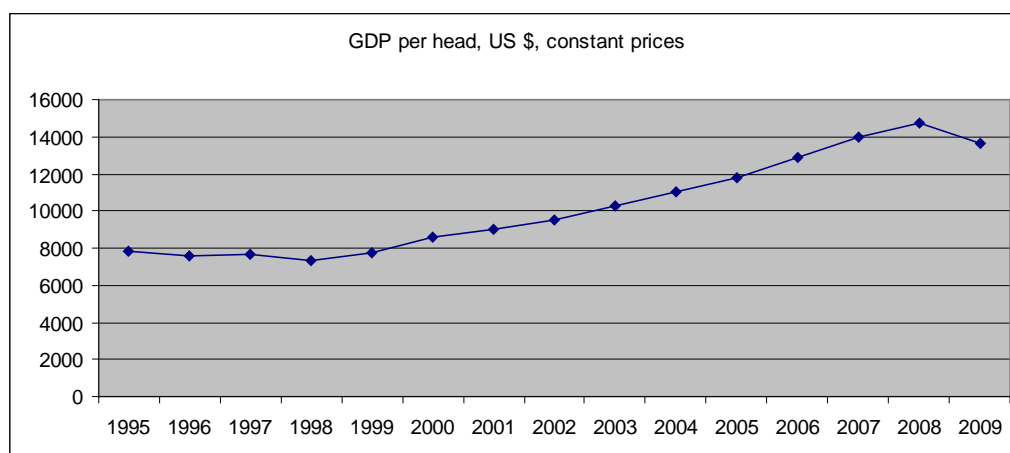

Source: OECD Statistics

Figure 2 Unemployment rate, 1993-2011
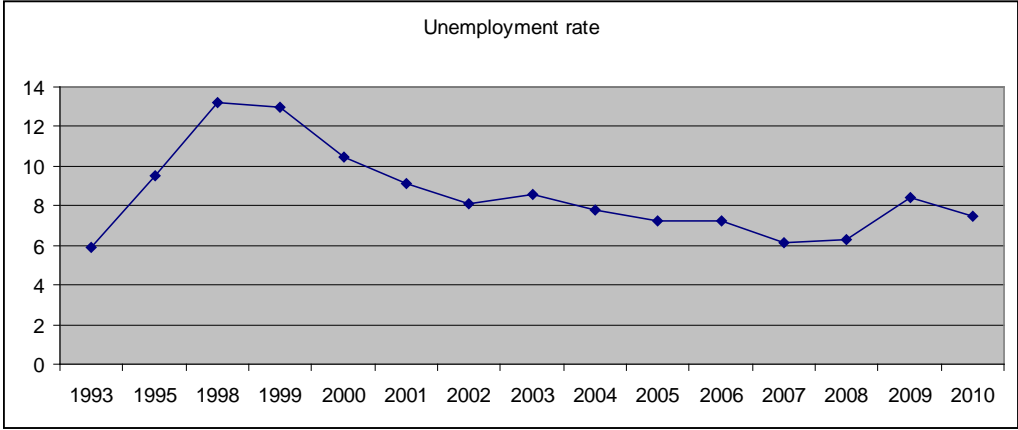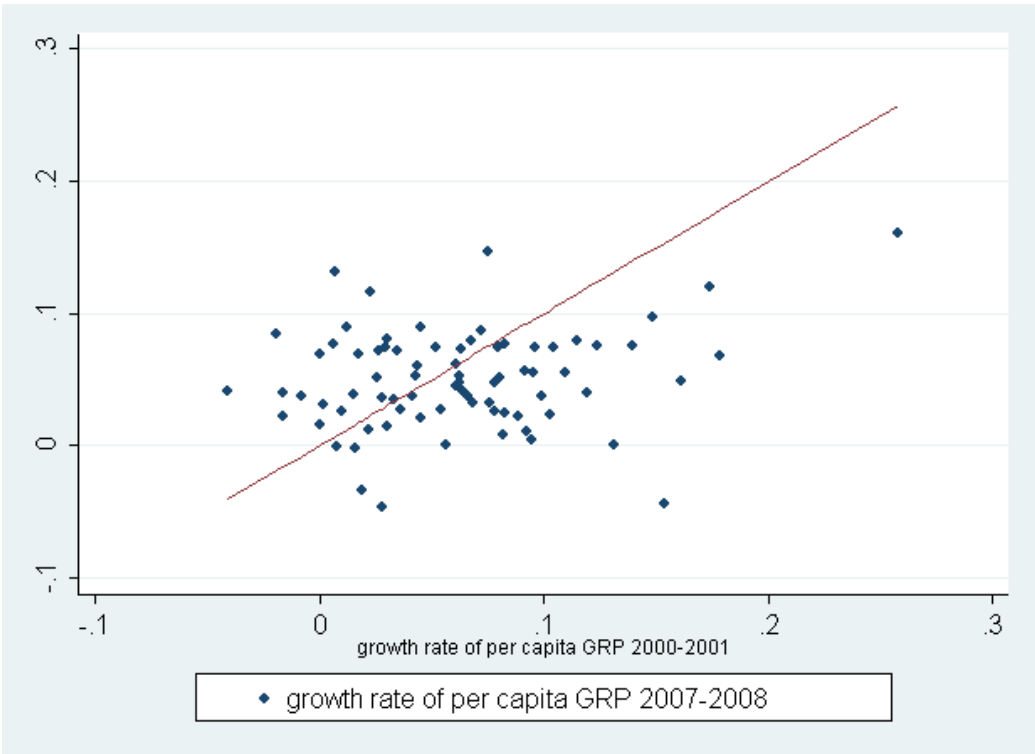

Source: Rosstat data

This overall good performance regarding GDP and unemployment is, however, accompanied by a large and rising inequality in income and a persistent variation in regional performance. Denisova (2012) estimates an increase in the Gini coefficient of incomes from 0.39 to 0.42 during 2001-2009, and a decline of the income share of the poorest quintile of the population from 10 percent in 1990 to 5 percent in 2008. In the richest areas of the country like, e.g., Moscow, this income share has declined even more to between 3 and 4 percent (Akhmedjonov et al 2013). The persistent variation in regional performance can be seen in the graphs 3a-b where we plot the growth rates of gross regional product (GRP) in 1995-1996 against those in 2007-2008 and 2000-2001 against those in 2007-2008. Regions located above the 45-degree diagonal had higher growth rate in the second period with respect to the first one, and vice versa for those regions below the 45-degree line. Figure 3a plots the growth rate in 1995-1996 against 2007-2008: unsurprisingly most regions are plotted above the diagonal line, i.e. they had higher growth rates in recent years than at the beginning of transition. Notice that in the first phase of transition, most regions had a negative growth rate of per capita GRP and growth rates appears rather dispersed. Less variability in growth rates is found in figure 3b, in the second phase of transition. Starting from the year 2000, most of the regions had a sustained growth path over the next decade with increasing concentration of regions around the 45-degree line.

Figure 3a Growth of GRP per head across Russian regions, 1995-2008
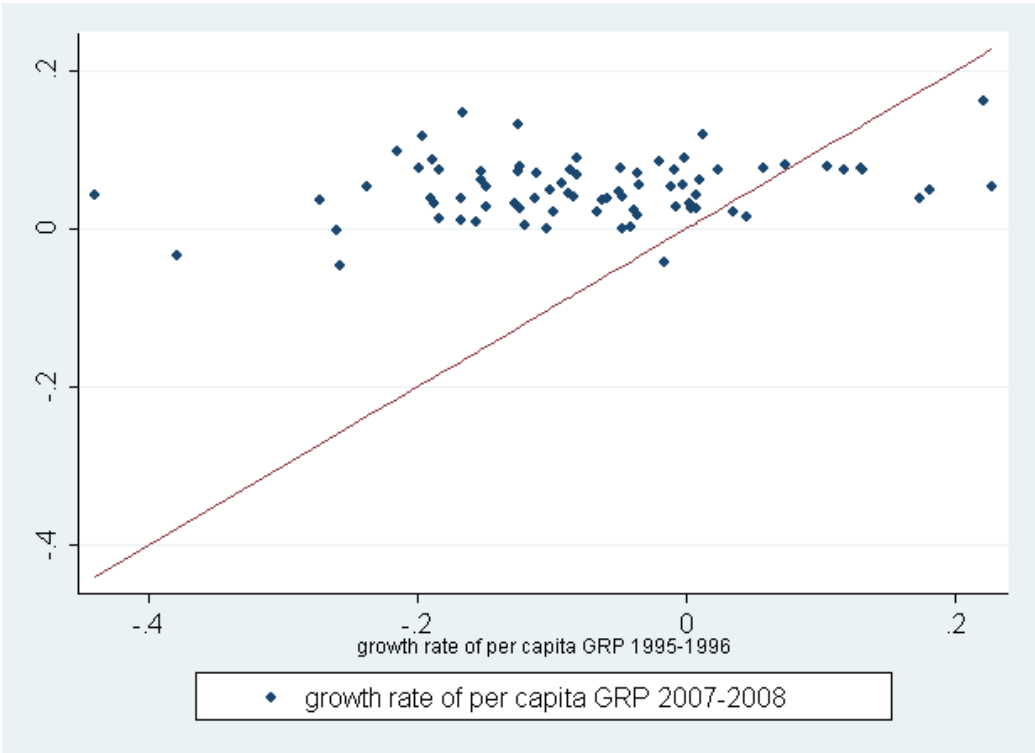
Source: Authors' calculations using Rosstat data.

Figure 3b Persistence of GRP per head growth rates across Russian regions, 2001-2008

Source: Authors' calculations using Rosstat data.

The structural changes over the recessionary and expansionary phases of the transition cycle (i.e. before and after the financial crisis of 1998) not only affected per capita GDP and the distribution of income, but also the productive structure of the country as the shares of gross value added were 5.7 and 4.2 for agriculture, and 30.3 and 17.7 for manufacturing in 1998 and 2010 respectively. The change in this structure was above all driven by two core areas of the Russian Federation: the Moscow and the Urals regions[1]. These two regions above all have been attracting large investments and have been contributing very substantially to the growth path of the Russian economy over the last 15 years. On the one hand, most businesses and financial activities are concentrated in Moscow, where ministries and agencies are also located. Fifty percent of gross value added of the city is accounted for by wholesale and retail trade, and real estate and business activities. On the other hand, the Urals area is the centre of extractive activities. The Tyumen region and its two autonomous provinces owe more than 50 percent of gross value added to mining and quarrying, whereas in the Sverdlovsk region the share of production of electricity, gas, and water in gross value added was slightly less than 40 percent in 2010. Though the core of extractive activities are located in the Urals, these activities are performed also in other areas, e.g. the positive economic results of the Far Eastern district in recent years are driven by the performance of Chukotka and Sakhalin. Two main drawbacks are related to extractive activities: first, they are geographically concentrated and resource-dependent, and benefits do not diffuse in surrounding regions unless these areas are involved in the same type of activity. Second, rentier activities[2] exacerbate disparities with non-rentier regions.

The aim of this paper is to analyze the patterns of regional convergence (or the lack thereof) and regional inequality in Gross Regional Product (GRP) in Russian regions from 1995 to 2010, taking the described specific agglomeration patterns into consideration. A deeper understanding of both the transition path and the current economic situation of the Russian Federation from a regional perspective is interesting on several grounds. First, despite the overall economic growth experienced at the national level, severe inter-regional disparities in living standards are persistent: shedding light on the regional recovery patterns is crucial to identify the reasons why certain regions still lack behind and design appropriate policy interventions. Second, the great role played by the increase in oil exports and oil prices in the process of recovery from 1999 raises concerns on a Russian Dutch-disease and its

---

[1] St. Petersburg is also a high growth region, but not an outlier like Moscow or the Urals area
[2] Desai et al (2005), Freinkman and Plekhanov (2009) define *rentier regions* as those which derive a major part of revenue from natural resource production or budgetary transfers from the central government

implications for investments in other sectors such as manufacturing. Due to the geographical concentration of extractive activities previously highlighted, it is worth comparing the economic performance of rentier regions vis-à-vis non-rentier regions.

The paper is organized as follows. In section 2 we review the literature on sigma and beta convergence, drawing reference to the literature on Russian regions in particular. Section 3 presents data and methods. We use aggregate data published annually by Rosstat. As far as methods are concerned, estimates of unconditional and conditional convergence are performed using the regression approach in line with the seminal work of Baumol (1986). We also employ the distribution dynamics approach, first proposed by Quah (1993). Technical aspects of our empirical methodologies are described in a technical appendix at the end of the paper. We then proceed to analyze the dynamics of regional inequality using generalized entropy measures and their decompositions. Following Shorrocks (1982, 1984), inequality is decomposed by groups and industrial sectors to understand the drivers of non-convergence. Our results are presented in sections 4-7, employing the various methods mentioned above. A final section concludes.

2   Related literature

Convergence refers to two concepts: sigma and beta convergence. Sigma convergence measures income dispersion. There is evidence of sigma convergence if dispersion, measured by the standard deviation, coefficient of variation or other indices, is declining over time. On the other hand, there is evidence of beta convergence if the relationship between the growth rate of income and the initial level of income is negative, in which case, e.g., poorer countries tend to grow faster than richer ones and thus catch-up eventually occurs.

Beta convergence can be unconditional , with all countries reaching the same steady state, or conditional, i.e. countries with a similar set of institutions and other variables determining growth can only reach the same steady state. Two main strands of research can be identified as far as beta convergence is concerned: the regression approach and the non-parametric distribution dynamics approach. The literature on the regression approach to beta convergence started with Baumol (1986) based on the seminal work of Solow (1956) and has developed into a vast literature applying cross-sectional, time-series, static and dynamic panel data techniques among others (see for example, Azariadis and Drazen 1990; Barro 1991; Barro and Sala-i-Martin 1991; Mankiw et al 1992; Armstrong 1995; Caselli et al 1996; Lee et al 1997; Eberhardt et al 2011; Barro 2012; Dobson et al 2012). The non-parametric approach

dates back to Quah (1993), Durlauf and Johnson (1995), Quah (1996), Bianchi (1997), Quah (1997) among others. This approach is critical towards the view implicit in the Barro-type growth regressions that each country has a steady-state growth path, which is well approximated by a time trend (Barro and Sala-i-Martin 1992). Instead, it investigates the possibility of multimodality and club convergence through kernel density estimators and Markov transition probability matrices.

The literature on Russian regional performance after transition is substantial[3], where the issues of convergence and of the growth determining factors have been analyzed from various perspectives. Different models and specifications have been tested to analyze the pattern of convergence (or divergence) in income and gross regional product across Russian regions, both with parametric and non-parametric methodologies. Results somehow vary, depending on the period of analysis and the methodology, but no consistent evidence of convergence has been found so far since the beginning of transition. We review the empirical studies on regional convergence in Russia and compare them with the most recent results for other countries.

The factors which can explain regional disparities in Russia refer to two main domains: initial conditions, in particular the availability of extractive activities, and the effectiveness of the reforms implemented during transition. Among the first studies on the economic performance of Russian regions, Mikheeva (1999) and Popov (1999) highlight the role of favourable initial conditions in the form of shares of exports and availability of natural resources, in explaining regional growth differentials. For the early years of transition, 1991-1997, Dolinskaya (2002) applies transition matrix methodology with personal incomes data and finds a tendency for the majority of Russian regions to be trapped at low incomes. Those few successful regions instead have had the availability of natural resources and a favourable industrial structure at the beginning of transition. For the time period 1992-2001, Solanko (2003) uses cross-sectional regressions controlling for education, agriculture, and the mining sector: the author finds both evidence of increasing income dispersion, i.e. sigma divergence, and beta convergence, signalling the presence of club convergence between homogenous groups of regions. In an updated analysis, Solanko (2008) finds convergence in the club of rich regions. Evidence in favour of club convergence splitting the sample in high and low-income regions is found also in static and dynamic panel specifications by Ledyaeva and

---

[3] See Gluschenko (2010) for a complete review

Linden (2008) in the period 1996-2004. Ledyaeva (2009) performs a spatial analysis on the factors attracting foreign direct investment (FDI) at the regional level: Her results show that the presence of big cities and ports, hydrocarbons extraction, and market size are drivers of inward FDI. Ahrend (2008) performs an extreme-bound analysis in a panel of 77 regions before and after the financial crisis of 1998, i.e. he analyses the intervals 1993 to 1998 and 1998 to 2004. The author does not find a robust relationship between initial levels of gross regional product and its growth rate. He identifies which factors affect economic growth during periods of crisis and growth: in the former scenario, initial conditions, in particular resource endowments and industrial structure, explain the differences in regional growth, whereas only natural resources remain significant in periods of growth, when other factors such as political reforms contribute to the path. In previous papers (Ahrend 2005, 2006; Ahrend and Tompson 2005), no evidence of convergence can be established controlling for initial industrial production, though convergence is found in real income. Guriev and Vakulenko (2012) perform inequality decompositions by income sources and study convergence in Gross Regional Product, incomes and wages. They find convergence in incomes and wages, but not in Gross Regional Product. Berkowitz and DeJong (2003) perform OLS and 2SLS regressions using data for the period 1993-1997 for 48 regions. They find that different regional implementation of price liberalization and privatizations were beneficial for growth and the formation of new enterprises, whereas no evidence of convergence or divergence in real income is found. In another study, Berkowitz and DeJong (2005) with a sample of 70 regions point to entrepreneurial activity as a driver of regional growth given favourable initial conditions and policy reforms.

Another strand of the literature is related to the New Economic Geography (Buccellato 2007, Lugovoy et al. 2007, Kholodilin et al. 2009). It focuses on the spatial structure of the country through various cross-sectional and panel data methodologies in the period around 1998-2006. The cited authors find evidence, albeit limited, of conditional convergence controlling for the spatial correlation between neighbouring regions and variables such as investments and extractive activities. Akhmedjonov et al (2013) investigate convergence in the time period 2000-2008 using a nonlinear panel unit root test and non- linear cross-sectional Augmented Dickey-Fuller (ADF) regression models. The authors do not find evidence of unconditional beta convergence, though conditional convergence is found within similar neighbouring regions. Finally, analysis of inequality through generalized entropy indices (Galbraith et al 2004, Mahler 2011) provides evidence of divergence up to the recent global

crisis of 2008 with the highest increase in inequality and regional differentials during the early years of transition.

Turning to similar literature studying China, the U.S. and the European Union, Lau (2010a,b) performs both linear and non linear panel unit root tests as Akhmedjonov et al (2013) with Chinese provincial data 1952-2003 and US data 1929-2005. The evidence with the Chinese data is similar to the Russian case, i.e. Lau finds increasing inequality since 1978, i.e. since the beginning of the Chinese reforms, whereas for the US the scenario is different, showing stationary time series and evidence of both sigma and beta convergence. Prochniak et al (2013) find strong beta convergence in GDP for various EU subsamples during 1970s-2010 using Bayesian Moving Averaging and a GMM estimator. Rapacki et al (2009) assess sigma and beta convergence in 27 countries in Central-Eastern Europe and the CIS during 1990-2005 with ambiguous results for different groups. The authors conclude that, due to the high heterogeneity of the sample, for the whole group the convergence hypothesis cannot be confirmed.

3   Data and methods

3.1   Regional Rossstat Data

We use aggregate data from the volumes *Regions of Russia* published annually by the Russian Federal State Statistics Services Rosstat. Data for nominal per capita GRP and volume indices at the regional level are available on a yearly basis since 1995. GRP is expressed in Russian roubles in constant 1995 prices from 1995 to 2010. Data for gross value added of industrial sectors as a percentage of the total are available from 1998 to 2010 in current prices: we use these data for the decomposition analysis of inequality.

The Russian Federation is divided in 83 regions, called Subject of the Federation, which can be sub-divided into the following types: as of 2012, there are 21 Republics, 9 Krais, 46 Oblasts, 2 Federal Cities (Moscow and Sankt Petersburg), 1 Autonomous Oblast and 4 Autonomous Okrugs. For the analysis of inequality, we use as units the regions[4] and further collect them according to the system of the Federal Districts, established in 2000 with the aim of ensuring compliance by the regional laws to the federal ones. The seven Federal Districts are Central, North-Western, Southern[5], Volga, Urals, Siberian and Far East District[6]. This

---

[4] We excluded Kaliningrad for territorial reason, Ingushetia and Chechnya due to data availability.
[5] The Southern Federal District has been divided in Southern District and District of the North Caucasus in 2010. However, the time span of our analysis stops at 2010 so we use the previous division in seven districts.
[6] See Appendix for the list of the regions and the federal districts

level of aggregation allows us to combine the regions in a few macro-areas and to analyze the patterns of inequality between regions.

The analysis of the regional Rossstat data aims at highlighting the relationship between inequality in regional GRP levels and regional industrial specialization, and it is motivated by several peculiarities which emerged over time as far as the geographical location of both population and economic activities are concerned. The share of the financial sector in GRP is rather limited and businesses of the financial sector are concentrated in Moscow. Extractive activities are mostly concentrated in the Urals District, which accounts for 90 percent of the total national production of natural gas, 70 percent of the oil production and 10 percent of the production of electricity, whereas agriculture is more diffused in the Southern and Volga Districts due to favourable weather conditions. The very uneven distribution of the population across the Russian Federation has been widely discussed[7]. In 2010, 70 percent of total population, that is 143 million people, lived in the European part of the country, which covers a mere 25 percent of the total national territory, and almost ten percent of the total population lived in the Moscow area. For these reasons, in the inequality metrics analyses we use population-weighted indices and decompose GRP by industrial sectors to highlight to role of extractive activities in inequality between regions.

3.2   Regression approach to convergence

Our dataset is composed of 77 Russian regions over the period 1995-2010. To test for the presence of convergence in real per capita GRP, we apply several regression models and consider different sub-periods. In this section we briefly describe each regression model, while the methodology of dynamic models is described in detail in the technical appendix and results will be presented and discussed in section 4[8].

In his seminal work, which serves as a general reference to parametric methodologies to model convergence, Baumol (1986) estimates convergence across a sample of countries from the Maddison dataset over 1870-1979 within the regression framework described by equation [1]:

$$ln y_{i,t} - ln\ y_{i,0} = a + \beta \ln y_{i,0} + \varepsilon_{i,t} \qquad [1]$$

where the dependent variable is the average growth rate of GDP over the period ($0, t$) and ln y_0 is the GDP at the beginning of the period, while $\varepsilon_{i,t}$ is a white noise error term. When the

---

[7] See Kontorovich (2010) for the case of the Far Eastern district
[8] Caselli et al (1996), Islam (2003), Barro and Sala-i-Martin (2004), Magrini (2004), Barro (2012) provide reviews of the convergence debate and provide critical discussions of the various models used.

estimate of β is negative and significant, this is interpreted as evidence of unconditional convergence.[9] We perform cross-section regressions in line with equation [1] for the whole period 1995-2010 across N=77 Russia's regions. To take account of the different growth performance in the 1990s and in the first decade of the new century we also split the data into the intervals 1995-2000 and 2001-2010.

With the availability of richer long-run databases, a further step of the literature has been the adoption of panel data models with country-specific and time-specific fixed effects[10] to combine the cross-sectional and the dynamic dimension and solve partially for the issues of omitted variable bias and unobserved heterogeneity. The use of panel data models for the study of convergence dates back to Islam (1995), who refers to the neoclassical framework of Mankiw et al (1992), Caselli, Esquivel and Lefort (1996). The typical estimated equation to assess unconditional or conditional convergence takes the form:

$$ln\ y_{i,t} - lny_{i,t-1} = \beta ln\ y_{i,t-1} + \delta W_{i,t-1} + \mu_i + \eta_t + \varepsilon_{i,t} \qquad [2]$$

where the dependent variable and $ln\ y_{i,0}$ have the same interpretation as in equation [1] and μ_i are country-specific fixed effects to control for unobserved time-invariant heterogeneity, $W_{i,t-1}$ is a row vector of economic growth determinants, which is included when testing the hypothesis of conditional convergence, and η_t are time dummies. The panel data estimator that we adopt for our analysis is the Least Square Dummy Variable (LSDV) estimator, which adds to the growth equation a full set of region-specific dummies in the form of fixed effects μ_i.

The literature on empirical economic growth adopts two methods when dealing with panel data.[11] The first method takes averages of the dependent variable over non-overlapping time intervals of the whole period and controls for the initial level of GDP for each interval, thus reducing short-term effects in the estimates of long-run convergence; the second method is the one immediately represented by equation [2] with periods *(t)* and *(t-1)*: this method uses

---

[9] Starting with Barro (1991) and Barro and Sala-i-Martin (1991, 1992, 1995), various specifications have been applied to estimate equation [1], e.g. year *(t)* and *(0)* can be two consecutive years , i.e. *(t)* and *(t-*1), or several years apart, i.e. *(t)* and *(t-τ)* with τ>1.

[10] The inclusion of time dummies removes common trends for it is a transformation of the data in term of deviations from time-specific averages. See Arellano and Bond (1991), Blundell, Bond and Windmeijer (2000).

[11] See Barro and Sala-i-Martin (2004) for a complete review, Borys, Polgar and Zlate (2008) for an application to EU candidate countries.

annual data with the annual growth rate of GDP as the dependent variable, and controls for the level of GDP prior to the one in which the growth rate is measured. The first method incorporates the advantage of reducing business cycle effects, however the averaging procedure over time intervals also reduces the number of observations in the dataset.

We estimate equation [2] over the whole period both averaging the growth rate of GRP over $\tau=4$ year intervals[12] and annually. As a sensitivity analysis, for the first method we also estimate the model using as a control variable the GRP of the year prior to the beginning of each interval[13]. Equation [2] is estimated also averaged over the two sub-periods 1995-2000, 2001-2010. We also estimate annually the two sub-periods 1995-2000, 2001-2010.

To address the endogeneity issues related to the static regression approach we use the Anderson-Hsiao (1981) estimator and a System-GMM estimator[14]. Anderson and Hsiao (1981) suggest to estimate the growth equation in first differences using lags as instruments for the endogenous covariates. Rearranging it in terms of $lny_{i,t}$, equation [2] can be written:

$$lny_{i,t} = (1+\beta)ln\ y_{i,t-1} + \delta W_{i,\ t} + \mu_i + \eta_t + \varepsilon_{i,t} \qquad [3]$$

with the lagged dependent variable on the right-hand side.

Anderson and Hsiao propose to transform the model into first-differences to eliminate the fixed effects μ_i :

$$\Delta lny_{i,t} = (1+\beta)(ln\ y_{i,t-1} - ln\ y_{i,t-2}) + \delta(W_{i,t-1} - W_{i,\ t-2}) + (\eta_t - \eta_{t-1}) + (\varepsilon_{i,t} - \varepsilon_{i,t-1}) \qquad [4]$$

using $lny_{i,t-2}$ and lagged differences such as $(lny_{i,t-2} - lny_{i,t-3})$ as valid instruments for $(ln\ y_{i,t-1} - ln\ y_{i,t-2})$[15].

We estimate equation [4] annually for the whole time period 1995-2010 and the two sub-periods 1995-2000, 2001-2010.

Arellano and Bond (1991) and Blundell and Bond (1998) suggest the use of GMM estimators which are consistent for panels with small T and large N if the white noise disturbances are

---

[12] Thus having T = 4 periods and N=77*4 observations

[13] The whole sample is averaged over T=4 time periods 1995-1998, 1999-2002, 2003-2006 and 2007-2010. We run different specifications controlling for GDP 1995; 1999 and 1998; 2003 and 2002; 2007 and 2006.

[14] This outline is based on Roodman (2009). See Bond, Hoeffler and Temple (2001), Roodman (2009) for detailed explanations of GMMs and practical implementation using Stata package xtabond2. See the technical appendix for a detail description of dynamic panel data models.

[15] The same reasoning applies to the set of control variables $W_{i,t-1.}$

not correlated across observations. The Anderson-Hsiao estimator indeed is consistent but inefficient, while GMM estimators are efficient since all lags of the dependent variable can be used as instruments.

The system-GMM estimator was introduced to improve efficiency for *y* close to a random walk and it is now the conventional approach in the literature. This estimator differences the instruments to make them exogenous to the fixed effects and requires the additional assumption of no correlation between the first differences of the instrumenting variables and the fixed effects in the level equation. The validity of this additional condition can be tested using the difference-in-Sargan test.

Formally, the system-GMM is a system of two equations, one in first differences:

$$ln \ y_{i,t} - ln \ y_{i,t-1} = (1 + \beta)(ln \ y_{i,t-1} - ln \ y_{i,t-2}) + \delta(W_{i,t-1} - W_{i,t-2}) + (\varepsilon_{i,t} - \varepsilon_{i,t-1}) \qquad [5]$$

which uses lagged levels of $y_{i,t}$ and $W_{i,t-1}$ as instruments, and one in levels:

$$ln \ y_{i,t} = (1 + \beta)ln \ y_{i,t-1} + \delta W_{i,t-1} + \mu_i + \varepsilon_{i,t} \qquad [6]$$

which uses lagged first differences $\Delta lny_{i,t-1}$ and $\Delta W_{i,t-1}$ as instruments. In our analysis, we run several specifications using the system-GMM estimator and use different lags to construct the instruments set and report the Hansen and difference-in-Hansen tests for overidentifying restrictions and also Arellano/Bond tests for first-order (AR1) and second-order autocorrelation (AR2) in the white noise disturbance $\varepsilon_{i,t}$. AR(1) is generally found and points to the randomness of ε_{it}, whereas the AR(2) has to be insignificant for consistent estimates of the first-differences equations.

3.3 Distribution dynamics approach

In this section we briefly review the two most popular methodologies of non-parametric estimation of convergence: Markov transition probability matrices and kernel density estimators[16]. The advantage of non-parametric approaches consists in the possibility to estimate convergence without making assumptions about the distribution of the data. As Quah (1993) points out in his seminal paper, the idea underlying the regressions of growth on initial income is that the initial condition controls for transitory dynamics, while the conditioning

---

[16] This section is mainly based on Quah (1993, 1996, 1997) and Magrini (2004).

variables explain the trend. However, the assumption that every country has a stable growth path well-approximated by a time trend is maybe too strong to give reliable estimates of convergence. Consequently, we also look at convergence of Russian regions employing non-parametric methods.

Let F_t denote the distribution of per capita GRP across 77 Russian regions at time t. The evolution of F_t is described by the law of motion:

$$F_{t+1} = M \cdot F_t \qquad\qquad [7]$$

where M maps the current distribution at time t into a future distribution at time $t+1$. By iteration of equation [7], the process gives a predictor for the future cross-section distributions:

$$F_{t+s} = (M \cdot M \cdot ... M) * F_t = M^s \cdot F_t \qquad\qquad [8]$$

Allowing $s \to \infty$, it is possible to characterize the long-run cross-sectional distribution of incomes. If F_{t+s} tends toward a mass point, then the evidence is in line with the convergence hypothesis. If F_{t+s} instead tends toward bi- or multimodality, there is evidence of polarization. Finally, persistence could be signalled by immobility of the relative positions. In our empirical analysis, per capita GRP is normalized by the national average and the relative incomes set is divided into quintiles, so that M is described by a 5x5 Markov chain transition matrix.

The stochastic kernel is the continuous version of the transition probability matrix. The discrete nature of Markov chain transition probability matrix has been criticized (see Magrini 2004) because it is assumed that the data-generating process is time invariant and satisfies the Markov property, whereas the discretisation of a first-order continuous Markov process as is done in equation [8] is likely to remove the Markov property. An alternative method to discretisation is to interpret M as a stochastic kernel, where three-dimensional plots of relative incomes are analyzed to assess convergence. The definition of a kernel estimator can be defined as follows. Let $f = f(x)$ denote the continuous density function of a random variable X at point x and $(x_1, x_2, \ldots x_n)$ be the n observations from f. The kernel estimator $\hat{f}(x)$ is then defined as

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x_i - x}{h}\right) \qquad [9]$$

where K is the kernel and $h>0$ is a smoothing parameter called bandwidth. While Silverman (1986) demonstrates that the choice of the kernel between a normal, uniform, triangular of Epanechnikov function is a minor issue for estimation, he also shows that the choice of the bandwidth h is crucial. The common optimality criterion is to minimize the mean integrated squared error and Silverman (1986) shows that the minimum occurs when h is chosen as:

$$h = 0.9[min(\hat{\sigma}, R/1.34)]n^{-1/5} \qquad [10]$$

where $\hat{\sigma}$ is the sample standard deviation and R is the inter-quartile range.

3.4 Generalized Entropy Indices

We measure regional inequality applying various indices and decomposition techniques. The hierarchical structure is *federal district–region* for the analysis of the regional Rosstat data. First, we measure inequality over the period 1995-2010 through the three population-weighted indices related to the Generalized Entropy (GE) measures of inequality: Theil L (mean logarithmic deviation), the Theil T index and half the squared coefficient of variation. Throughout the paper, results are discussed for half the squared coefficient of variation. Inequality in GE indices is decomposed in a *within* and a *between* component following Shorrocks (1984), and by industrial sectors over the period 1998-2010 following Shorrocks (1982), Jenkins and Van Kerm (2008).

Generalized Entropy measures of inequality are determined using the following general formula:

$$GE(\alpha) = \frac{1}{\alpha(\alpha - 1)}\left[\frac{1}{N}\sum_{i=1}^{N}\left(\frac{y_i}{\bar{y}}\right)^{\alpha} - 1\right] \qquad \alpha \neq 0,1 \qquad [11]$$

with y_i representing income of observation i, $\bar{y}$ representing mean income and α a sensitivity parameter.

For α = 0, 1, 2 we have the mean logarithmic deviation (Theil L), Theil T, and half the squared coefficient of variation respectively. General Entropy varies between 0 and 1, with increasing inequality as the index approaches value 1. Parameter α reflects the sensitivity of the index to income differences in different parts of the income distribution: for values of α close to 0, the GE index is more sensitive to changes in the lower tail of the distribution, and vice versa for α close to 1. The General Entropy indices satisfy four properties called for when dealing with income inequality metrics: anonymity, scale independence, population independence, and the transfer principle (Shorrocks 1980, Cowell and Kuga 1981).

Substituting into [11], half the squared coefficient of variation can be written as:

$$GE(2) = \frac{1}{2N \overline{y}^2} \sum_{i=1}^{N} \left(y_i - \overline{y} \right)^2 \qquad [12]$$

As shown by Shorrocks (1982), half the squared coefficient of variation is additively decomposable both by population subgroups and by GRP components or income sources[17]. Decomposition by population subgroups allows to identify the role of *within-group* and *between-group* inequality, whereas the decomposition by factors shows the contribution of each factor to inequality. Jenkins (1995) performs both types of decomposition for the UK through the 1970s and 1980s, while Brewer et al (2009) extend the analysis up to 2007. Hoshino (2011) measures regional disparities in China since 1979, while Akita et al (2011) perform a bi-dimensional decomposition using data for Indonesia. Our study is related to Galbraith et al (2004) who find increasing between-groups inequality in Russia and China during the Nineties decomposing the Theil T index by industrial sectors until the year 2000.

This decomposition can be derived as follows. Consider a sample with *j={1,...,M}* groups where each unit *i={1,...,N}* belongs exclusively to one group[18]. Weights are defined as the share of the population belonging to unit *i* over total population *T*. Substituting into [12], the population-weighted half the squared coefficient of variation (WCV) can be written as:

---

[17] On the other hand, the Gini coefficient and the Theil indices can not be easily additively decomposable by income sources. See Aaberge et al (2005), Araar (2006), Jenkins and Van Kerm (2008) for further references.
[18] Groups and units are Federal Districts and regions in Rosstat data.

$$\text{WCV} = \frac{1}{2\overline{\overline{y}}^2} \sum_{j=1}^{M} \sum_{i=1}^{N_j} \frac{T_{ji}}{T} \left(\overline{y_{ji}} - \overline{Y} \right)^2 \qquad [13]$$

with $\overline{y_{ji}}$ representing per-head GRP of unit *i* belonging to group *j*; T_{ji} the population of unit *i* in group j and $\overline{Y}$ GRP per capita of all units.

Formula [13] can be additively decomposed in a *within-group* inequality component and a *between-group* inequality component as follows:

$$\text{WCV} = \frac{1}{2} \left\{ \sum_{j=1}^{M} \left(\frac{T_j}{T} \right) \left(\frac{\overline{Y_j}}{\overline{\overline{Y}}} \right)^2 CV(Y_j)^2 + CV(\overline{Y})^2 \right\} \qquad [14]$$

with the two components respectively expressed as follows:

$$I_W = \sum_{j=1}^{M} \left(\frac{T_j}{T} \right) \left(\frac{\overline{Y_j}}{\overline{\overline{Y}}} \right)^2 CV(Y_j)^2 \qquad \textit{within-group} \text{ inequality component} \qquad [14a]$$

$$I_B = CV(\overline{Y})^2 \qquad \textit{between-group} \text{ inequality component} \qquad [14b]$$

where [14a] represents a weighted sum of within-group inequalities and [14b] inequality across groups.

Shorrocks (1982) shows that there exists a unique decomposition rule such that inequality in total income can be expressed as the sum of the contribution to inequality by each factor component, also satisfying some basic axioms.

The contribution of each factor *k* to inequality in total income thus can be written as:

$$s_k = \frac{COV(Y^k, Y)}{\sigma^2(Y)} = \rho(Y^k, Y) x \frac{\mu_Y^k}{\mu_Y} x \left(\frac{GE_2(Y^k)}{GE_2(Y)} \right) \qquad [15]$$

where GE_2 is half the squared coefficient of variation.

Total inequality can be expressed as the correlations of the factors with total income, the factor shares in total income and the factor inequalities summarized by the GE index.

Those factors with positive ρ and therefore positive s_k have an augmenting effect on inequality, whereas the opposite is true for negative values.

4 Regression results

Results of the growth regressions are presented in tables 1-3. The results of the cross-sectional specifications in table 1, which estimate unconditional convergence, do not lend support to the convergence hypothesis. While all estimated coefficients are negative as expected if a convergence process is in place, the results are significant – albeit at the 10% level – only for the more recent time period 2005-2010 (column 5).

Table 1 Growth regression results, cross-sectional data

Dependent variable: average growth rate of per capita GRP					
	(1) 1995-2010	(2) 1995-2000	(3) 2001-2010	(4) 2000-2005	(5) 2005-2010
Initial GRP per capita	-.12 (.09)	-.04 (.05)	-.05 (.05)	-.004 (.04)	-.04 (.02)*
Observations	77	77	77	77	77
R^2	.05	.02	.02	.0002	.04

Notes: all variables are in natural logarithm. Symbols denote: * significance at 10%; ** significance at 5%; *** significance at 1%. Robust clustered S.E. at the regional level in parenthesis.

However, simple cross-section growth equations are generally not very informative because they do not control for unobserved heterogeneity and different steady states across regions. The LSDV models in table 2 provide estimation results of equation [3] using both annual data and averages over time intervals.

Table 2 Growth regression results, static panel data models

Dependent variable: growth rate of per capita GRP					
	Annual data			Interval averages		
	(1) 1995-2010	(2) 1995-2000	(3) 2001-2010	(4) 1995-2010	(5) 1995-2000	(6) 2001-2010
GRP per capita	.14 (.04)***	.65 (.17)***	.10 (.06)**	.09 (.03)***	.62 (.06)**	.10 (.03)***
Observations	1155	385	770	308	154	154
R^2	.52	.63	.50	.81	.64	.38

Notes: all variables are in natural logarithm. Symbols denote: * significance at 10%; ** significance at 5%; *** significance at 1%. All specifications control for regional and time effects. Clustered S.E. robust to heteroskedasticity and contemporaneously correlated disturbances in parenthesis. Robust clustered S.E. in columns (5-6) in parenthesis. Column (2) is averaged over 4-year intervals (T=4), column (5) is averaged over 3-year intervals (T=2), column (6) is averaged over 5-year intervals (T=2). In columns (1-3) GRP per capita is the GRP in the year prior to the one for which the growth rate is measured. In columns (4-6) GRP per capita is the GRP at the initial year of each interval. Results in columns (4-6) are robust to controlling for the GRP of the year prior to the beginning of the interval.

We find evidence against the unconditional convergence hypothesis in the whole period 1995-2010 both with annual data (column 1) and with data in 4-year time intervals[19] (column 4): the coefficients are positive (.12 and .09 respectively) and highly significant. In columns 2 and 3 and 5 and 6, the results over the sub-periods 1995-2000 and 2001-2010 show that though still not converging, indeed Russia's regions are less diverging over time, since the positive coefficients, indicating divergence, are .65 and .62 for the early years of crisis immediately after the collapse of the Soviet Union, whereas they have the value of .10 in the period of economic growth.

Further refining the analysis of unconditional convergence, we estimate several dynamic panel models with annual data employing Anderson-Hsiao and System-GMM estimators (Table 3).

Table 3 Growth regression results, dynamic panel data models

Dependent variable: annual growth rate of per capita GRP						
	Anderson-Hsiao			System-GMM		
	(1) 1995-2010	(2) 1995-2000	(3) 2001-2010	(4) 1995-2010	(5) 1995-2000	(6) 2001-2010
Lagged dep. variable	-.007 (.004) *	-.015 (.009)*	-.004 (.004)	.07 (.03)***	-.01 (.04)	.15 (.05)***
AR(1)				.00	.00	.00
AR(2)				.77	.40	.67
Hansen test				.10	.21	.10
Diff-in-Hansen				.23	.16	.03
N. instruments				43	13	31
Observations	1078	308	770	1155	385	770

Notes: all variables are in natural logarithm. Symbols denote: * significance at 10%; ** significance at 5%; *** significance at 1%. All specifications include time effects. Robust S.E. in parenthesis. The lagged dependent variable in the Anderson-Hsiao estimators columns (1-3) is instrumented with $y_{i,t-2}$. In system GMM we use robust Windmeijer S.E. to reduce finite sample bias. AR(1) and AR(2) report p-values for the Arellano and Bond's tests for first and second order residual serial correlation. The null hypothesis is no autocorrelation. Hansen and difference-in-Hansen report p-values for the tests for overidentifying restrictions. Difference-in-Hansen tests for the additional orthogonality conditions required by system-GMM i.e. that the instruments for the level equation are uncorrelated with the fixed effects. The null hypothesis is instrument exogeneity. In the baseline specification (4-6), instruments for the equation in differences are log per capita GRP lagged twice, instruments for the levels equation are log per capita GRP lagged and differenced once.

The negative coefficient of the lagged dependent variable in the Anderson-Hsiao estimators in columns 1-3 can be considered indicative of a very slow convergence process across Russia's regions. However, the results can be inefficient on two grounds. On the one hand, the results are not robust to lags going back further in time, on the other hand, the IV estimator for the

---

[19] The 4-year time intervals in column (4) are 1995-1998, 1999-2002, 2003-2006 and 2007-2010.

model in first differences does not make use of all possible orthogonality conditions nor does it take into account the structure of the differenced error term. Thus in columns 4-6 we report estimates of system-GMM regressions. To avoid the biases of instrument proliferation as *T* increases we reduce the instrument count (Roodman 2009). We also apply Windmeijer (2005) robust standard errors to reduce finite sample bias. In columns 4-6 we consider the lagged dependent variable as endogenous. For the whole period 1995-2010, both the Hansen and the difference-in-Hansen tests for overidentifying restrictions confirm that the instrument set is valid. There is first-order but no second-order autocorrelation according to the Arellano and Bond AR(1) and AR(2) tests, thus our model specification is correct. Finally, the p-values of all tests are not too high to suggest the invalidity of the system-GMM models (Roodman 2009).

The coefficient of the lagged dependent variable is statistically significant and provides evidence that Russia's regions are not yet converging. In comparison with the LSDV coefficient in column 1 of table 2 the estimated coefficient of the system-GMM indicates that the diverging process is indeed slower with respect to a static framework. Results for unconditional convergence the two sub-periods are instead weaker and the use of different instruments sets does not improve results: in column (5) the lagged dependent variable is not significant, whereas in column (6) the additional overidentifying conditions for the level equation are not valid. Overall, the results testing for unconditional convergence show that an unconditional convergence process is not yet found across Russia's regions, and that instead we have a process of divergence. This process has however slowed down over time.

Having established that Russia's regions are not unconditionally converging, we estimate conditional convergence controlling for a set of steady-state controls. Studying conditional convergence, we are interested at disentangling the role of resource-abundant regions[20] and fiscal federalism.

The Russian oil sector has been studied extensively at the national level[21] to assess whether the Russian economy suffers from the so-called Dutch disease, i.e. a pattern characterized by four symptoms: real appreciation of the currency, reduction of the manufacturing sector, development of the service sector, and increasing real wages. The evidence so far is mixed.

---

[20] Through the section, *resource* is used interchangeably with *oil, oil and gas*. We do not refer here to availability of other natural resources such as coal, iron, wood, gold because numerous Russia's regions are abundant of those raw materials. Moreover, *rentier regions* commonly identifies oil and gas abundant regions.

[21] See e.g. Beck, Kamps and Mileva (2007), Dobrynskaya and Turkisch (2010).

Becks et al (2007) do not find clear evidence of a decline of the manufacturing sector nor a consistent shift of the labour force from manufacturing to the mining sector through the last two decades. Consistent evidence is found by Dobrynskaya and Turkisch (2010) who also find an increase in manufacturing production, which posits against the existence of the Dutch disease. Overall the Russian industrial potential is still strong and the development of the service sector is driven by the transition process rather than being a Dutch disease symptom. The number of resource rich regions is limited in Russia and they are geographically bounded. To test whether the effect of being a resource abundant region has been constant over time, we control for a full set of interaction terms between a dummy variable for resource abundant regions and time effects.

As for fiscal federalism and federalism in general, during the last 20 years, Russia has gone through two opposite phases of de-centralization and re-centralization of power with the threshold being Putin's first Presidential term[22]. The whole time period is again separated in two periods 1995-2000 and 2001-2010 to disentangle different effects between the early phase of decentralization in the Nineties and the phase of recentralization of power which started with President Putin since the year 2000. We augment the basic specification of unconditional convergence with revenues (tax revenues, non-tax revenues and received grants) of the consolidated regional budget[23] as a percentage of GRP and the regional budget surplus. Since it is not possible to include simultaneously revenues and expenditures due to collinearity reasons, we include the budget surplus as a proxy for the regional budget constraint[24]. The vector of steady states controls also includes population size, consumer price index to control for price changes, and investments in fixed capital per capita[25].

Results for an annual fixed effects specification are shown in table 4.

Table 4 Results for conditional convergence, panel data - annual fixed effects specification.

Fixed effects annual panel data model			
Dependent variable: annual growth rate of per capita GRP			
	(1) 1995-2010	(2) 1995-2000	(3) 2001-2010
GRP per capita	.13 (.02) ***	.80 (.12) ***	.07 (.03) **
Population size	-.13 (.04) ***	-.42 (.13) ***	-.45 (.10) ***
Investments p.c.	.02 (.007) ***	-.03 (.01) *	.012 (.009)

[22] On the topic of federalism, see Desai, Freinkman and Goldberg (2005), Zhuravskaya (2010), Martinez-Vazquez and Timofeev (2010) among others.
[23] Revenues are collected within the region and credited to the federal, regional, and local budgets.
[24] See Hiebert et al (2012) for an application to the European Union
[25] Personnel engaged in R&D on 10,000 employed population, students in higher education institutions on 10,000 population as a proxy for human capital, number of hospital beds on 10,000 population as a proxy for public goods provision are not significant, excluded from regression results.

Consumer price index	.07 (.15)	.01 (.15)	.21 (.14)
Revenues (%GRP)	-.001 (.0007) *	-.001 (.001)	-.003 (.001) ***
Budget constraint (%GDP)	.001 (.001)	.001 (.001)	.002 (.001) **
Observations	1155	385	770

Notes: all variables are in natural logarithm. Symbols denote: * significance at 10%; ** significance at 5%; *** significance at 1%. All specifications include time effects. Robust S.E. in parenthesis. The interaction terms between a dummy for resource abundant regions and time effects are not shown for space reason. Authors' calculations using Rosstat data

On the convergence side, results confirm the absence of convergence of the unconditional analysis. The inclusion of additional explanatory variables slightly decreases the diverging path of Russia's regions, while population and investments are significant with the expected signs. The estimates of the interaction terms for resource abundant regions and a joint F test are significant and coefficients have negative sign, meaning that as time goes by, the effect of being a resource rich region becomes smaller with respect to the base year. Interesting insights emerge also from the conditional analysis, in particular for the sub-period 2001-2010: the negative sign of the tax revenues shows that an increase in taxation revenues and transfers is significantly associated with decreasing economic growth even controlling for the positive effect of a regional budget surplus, in the decade when most of regional autonomy has been recentralized and tax retention rates lowered. Our result relates to Desai et al (2003), who use data on tax retention and find a positive effect of the tax retention rate (as a proxy for fiscal autonomy) on regional growth in the period 1996-1999, before the process of recentralization began in the year 2000.

5 Distribution dynamics

Table 5 shows the one-step annual transition probability matrix, which contains the transition probabilities of moving between any two quintiles of relative income. It is constructed as the average of the annual transition probabilities over the whole period, based on transitions from end-of year to end-of-year. The sample is composed of 77 Russian regions over the period 1995-2010, for a total of 1,155 observations. Each cell shows both the number of observations which moved into that relative income quintile over the period and the transition probability[26].

---

[26] E.g., the first row shows that of the 230 observations in the first income quintile in 1995, 218 observations (i.e. 94.8 percent) could be found in the same quintile in 2010.

Table 5 Transition probability matrix, annual average transitions between 1995-2010

Origin quintile	Destination quintile				
Number (probability)	1st	2nd	3rd	4th	5th
230 1st	218 (94.8)	12 (5.2)	0	0	0
230 2nd	17 (7.4)	198 (86.1)	15 (6.5)	0	0
232 3rd	0	22 (9.5)	196 (84.5)	14 (6)	0
233 4th	0	0	16 (6.9)	203 (87.1)	14 (6)
230 5th	0	0	0	15 (6.5)	215 (96.5)

Authors' calculations using Rosstat data

The diagonal shows that the general feature is persistence, with the vast majority of observations being immobile in their initial quintile. Probabilities to move into other quintiles are always below 10 percent and transitions only take place between the two nearest quintiles. Moreover, the probability to move to the next higher income quintile is lower than the probability to move downward. Thus regions if they move in the distribution of GRP become poorer rather than richer in relative terms.

To get a long-run view of reallocation of regions in the income distribution we present the probability transition matrix for the whole period, i.e. transitions between t =1995 and t+τ =2010 (Table 6).

Table 6 Transition probability matrix, 16 years transitions between 1995-2010

Origin quintile	Destination quintile				
Number (probability)	1st	2nd	3rd	4th	5th
12	10 (83.3)	2 (16.7)	0	0	0
14	6 (42.9)	5 (35.7)	3 (21.4)	0	0
20	1 (5)	7 (35)	7 (35)	4 (20)	1 (5)
14	0	2 (14.3)	1 (7.14)	5 (35.7)	6 (42.9)
17	0	0	4 (23.5)	4 (23.5)	9 (52.9)

Authors' calculations using Rosstat data

The dynamics are more pronounced, in particular for those regions that start in the upper income quintiles, while the poorest regions, i.e. those in the first income quintile, remained trapped in relative poverty. Only 2 out of 12 regions moved out of the first into the second income quintile. On the other hand, only half of the richest regions in 1995 were still in the same position in 2010. Regions finding themselves in 1995 in the third and fourth income quintiles seem particularly mobile. For example, being in 1995 in the third quintile 7 regions out of 20 remained in their relative position, while 1 reached the highest income state and 1 fell to the lowest relative state. The overall evidence shows a tendency toward immobility for regions in low income quintiles and substantial mobility for regions higher up in the income distribution in 1995.

We also estimate transition probability matrices for the sub-periods 1995-2000 (table 7) and 2001-2010 (table 8) to capture the same time intervals that we employed in the regressions performed in the previous section. In the first phase after the break-up of the Soviet Union, the evidence shows a tendency towards the formation of two clubs, one of poor regions mainly entrapped in the first income quintile (only 2 out of 10 moved to the second quintile) and another one of rich regions in the top income class. These results indeed validate the regression results of the LSDV model in table 2, insofar as they show a very large divergence in the early phase of transition. In addition, inspection of the transition probability matrix allows us to infer that this divergence actually resulted in the formation of two clubs.

Table 7 Transition probability matrix, 6 years transition between 1995-2000

Number (probability)	1st	2nd	3rd	4th	5th
12	10 (83.3)	2 (16.7)	0	0	0
14	6 (28.6)	6 (42.9)	4 (28.6)	0	0
20	0	6 (30)	9 (45)	5 (25)	0
14	0	0	4 (28.6)	8 (57.1)	2 (14.3)
17	0	0	0	2 (11.8)	15 (88.2)

Authors' calculations using Rosstat data

During the second phase of economic transition, shown in table 8, the existence of the club of the poorest regions in the first income quintile is confirmed and actually has three more regions than in the early phase of transition. On the other hand, the club of regions with top

income is now more diversified since the probability to remain in the top quintile, which was almost 90 percent in the early phase of transition, declined to 70 percent, with 2 regions moving also to the second nearest income class. These dynamics are in line with the regression results that show decreasing divergence over time.

Table 8 Transition probability matrix, 10 years transition between 2001-2010

| Number (probability)	1st	2nd	3rd	4th	5th
15	14 (93.3)	1 (6.7)	0	0	0
15	3 (20)	8 (53.3)	4 (26.7)	0	0
16	0	7 (43.7)	7 (43.7)	2 (12.5)	0
17	0	0	2 (14.3)	9 (14.3)	6 (35.3)
14	0	0	2 (14.3)	2 (14.3)	10 (71.4)

Authors' calculations using Rosstat data

The stochastic kernel plot is the continuous representation of the probability transition matrix. The plot shows how the distribution of relative incomes evolved over the period. In figures 4-6 we represent results for the whole period 1995-2010 and the two sub-periods 1995-2000, 2001-2010 with relative income of the initial year in the x axis, relative income of the final year in the y axis, and the joint density in the z axis. The figures show multimodal distributions, which signal the absence of unconditional convergence. Looking in particular at figure 4 and 5, the z axis of the joint density shows that for a region with relatively low income at the beginning of the period, the probability of remaining at low income levels at the end of the period is high. Moreover, since most of the stochastic kernels are located in the 45° diagonal, mobility across the income distribution appears weak.
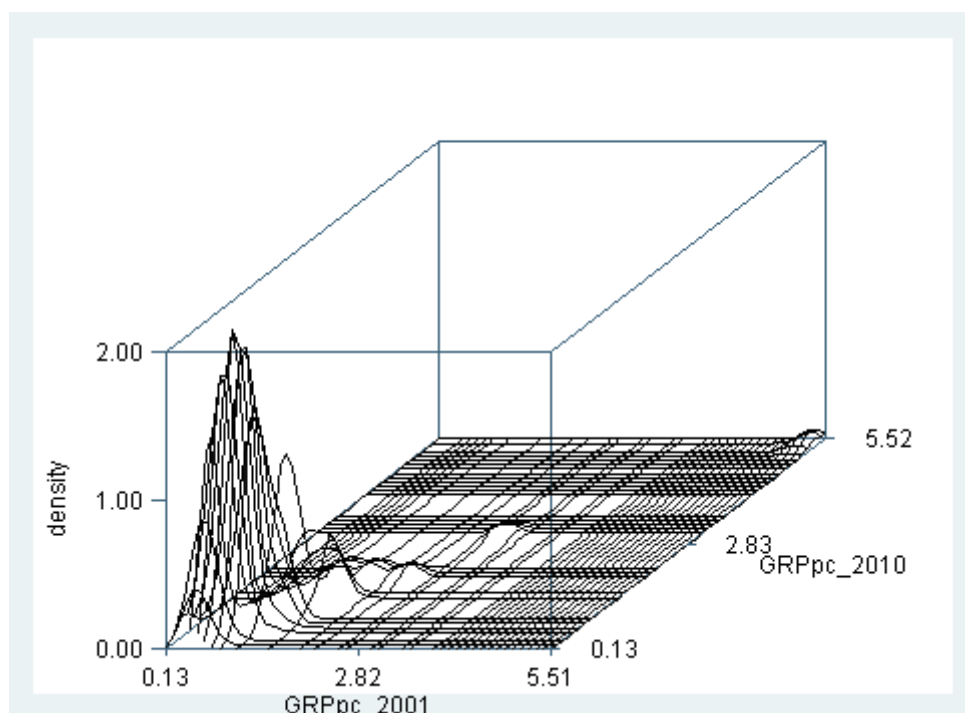
Figure 4 Stochastic kernel plot 1995-2010
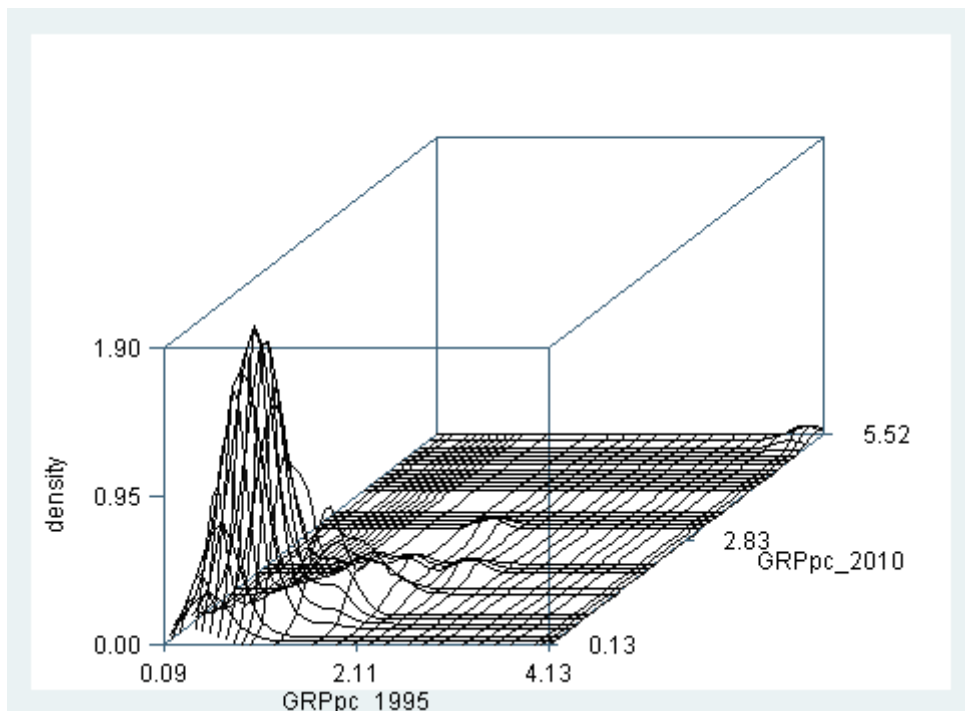
Authors' calculations using Rosstat data

Figure 5 Stochastic kernel plot 1995-2000
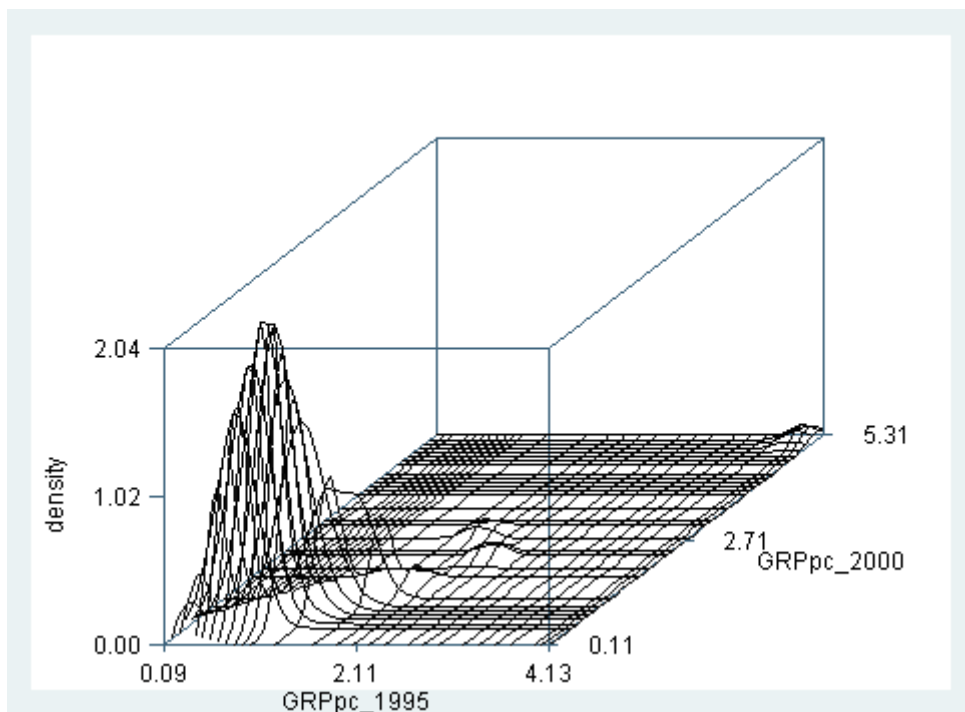
Authors' calculations using Rosstat data

Figure 6 Stochastic kernel plot 2001-2010

Authors' calculations using Rosstat data

6   Robustness checks

As a robustness check, we perform the analysis of regional convergence using the series of average annual income per capita which, with respect to GRP capita, may be a more precise proxy for regional living standards. Nominal income is converted into real using the regional consumer price index, at constant 1995 prices.

The evidence of regional divergence found in per capita GRP is confirmed with real regional income in all empirical models apart from the cross-section regressions and Anderson-Hsiao estimates.[27] The replicated cross section regressions of table 1 as well as the Anderson-Hsiao estimates using income instead of GRP show regional convergence, above all in the years 2001-2010 (columns 3 of tables 9 and 11). Guriev and Vakulenko (2012) find the same result for the cross section set-up, namely regional GDP-divergence together with income-convergence. They try to explain this puzzle estimating regional migration flows using a gravity model. Their central idea is that there exists a minimum income threshold in the sending region below which people are too poor to migrate. The overall economic growth in the 2000s allowed regions to exit from this poverty trap and thus resulted in income

---

[27] These latter estimates are, however, barely significant.

convergence in the 2000s. On the other hand, the authors justify divergence in GDP with total factor productivity differentials due to political and economic institutions, geography, etc., although a lack of detailed panel data on capital stocks and flows as well as on investment and the business climate makes it very difficult to test this hypothesis empirically.

Table 9 Growth regression results, cross-sectional income data

	Dependent variable: average growth rate of per capita income				
	(1) 1995-2010	(2) 1995-2000	(3) 2001-2010	(4) 2000-2005	(5) 2005-2010
Initial income per capita	-.35 (.06)***	-.05 (.05)	-.32 (.03)***	-.08 (.03)**	-.24 (.01) ***
Observations	77	77	77	77	77
R^2	.41	.02	.49	.13	.08

Notes: all variables are in natural logarithm. Symbols denote: * significance at 10%; ** significance at 5%; *** significance at 1%. Robust clustered S.E. at the regional level in parenthesis.

On the other hand, the replication of the models with panel data and the transition matrices computed with income do not provide evidence of convergence.

In the former setting, the convergence coefficient is positive like in the regressions with per capita GRP[28], while in the latter setting the probabilities in the diagonals are indeed still high, signalling persistence[29].

Table 10. Growth regression results, static panel data model

	Dependent variable: growth rate of per capita income		
	Annual data		
	1995-2010	1995-2000	2001-2010
Income per capita(t-1)	.21 (.04)***	.60 (.20)***	.21 (.07)***
Observations	1155	385	770
R^2	.88	.94	.47

Notes: all variables are in natural logarithm. Symbols denote: * significance at 10%; ** significance at 5%; *** significance at 1%. All specifications control for regional and time effects. Clustered S.E. robust to heteroskedasticity and contemporaneously correlated disturbances in parenthesis

---

[28] See footnote 28 for Anderson-Hsiao estimators
[29] Transition matrices available upon request

Table 11 Growth regression results, dynamic panel data models

	Dependent variable: annual growth rate of per capita income					
	Anderson-Hsiao			System-GMM		
	(1) 1995-2010	(2) 1995-2000	(3) 2001-2010	(4) 1995-2010	(5) 1995-2000	(6) 2001-2010
Lagged dep. variable	-.02 (.006)***	-.006 (.01)	-.03 (.006)***	.04 (.02)**	-.10 (.12)	.05 (.02)**
AR(1)				.00	.00	.00
AR(2)				.71	.14	.66
Hansen test				.09	.55	.05
Diff-in-Hansen				.43	.64	.20
N. instruments				43	13	31
Observations	1078	308	770	1155	385	770

Notes: all variables are in natural logarithm. Symbols denote: * significance at 10%; ** significance at 5%; *** significance at 1%. All specifications include time effects. Robust S.E. in parenthesis. The lagged dependent variable in the Anderson-Hsiao estimators columns (1-3) is instrumented with $y_{i,t-2}$. In system GMM we use robust Windmeijer S.E. to reduce finite sample bias. AR(1) and AR(2) report p-values for the Arellano and Bond's tests for first and second order residual serial correlation. The null hypothesis is no autocorrelation. Hansen and difference-in-Hansen report p-values for the tests for overidentifying restrictions. Difference-in-Hansen tests for the additional orthogonality conditions required by system-GMM i.e. that the instruments for the level equation are uncorrelated with the fixed effects. The null hypothesis is instrument exogeneity. In the baseline specification (4-6), instruments for the equation in differences are log per capita income lagged twice, instruments for the levels equation are log per capita income lagged and differenced once.

Table 12 Results for conditional convergence, panel data - annual fixed effects specification.

Fixed effects annual panel data model			
Dependent variable: annual growth rate of per capita income			
	1995-2010 (1)	1995-2000 (2)	2001-2010 (3)
Income per capita	.22 (.21)***	.55 (.05)***	.22 (.03)***
Population size	-.16 (.04)***	-.75 (.09)***	-.18 (.15)
Investments p.c.	-.006 (.007)	-.001 (.01)	-.01 (.01)
Consumer price index	-.82 (.06)***	-.69 (.07)***	-.84 (.14)***
Revenues (%GRP)	-.0001 (.0006)	.0001 (.0007)*	-.001 (.0007)
Budget constraint (%GDP)	-0002. (.0006)	-.001 (.0007)*	.0006 (.0004)
Observations	1155	385	770

Notes: all variables are in natural logarithm. Symbols denote: * significance at 10%; ** significance at 5%; *** significance at 1%. All specifications include time effects. Robust S.E. in parenthesis. The interaction terms between a dummy for resource abundant regions and time effects are not shown for space reason.

## 7   Decomposition of aggregate inequality measure

The aim of the decompositions of inequality is to shed light on two policy-related issues: (i) to identify the presence of geographical patterns in inequality, and (ii) to disentangle the contribution of sectoral GRP components in inequality throughout 1995-2010 to understand the drivers of non-convergence. These two issues can indeed be correlated in the Russian case, since the quarrying and mining sector is geographically bounded. The general trends of interregional inequality are discussed in section 7.1, the decompositions of half the squared coefficient of variation by districts and by industrial sectors are presented and discussed in sections 7.2 and 7.3 respectively. The results are discussed for five selected years: 1995, 1998, 2002, 2007 and 2010. These years encompass the two main phases of transition (i.e. the transition before and after the financial crisis of 1998), the period of economic growth, and the aftermath of the recent global financial crisis of 2008[30]. Inequality decomposition by factor components is not performed before 1998 due to data limitations.

7.1   General trends of interregional inequality

Figure 7 shows the dynamics of inequality in GRP per capita measured through Theil L, Theil T and the coefficient of variation between 1995 and 2010.

---

[30] The results for all years are available upon request.

Figure 7 Overall interregional inequality: population weighted Theil L, Theil T and CV
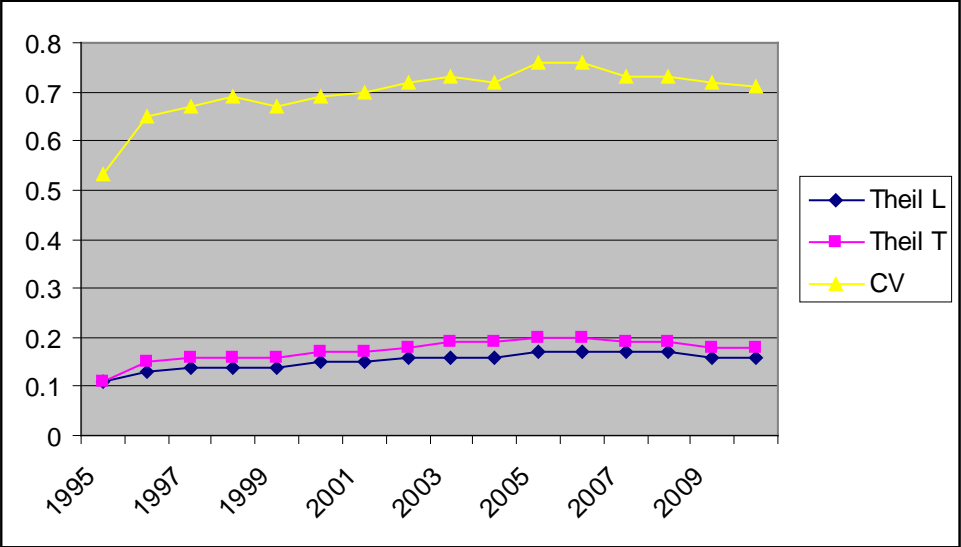
Source: authors' elaboration employing Rosstat data

Overall interregional inequality in per capita GRP exhibits an increasing trend since the beginning of transition from the planned economy up to the years 2000s. The steepest gradient occurs in the early stages of transition: between 1995 and 1998, as the Theil indices increased from .11 to .16 and the coefficient of variation jumped from .53 to .69. The trend then has been one of stable or slightly increasing inequality until the more recent years, which show slowly decreasing inequality in the wake of the global crisis of 2008.

7.2 Decomposition by groups

The hierarchical structure used for the decomposition by population groups is *federal districts–regions*. We adopt the pre-2010 federal structure, dividing Russia into seven federal districts. The partition of the regions into the federal districts is mutually exclusive and invariant over the period of analysis. Table 13 shows within-district inequality, which is calculated as inequality between regions within each district, as well as between-district inequality and total inequality for the five selected years

Table 13 Decomposition of population weighted half the squared coefficient of variation by groups

Decomposition of the population-weighted half the squared coefficient of variation by federal districts									
District	*Central*	*North-W*	*Southern*	*Volga*	*Urals*	*Siberian*	*Far East*	*WD*	*BD*	*Tot*
1995	.10	.04	.04	.04	.21	.04	.06	.10	.03	.14
1998	.20	.03	.04	.04	.35	.05	.06	.18	.05	.24
2002	.24	.02	.03	.05	.35	.04	.07	.21	.05	.26
2007	.26	.03	.02	.04	.30	.03	.07	.21	.06	.27
2010	.22	.03	.01	.04	.32	.03	.06	.19	.05	.25

Source: authors' elaborations of Rosstat data.

Inequality in GRP has been increasing over time in both components. The index was .14 in 1995 and reached .25 in 2010, with the highest value of .27 soon before the recent global financial crisis. Splitting the whole time period in phase 1 (1995-1998), phase 2 (1999-2007), and phase 3 (2008-onwards), most of the increase in inequality refers to phase 1, i.e. the period of crisis soon after the beginning of transition from the planned economy. In the following years in fact, inequality has maintained an increasing trend, but at a lower pace: the index increased from .14 to .24 in phase 1, from .22 to .27 in phase 2. Eventually, the trend seems to be reversing with the recent financial crisis, but the period of analysis is too short to capture any consistent trend.

Most of overall interregional inequality is due to the within-district (i.e. between regions) component, whereas the between-districts component is increasing over time, but its magnitude is lower. Among the seven federal districts, the Central and the Urals districts show the highest contribution to within-district between-regions inequality and overall interregional inequality. Inequality within the Central district doubled from value .10 in 1995 to .20 in 1999, at the beginning of phase 2. For the Urals district, the values are .21 and .33 respectively. Indeed these districts represent the two diverging factors of the Russian economy: the concentration of activities and businesses in the Moscow area and the oil extractive sector in the Urals. On the other hand, these two regions are on the opposite sides as for population shares since ten percent of the total Russian population lives in the Moscow area (and 26 percent in the whole Central district) whereas only 8 percent of the Russian population live in the four regions of the Urals.

To understand the contribution of Moscow and Tyumen to inequality patterns, the analysis is repeated excluding these two regions. The results show that once these areas are not in the sample, the patterns of the seven districts are rather homogeneous over time and space. Excluding Moscow, the index within the Central district decreases to .02 in 1995 and .03 in

2010, and the within-district inequality component also decreases yearly. Excluding Tyumen, the values for the Urals district are .02 and .03 respectively, with also a decrease in the within-district inequality component[31].

The decomposition by groups thus allows us to answer our first policy-related issue regarding spatial inequality: the districts with the lowest between-regions inequality are those which, after the collapse of the Soviet Union and the dismantling of the large state-owned firms, have succeeded in the diversification of their economic activities. This is indeed the case for the North-Western and the Volga districts, where there are no outliers with concentration of economic activities and businesses, nor full specialization. On the other hand the financial sector, public agencies and ministries are concentrated in Moscow, and the Urals area is the geographical core for oil-related activities. The Southern district instead, in the Caucasus, is specialized in agriculture in a homogeneous pattern across the regions and shows limited within-district inequality, but the unstable political situation of a few regions like Chechnya is highly detrimental to the creation of business activities. It is thus worth stressing that results for the Southern district, which is the poorest and mostly populated one, crucially depend on the political situation of each regions, given that instability and wars over time have prevented the recovery and development of the area[32]. Taking a convergence perspective, the patterns of inequality, which has been increasing at a decreasing pace, confirm the regressions' results, which show that regions were indeed diverging throughout 1995-2010 but at a decreasing rate after the financial crisis of 1998.

7.3 Decomposition by GRP factors

Tables 14-17 show results for the decomposition of total inequality by gross value added at the industrial level. The reported results are the proportional contribution of each factor to total inequality s_k, the share of each component in total GRP $share\_k$, and half the squared coefficient of variation for each factor WCV_k. The results are weighted by regional population shares. The structure of gross value added has been extended after 2002, leading to more detailed decompositions. In particular, for the years 1998-2002 the energy sector (i.e. mining and quarrying, electricity, gas, and water) is not identified and subsumed under the broad section "physical goods", which in turn is totally decomposed into manufacturing, agriculture, and construction. For "intangible services", we define as a residual sector "other

---

[31] These results are not shown here but available upon request.
[32] Due to data availability, Chechnya and Ingushetia are excluded from the sample. The exclusion of the two poorest regions of the district probably biases inequality within the Southern district downwards.

services", i.e. the share of gross value added not related to transport, communications, and trade and catering.

For the years 2007 and 2010, the energy sector is divided into mining and quarrying activities, and electricity, gas, and water. Social services are divided into public administration and defence, education, health, and residual services. Due to collinearity problems, the financial sector is not included in the decomposition. However, the financial sector contributes in a very limited fashion to total GRP (the share at the national level was 0.6 percent in 2010) and its value is nil in large parts of the country, with the exception of the Moscow area.

In 1998 (table 14), the largest component of gross value added is the manufacturing sector (33 percent), followed by the residual sector "other services" and "trade & catering." Manufacturing also has the largest proportional contribution to inequality according to Shorrocks' decomposition rule in equation [15], followed by agriculture. According to equation [15], the proportional contribution to inequality is determined by the share of the factor in total GRP, the correlation of the factor with total GRP, and its own inequality. However the inequality augmenting effect of agriculture is cancelled out by the equalizing effect of the trade & catering sector.

Table 17 Inequality decomposition of the GE(2) index by factor components, year 1998

GRP factors year 1998	s_k	share_k	WCV_k
Manufacturing	.76	32.56	.05
Agriculture	.36	8.4	.23
Constructions	-.085	7.45	.03
Transports	.12	9.84	.10
Communications	-.022	1.97	.08
Trade&catering	-.33	13.81	.09
Other services	.19	25.9	.01
Total	1	100	-

Notes: authors' calculations using Rosstat data

The factor with the highest inequality measure (WCV_k) is agriculture: this sector is unevenly distributed over the country and its contribution to gross value added is higher in the poorest area, i.e. the densely-populated Southern district. Results of the decompositions in table 13 show that for the whole period of analysis, within-district inequality in the Southern district is very limited and lower than the overall between-district component: it is thus likely that the high contribution to inequality of the agricultural sector is signalling that the whole area is lagging behind in the process of catching-up with the other Russian regions. In the year 2002 and the years in between, the picture is generally unchanged, with the exception of the construction sector whose contribution to inequality s_k changes sign, loosing its characteristic of equalizing effect through the years of economic growth. For both years, the residual category "other services" (i.e. all services excluding transports, communications, trade&catering) covers a share around one forth of the total gross value added, though it is not possible to identify any sub-sector.

Table 15 Inequality decomposition of the GE(2) index by factor components, year 2002

GRP factors year 2002	s_k	share_k	WCV_k
Manufacturing	.93	31.37	.07
Agriculture	.51	9.93	.23
Constructions	.07	7.81	.06
Transports	.34	8.88	.12
Communications	-.06	1.8	.11
Trade&catering	-1	17.3	.21
Other services	.21	22.8	.01
Total	1	100	-

Notes: authors' calculations using Rosstat data

Manufacturing gives the largest contribution to gross value added also in 2007 and 2010, followed by wholesale, retail trade, and repairs. The latter has the largest equalizing contribution to the reduction of inequality (s_k = -8.83 and s_k= -23 in 2007 and 2010

respectively); on the other hand, though not geographically concentrated, the proportional contribution of manufacturing is inequality augmenting, but the values are rather limited ($s_k =$ +2.92 in 2010).

Table 16 Inequality decomposition of the GE(2) index by factor components, year 2007

GRP Factors year 2007	s_k	share_k	WCV_k
Agriculture and fishing	4.13	7.40	.24
Mining	3.01	6.79	1.58
Manufacturing	1.23	21.83	.11
Electricity, gas, water	.84	4.05	.12
Constructions	1.36	7	.07
Wholesale retail trade	-8.83	18.26	.08
Hotels	-.03	1.09	.11
Trasports&communications	1.20	11.09	.07
Real estate	-4.35	8.14	.11
Public administration	1.54	5.42	.10
Education	.81	3.38	.06
Health	.83	4.05	.05
Other services	-.76	1.43	.13
Total	1	100	-

Notes: authors' calculations using Rosstat data

Table 17 Inequality decomposition of the GE(2) index by factor components, year 2010

GRP Factors year 2010	s_k	share_k	WCV_k
Agriculture and fishing	9.38	6.71	.26
Mining	15.91	7.05	1.46
Manufacturing	2.92	19.83	.10
Electricity, gas, water	.95	4.95	.10
Constructions	6.10	7.73	.15
Wholesale retail trade	-23	17.23	.09
Hotels	.12	1.14	.14
Trasports&communications	1.27	11.1	.05
Real estate	-15.8	9.10	.14
Public administration	2.99	5.98	.08
Education	1.11	3.56	.05
Health	.73	4.30	.04
Other services	-1.74	1.25	.08
Total	1	100	-

Notes: authors' calculations using Rosstat data

The real estate sector, which was not identified in the years 1998 and 2002, covers about 10 percent of gross value added with a large contribution to the reduction of inequality, which is increasing over the time period of growth. The detailed decomposition of the service sector in the years 2007 and 2010 shows that public administration, education, and health all together contribute to slightly less than 15 percent of gross value added and have a limited, but

positive impact on inequality, meaning that social services are not a converging factor neither during the crisis nor in periods of growth. The effect is however partly offset by the equalizing contribution of "other personal and social services" in recent years (s_k = -.76 and s_k= -1.74 in 2007 and 2010 respectively).

The energy sector, which was also not identified in the two previous years, accounts for slightly less than 15 percent of gross value added in 2007 and 2010, with the highest contribution from mining and quarrying activities (7 percent in both years). On the one hand, the mining sector *per se* does not appear to have a peculiar role in the global composition of GRP, in fact its magnitude is comparable to those of the agriculture and fishing sectors or constructions; on the other hand, the proportionate contribution to overall inequality and the factor WCV show that mining is the first contributor to inequality. Looking at the *WCV_ks* in recent years, the value of the mining sector is the highest one with values 1.5-1.6, whereas the values for other sectors are below .15, with the exception of agriculture as in previous years.

From a convergence perspective, the decomposition of inequality by GRP components confirms the results of the decomposition by groups and the regression results. Moreover, it helps to shed light on the reasons why Russia's regions are not yet converging, and on which areas and industrial sector policy intervention should be concentrated. The Russian convergence pattern is characterized by the formation of two clubs, and the two driving sectors of inequality are agriculture and mining. One club is composed by poor, highly-populated regions specialized in agriculture which indeed show low inequality between themselves but contribute highly to the overall level of inequality; the other club is instead composed by those regions which are resource abundant and scarcely-populated, and the Moscow area. Finally, despite the reforms to equalize the provision of public goods across Russia, the social services' sector of the public administration, education, and health still do not convey the expected equalizing effect across regions: they are still – albeit in a limited proportion – contributing to inequality across regions rather than attenuating them.

8. Conclusions

Using regional-level data provided by Rossstat, this paper analyzes convergence in per capita gross regional product of Russia's regions during the period 1995-2010, when reliable regional data are available. We both analyze beta- and sigma-convergence in the paper.

Using various models of the determination of growth of per capita gross regional product and different time intervals, i.e., the whole period as well as periods before and after the 1998 financial crisis, we do not find unconditional or conditional beta-convergence of Russian regions. The models that we employ span the gamut of simple static OLS models to system GMM panel data models. The findings derived from these different models are confirmed with robustness checks that employ relative real regional income instead of gross regional product.

Since Barrow-type regressions that we employ are criticized for assuming a stable growth trend for all economies (or for all regions) we also test convergence within a non-parametric framework. To this purpose we estimate Markov transition probability matrices as well as stochastic kernels of relative regional income. The transition probability matrices are estimated over the whole period (1995-2010) but also over the time interval of the late nineties as well as over the first decade of the new century. The diagonal elements dominate strongly in all these estimated transition probability matrices, thus pointing to strong persistence in the relative distribution of income and a lack of convergence. The stochastic kernel plots show very similar patterns of the dynamics of relative regional income.

As far as sigma-convergence is concerned, decompositions of regional income and gross regional product also find no convergence of this type for Russian regions. These decompositions point to the geographical concentration of extractive activities in the Urals and of business services and of the public administration in the Moscow area as the main culprit for this lack of convergence. They also establish that despite reforms to equalize provisions of public goods across Russia, the social services sector of the public administration, education and health still do not have the expected equalizing impact on regional income, but rather contribute to inequality across Russia's regions.

References

Aaberge R., Bjerve S. and Doksum K. (2005). Decomposition of Rank-Dependent Measures of Inequality Across Groups. *International Journal of Statistics* 2005, Vol. LXIII, n. 3, pp 493-503.

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

Ahrend R. (2006). How to Sustain Growth in a Resource Based Economy? The Main Concepts and their Application to the Russian Case. VOPROSY ECONOMIKI, N.P. Redaktsiya zhurnala "Voprosy Economiki", vol. 7.

Ahrend R. (2008). Understanding Russian Regions' Economic Performance During Periods of Decline and Growth – An Extreme Bound Analysis Approach. Economics Department, Working Paper 52, OECD.

Ahrend R., Tompson W. (2005). Fifteen Years of Economic Reform in Russia: What Has Been Achieved? What Remains To Be Done? *OECD Economics Department*, Working Paper No. 430.

Akhmedjonov A., Lau M. C. K. and Izgi B. B. (2013). New Evidence of Regional Income Divergence in post-reform Russia. *Applied Economics*, Volume 45, Issue 18.

Akita T., Kurniawan P.A. and Miyata S. (2011). Structural Changes and Regional Income Inequality in Indonesia: A Bidimensional Decomposition Analysis. *Asian Economic Journal* 2011, Vol. 25, No. 1, 55-77.

Anderson T.W. and Hsiao C. (1981). Estimation of Dynamic Models with Error Components. *Journal of American Statistical Association* 76, 598–606.

Araar A. (2006). On the Decomposition of the Gini Coefficient: an Exact Approach, with an Illustration using Cameroonian Data. *CIRPEE* WP 06-02.

Arellano M. and Bond S. (1991). Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations. *Review of Economic Studies*, 58, 277-297.

Arellano M., and Bover O. (1995). Another Look at the Instrumental Variable Estimation of Error-components Models. *Journal of Econometrics* 68: 29–51.

Armstrong H. (1995). Convergence Among Regions of the European Union, 1950-1990. *Papers in Regional Science,* Vol. 74, Issue 2, pp. 143-152, April 1995.

Azariadis C. and Drazen A. (1990). Threshold Externalities in Economic Development. *Quarterly Journal of Economics* MIT Press, Vol. 105(2), pages 501-26, May.

Barro R.J. (1991). Economic Growth in a Cross-Section of Countries. *Quarterly Journal of Economics*, 106, 407-443.

Barro R.J. (2012) Convergence and Modernization Revisited. *NBER Working Paper Series* WP 18295.

Barro R.J. and Sala-i-Martin X. (1991). Convergence Across States and Regions. *Brooking Papers on Economic Activity*, no. 1, 107-158.

Barro R.J. and Sala-i-Martin X.(1992). Convergence. *Journal of Political Economy,* 100, 223-251.

Barro R.J. and Sala-i-Martin X. (1995). *Economic Growth,* New York: Mc Graw Hill.

Barro R.J. and Sala-i-Martin X. (2004). *Economic Growth 2nd Edition*, Cambridge MA, MIT Press.

Baumol W. (1986). Productivity Growth, Convergence and Welfare: What the Long-Run Data Show. *American Economic Review*, 76, 1072-1085.

Beck R., Kamps A., and Mileva E. (2007). Long-term Growth Prospects for the Russian Economy. *ECB Occasional Paper Series* No. 58/March 2007.

Berkowitz D., and DeJong D. N. (2003). Policy Reform and Growth in Post-Soviet Russia. *European Economic Review 47*, pp. 337-352.

Berkowitz D., and DeJong D. N. (2005). Entrepreneurship and Post-socialist Growth. *Oxford Bulletin of Economics and Statistics,* 67, I, pp. 25-46.

Bianchi M. (1997): Testing for Convergence: Evidence from Non-parametric Multimodality Tests, *Journal of Applied Econometrics*, 12; pp. 393-409.

Blundell R., and Bond S. (1998). Initial Conditions and Moment Restrictions in Dynamic Panel Data Models. *Journal of Econometrics* 87: 115–143.

Blundell R., Bond S. and Windmeijer F. (2000). Estimation in Dynamic Panel Data Models: Improving on the Performance of the Standard GMM Estimator. *IFS Working Papers* W00/12, Institute for Fiscal Studies.

Bond S. , Hoeffler A. and Temple J. (2001). GMM Estimation of Empirical Growth Models. *CEPR Discussion Paper* No. 3048.

Borys M., Polgar E. and Zlate A. (2008). Real Convergence and the Determinants of Growth in EU Candidate and Potential Candidate Countries. A Panel Data Approach. *ECB Occasional Papers* No. 86/June 2008.

Brewer M., Muriel A. and Wren-Lewis L. (2009). Accounting for Changes in Inequality Since 1968: Decomposition Analyses for Great Britain. *ISER Working Paper Series* 2012-17, Institute for Social and Economic Research.

Buccellato T. (2007). Convergence Across Russian Regions: A Spatial Econometrics Approach. Research Paper 053_DP70 University of London, the Centre for Financial and Management Studies.

Caselli F., Esquivel G., and Lefort F. (1996). Reopening the Convergence Debate: A New Look at Cross-Country Growth Empirics. *Journal of Economic Growth*, 1996.

Cowell F.A. and Kuga K. (1981). Additivity and the Entropy Concept: An Axiomatic Approach to Inequality Measurement. *Journal of Economic Theory,* Elsevier, vol. 25(1), pages 131-143, August.

Denisova I. (2012). Income Distribution and Poverty in Russia. *OECD Social, Employment and Migration Working Papers,* No. 132, OECD Publishing.

Desai R.M., Freinkman L. and Goldberg I. (2005). Fiscal Federalism in Rentier Regions: Evidence from Russia. *Journal of Comparative Economics,* 33(4), pp. 814-834.

Dobrynskaya V. and Turkisch E. (2010). Economic diversification and Dutch disease in Russia. *Post-Communist Economies,* Vol. 22, No. 3, September 2010, 283-302.

Dobson S., Ramlogan-Dobson C. and Strobl E. (2012). Convergence or Divergence in Cross-Country Growth? *International Review of Applied Economics*, 26:3, 417-424.

Dolinskaya I. (2002). Transition and Regional Inequality in Russia: Reorganization or Procrastination? *IMF Working Paper* No. WP/02/169.

Durlauf S. N. and Johnson P.A. (1995). Multiple Regimes and Cross-Country Growth Behaviour. *Journal of Applied Econometrics* 10 (4), 365–84.

Eberhardt M. and Teal F. (2011). Econometrics for Grumblers: A New Look at the Literature on Cross-Country Growth Empirics. *Journal of Economic Surveys,* Vol. 25(1), pp.109-155.

Galbraith J.K., Krytynskaia L. and Wang Q. (2004). The Experience of Rising Inequality in Russia and China During Transition. *The European Journal of Comparative Economics* Vol. 1, n. 1, pp 87-106.

Gluschenko K. (2010). Methodologies of Analyzing Inter-Regional Income Inequality and Their Applications to Russia. *William Davidson Institute* Working Paper N. 984

Guriev S. M. and Vakulenko E. (2012). Convergence between Russian regions. *CEFIR/NES* WP No. 180, 2012.

He Q. and Sun M. (2013). Dances with Chinese Data: Are the Reform Period Chinese Provincial Panel Data Reliable? *Annals of Economics and Finance* 14-2, 529-551 (2013).

Hiebert P., Lamo A., Avila D. and Vidal J. (2012). Fiscal Policies and Economic Growth in Europe: an Empirical Analysis. Available at SSRN: http://ssrn.com/abstract=2094444.

Hoeffler A.E. (2002). The Augmented Solow Model and the African Growth Debate. *Oxford Bulletin of Economics and Statistics,* Vol. 64, Issue 2, pp.135-158, May 2002.

Hoshino M. (2011). Measurement of GDP per capita and Regional Disparities in China, 1979-2009. *Discussion Paper Series* DP 2011-17, Kobe University.

Islam N. (1995). Growth Empirics: A Panel Data Approach. *The Quarterly Journal of Economics,* Vol. 110, No. 4 (Nov 1995), pp. 1127-1170.

Islam N. (2003). What Have We Learned from the Convergence Debate? *Journal of Economic Surveys,* Vol. 13, No. 3.

Jenkins S. P. (1995). Accounting for Inequality Trends: Decomposition Analyses for the UK, 1971-86. *Economica*, New Series, Vol. 62, No. 245 (Feb., 1995), pp. 29-63.

Jenkins S.P.and Van Kerm P. (2008). The Measurement of Economic Inequality. Prepared for the Oxford Handbook on Economic Inequality edited by Brian Nolan, Wiermer Salverda and Tim Smeeding, 2008.

Kholodilin K.A., Oshchepkov A., Siliverstovs B. (2009). Russian Regional Convergence Process: Where Does It Go? *KOF* WP 216

Kontorovich V. (2010) Can Russia Resettle the Far East? *Post-Communist Economies,* 12:3, pp. 365-384.

Lau C. K.M. (2010a). Convergence across the United States: Evidence from Panel ESTAR Unit Root Test. International Advances in Economic Research, 16(1), 52-64.

Lau C.K.M. (2010b). New Evidence About Regional Income Divergence in China. China Economic Review, 21 (2010) 293 – 309.

Ledyaeva S. (2009). Spatial Econometric Analysis of Foreign Direct Investment Determinants in Russian Regions. *World Economy*, Vol. 32, Issue 4, pp. 643-666, April 2009.

Ledyaeva, S., and M. Linden (2008). Determinants of Economic Growth: Empirical Evidence from Russian Regions. *European Journal of Comparative Economics*, Vol. 5, n. 1, pp. 87-105.

Lee K., Pesaran M.H. and Smith R. (1997). Growth and Convergence in a Multi-country Empirical Stochastic Solow Model. *Journal of Applied Econometrics*, 12, pp. 357-392.

Lugovoy, O., Dashkeyev V., Mazayev I., Fomchenko D., Polyakov E., and Hecht A. (2007). Analysis of Economic Growth in Regions: Geographical and Institutional Aspect. *Moscow: IET*.

Magrini S. (2004). Regional (Di)Convergence. *Handbook of Regional and Urban Economics*, Volume 4, edited by Vernon Henderson and Jacques-François Thisse.

Mahler C. (2011). Divergence Fortunes: Recent Developments in Income Inequality Across Russian Regions. *Opticon 1826*, Issue 10, Spring 2011.

Mankiw G.N., Romer D. and Weil D. (1992). A Contribution to the Empirics of Growth, *Quarterly Journal of Economics,* 107, 407-437.

Martinez-Vazquez J. and Timofeev A.(2010). Inter-regional Equalization and Growth in Russia. Georgia State University, Andrew Young School of Policy Studies. International Studies Program, WP 10-11, April 2010.

Mikheeva, N. (1999). Analysis of Interregional Inequality in Russia. *Studies on Russian Economic Development*, 10 (5), pp. 514-521.

Popov V. (1999). Reform Strategies and Economic Performance of Russia's Regions. *World-Development*, 29(5), 865-86.

Prochniak M. and Witkowski B. (2013). Time Stability of the Beta Convergence Among EU Countries: Bayesian Model Averaging Perspective. *Economic Modelling* 30(2013) 322-333.

Quah D. (1993). Empirical Cross-Section Dynamics in Economic Growth. *European Economic Review*, Elsevier, Vol. 37, No.2-3, pp. 426-434.

Quah D. (1996). Twin Peaks: Growth and Convergence in Models of Distribution Dynamics. *The Economic Journal*, 106 (July), 1045-1055.

Quah D. (1997) Empirics for Growth and Distribution: Stratification, Polarisation and Convergence Clubs. *Journal of Economic Growth*, 2(1), 27-59.

Rapacki R. and Prochniak M. (2009). Real Beta and Sigma Convergence in 27 Transition Countries, 1990–2005. *Post-Communist Economies*, 21:3, 307-326.

Roodman D. (2009). How to do xtabond2: An introduction to difference and system GMM in Stata. *The Stata Journal (2009)* 9, Number 1, pp. 86-136.

Silverman, B. W. (1986). Density Estimation for Statistics and Data Analysis, Monographs on Statistics and Applied Probability, Chapman and Hall, London, UK.

Shorrocks A.F. (1980). The Class of Additively Decomposable Inequality Measures. *Econometrica,* Vol. 48, No. 3 (Apr. 1980), pp. 613-625.

Shorrocks A.F. (1982). Inequality decomposition by factor components. *Econometrica,* Vol. 50, No. 1 (Jan 1982), pp. 193-212.

Shorrocks A.F. (1984). Inequality Decomposition by Population Subgroups. *Econometrica,* Vol .52, No. 6 (Nov 1984), pp. 1369-1385

Solanko L. (2003). An Empirical Note on Growth and Convergence across Russian Regions. *BOFIT Discussion Papers* 9/2003, Bank of Finland, Institute for Economies in Transition.

Solanko L. (2008). Unequal Fortunes: A Note on Income Convergence Across Russian Regions. *Post-Communist Economies*, Vol. 20, No. 3, 287–301.

Solow R. (1956). A Contribution to the Theory of Economic Growth, *Quarterly Journal of Economics,* 70, 65-94.

Windmeijer F. (2005). A Finite Sample Correction for the Variance of Linear Efficient two-step GMM Estimators. *Journal of Econometrics* 126: 25–51.

Zhuravskaya E. (2010). Federalism in Russia. *CEFIR/NES* WP No. 141.

Appendix

List of Russian regions and Federal Districts

Central Federal District

Belgorod Oblast, Bryansk Oblast, Vladimir Oblast, Voronezh Oblast, Ivanovo Oblast, Kaluga Oblast, Kostroma Oblast, Kursk Oblast, Lipetsk Oblast, Moscow Federal City, Moscow Oblast, Orel Oblast, Ryazan Oblast, Smolensk Oblast, Tambov Oblast, Tver Oblast, Tula Oblast, Yaroslav Oblast

North-Western Federal District

Arkhangelsk Oblast, Vologda Oblast, Kaliningrad Oblast, Republic of Karelia, Komi Republic, Leningrad Oblast, Murmansk Oblast, Novgorod Oblast, Pskov Oblast, Saint Petersburg Federal City

Southern Federal District

Republic of Adygea, Republic of Dagestan, Republic of Ingushetia, Kabardino-Balkaria Republic, Karachay-Cherkessia Republic, Republic of North Ossetia-Alania, Republic of Chechnya, Stavropol Krai, Astrakhan Oblast, Volgograd Oblast, Republic of Kalmykia, Krasnodar Krai, Rostov Oblast

North Caucasian Federal District *established in 2010*

Starting from 2010, the regions of Dagestan, Ingushetia, Kabardino-Balkaria, Karachay-Cherkessia, North Ossetia-Alania, Chechnya and Stavropol form the North Caucasian Federal District

Volga Federal District

Republic of Bashkortostan, Republic of Tatarstan, Republic of Udmurtia, Chuvash Republic, Mari El Republic, Kirov Oblast, Republic of Mordovia, Nizhny Novgorod Oblast, Orenburg Oblast, Penza Oblast, Perm Krai, Samara Oblast, Saratov Oblast, Ulyanovsk Oblast

Urals Federal District

Kurgan Oblast, Sverdlovsk Oblast, Tyumen Oblast, Chelyabinsk Oblast

Siberian Federal District

Republic of Altai, Altai Krai, Republic of Buryatia, Irkutsk Oblast, Kemerovo Oblast, Tuva Republic, Republic of Khakassia, Krasnoyarsk Krai, Novosibirsk Oblast, Omsk Oblast, Tomsk Oblast, Zabaykalsky Krai

Far Eastern Federal District

Sakha-Yakutia Republic, Primorsky Krai, Khabarovsk Krai, Kamchatka Krai, Magadan Oblast, Amur Oblast, Jewish Autonomous Okrug, Sakhalin Oblast, Chukotka Autonomous Okrug

Technical Appendix
Dynamic panel data models

The static panel data model of equation [2] takes the form:

$$ln\ y_{i,t} - ln y_{i,t-1} = \beta ln\ y_{i,t-1} + \delta W_{i,t-1} + \mu_i + \eta_t + \varepsilon_{i,t}$$

This model faces two drawbacks: (i) the model is actually a dynamic one, where correlation between the lagged endogenous dependent variable $y_{i,t-1}$ on the right-hand side and the region-specific fixed-effects μ_i can lead to inconsistent estimates; (ii) some of the control variables used in the estimation of the conditional convergence models might also be endogenous.

Since the results of the LSDV regressions can be inconsistent because of the correlation between $ln y_{i,t-1}$ and the fixed-effects μ_i, Anderson and Hsiao (1981) suggest to estimate the growth equation in first differences using lags as instruments for the endogenous covariates. As seen in the text, rearranging it in terms of $ln y_{i,t}$, equation [2] can be written:

$$ln y_{i,t} = (1+\beta)ln\ y_{i,t-1} + \delta W_{i,t} + \mu_i + \eta_t + \varepsilon_{i,t} \qquad [3]$$

with the lagged dependent variable on the right-hand side.

Anderson and Hsiao propose to transform the model into first-differences to eliminate the fixed effects μ_i:

$$\Delta ln y_{i,t} = (1+\beta)(ln\ y_{i,t-1} - ln\ y_{i,t-2}) + \delta(W_{i,t-1} - W_{i,t-2}) + (\eta_t - \eta_{t-1}) + (\varepsilon_{i,t} - \varepsilon_{i,t-1}) \qquad [4]$$

The authors show that OLS estimates of equation [4] would be inconsistent because $ln y_{i,t-1}$ is correlated with $\varepsilon_{i,t-1}$. However $ln\ y_{i,t-2}$ is correlated with $(ln\ y_{i,t-1} - ln\ y_{i,t-2})$ but not with $(\varepsilon_{i,t} - \varepsilon_{i,t-1})$, unless the error term exhibits autocorrelation. Thus, it is possible to get consistent estimates of equation [4] using $ln y_{i,t-2}$ and lagged differences such as $(ln y_{i,t-2} - ln y_{i,t-3})$ as valid instruments for $(ln\ y_{i,t-1} - ln\ y_{i,t-2})$.

The Anderson-Hsiao estimator is consistent but inefficient, while GMM estimators are efficient since all lags of the dependent variable can be used as instruments. However, instrument

proliferation is indeed a critical point in GMM estimators (Roodman 2009) because a large instrument set can overfit endogenous variables. The rule of thumb is to keep the number of instruments lower than the number of individual units. In our analysis we test for the robustness of our results by reducing the instrument count and collapsing the set of instruments.

As seen for the Anderson-Hsiao estimator, the lagged dependent variable in equation [4] is endogenous and any control variable $W_{i,t-1}$ which is not strictly exogenous could also become endogenous through the error term. On the other hand, lags going further back in time are not correlated with the disturbance term and can be used as instruments in GMM: lags *t-2* and earlier lags are valid instruments for endogenous variables, lags *t-1* can be used as valid instruments for strictly exogenous and predetermined variables.

One known weakness of the difference-GMM method is that it could have low efficiency if the series are persistent - as it is the case for output - and past values do not convey much information on future changes[33]. The system-GMM estimator was introduced to improve efficiency for *y* close to a random walk and it is now the conventional approach in the literature.

---

[33] See Arellano and Bover (1995), Blundell and Bond (1998), Roodman (2009) among others. See Hoeffler (2002) for an application to African growth performance, He and Sun (2013) for a convergence study across Chinese provinces.

Alma Mater Studiorum - Università di Bologna
DEPARTMENT OF ECONOMICS

Strada Maggiore 45
40125 Bologna - Italy
Tel. +39 051 2092604
Fax +39 051 2092664
http://www.dse.unibo.it