Session: Plenary Session 1: Measuring and Monitoring Economic Well-being in Times of Rapid Change
Time: MONDAY, AUGUST 25, MORNING
Session Organizer (s): Stephen Jenkins and John Micklewright
Discussant: Steven Stillman

Paper Prepared for the 30th General Conference of
The International Association for Research in Income and Wealth

Portoroz, Slovenia, August 24–30, 2008

Through the Looking-Glass: What is behind official data on inequality in Russia over 1992-2003?

by

Ruslan Yemtsov

For additional information please contact:

Ruslan Yemtsov
The World Bank
1818 H street Washington DC 20433
Email: ryemtsov@worldbank.org

Through the Looking-Glass:
What is behind official data on inequality in Russia over 1992-2003?

Ruslan Yemtsov[1]

August 20, 2008

Abstract

Russia represents an interesting case in which officially published data on income inequality are not directly derived from any specific survey. Rather they come from combining four elements: household survey of expenditures, household balances of income and expenditures, macroeconomic statistics tracing changes in the aggregate household money incomes, and a parametric model (lognormal) of the distribution, calibrated using data from these sources. This method is called to deal with problems which are not unique to Russia: low response rate in household surveys and under-reporting of incomes by survey respondents. The paper describes the evolution of the method over the period under study. It finds several breakpoints during 1992-2006, affecting comparability of inequality statistics over time. The paper shows that the modeling assumptions influence the level and trends of published inequality indices, and that these assumptions can be challenged. The raw survey data are adjusted by the Russian statistical office in line with these assumptions. Based on documentation of the official survey the paper proposes ways for restoring crude survey weights, which is necessary for comparisons to other survey-based inequality statistics. Relying on raw data from three surveys for the same year 2003: the official expenditure survey (HBS), large alternative survey (NOBUS) and a well-known panel survey (RLMS), the paper shows that the choices of welfare indices have at least as large an influence on measured inequality as modeling assumptions. Hence, improvement in the consistency of welfare indices can help to improve the accuracy of inequality data in Russia. Initiative of Rosstat to open access to its survey data will help to make a significant progress in inequality analysis, and will motivate improved data quality. The paper advocates for better sampling techniques, expanded definitions of welfare aggregates, expanded open access to survey data, and the use of other data sources (e.g. tax data) to provide new, more reliable evidence on the levels and changes of inequality in Russia.

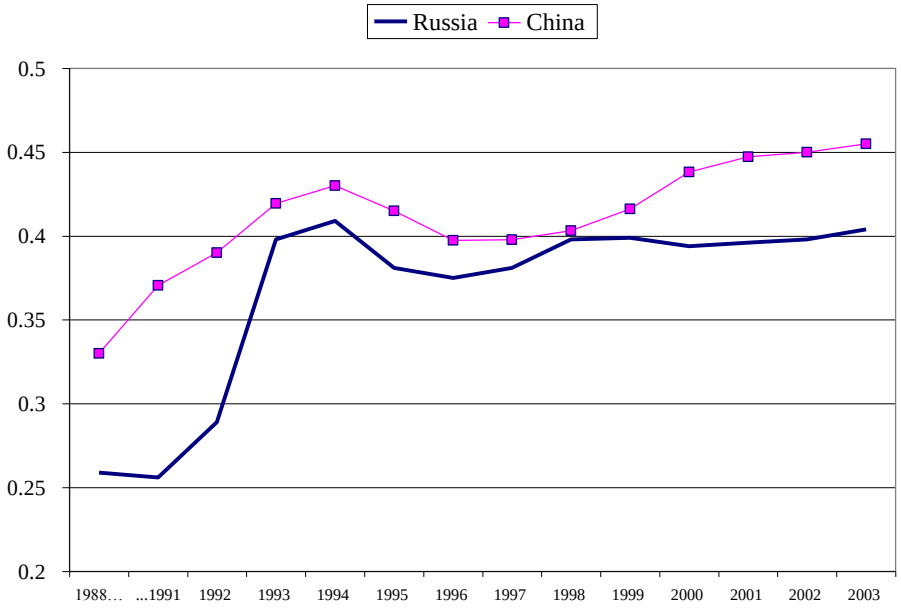
[1] This paper is prepared for The Conference of the **International Association for Research in Incomes and Wealth** (IARIW), August 24-30, 2008 for the session "Measuring and Monitoring Economic Well-Being in Times of Rapid Change". Some of its parts represent thoroughly revised and updated data section of an unpublished paper presented by the author at Cornell/LSE/WIDER Conference on Spatial Inequality and Development, on June 28-30, 2002. The author is indebted to Antony Shorrocks, who provided extensive comments on the earlier version, to Sergei Guriev and to Asad Alam for useful discussions and comments. Assistance provided by the staff of the Household Budget Survey Department of the Russian Federal Statistical Service (Rosstsat), is gratefully acknowledged, in particular the author thanks Elena Frolova and Tatyana Velikanova. The work was made possible by the Joint project by the Russian government, DFID, and the World Bank 'Enhancing Poverty Measurement, Monitoring, and Analysis of Poverty in Russia", 2002-2007. Through this work, with invaluable help form Olga Shabalina, project coordinator, the author had opportunities to discuss the issues presented in the paper with John Gibson and Graham Kalton. I am also grateful to Emil Tesliuc for providing materials on NOBUS. Victor Sulla helped the author to analyze household survey data from Rosstat. Zurab Sajaia provided crucial help with Rosstat and RLMS data at the earlier stage of the work. The paper does not represent any form of an official statement on the part of the World Bank.

Introduction

All scholars of inequality agree most of the time that former planned economies have suppressed the inequality. The transition to market economy was associated with a set of factors which had a potential to rapidly increase the variation in individual incomes and consumption. One does need any data to observe increased variability in individual material living standards. But by how much has inequality increased and where transition economies stand vis-à-vis other countries remains a source of debate.

Russia is often cited as an example of how fast and traumatic such an increase could be. These views are supported by a number of data sources, official and unofficial. It also supposedly shares this trend with other large economy which underwent a transition from a command to a market economy – China. The Figure 1 below based on official data from both countries on Gini for per capita incomes seems to corroborate the view of transition leading to sharp increase in inequality. By mid-2000 in both countries Gini indices were in excess of 0.4, a high level by any standard. The increase in Gini by 10-15 percentage point value is well above the threshold of 3 percentage points threshold proposed by Atkinson as "significant change" (Atkinson 2003).

Figure 1. Gini index for per capita incomes in China and Russia, 1988-2003



Notes: for both countries Gini is for nominal per capita incomes, without cost of life adjustments.
Sources: Russia, Rosstat, Social'no-ekonomicheskoe polozhenie Rossii, 2003 and 2007, China- China- data from Ravallion and Chen (2004), quoted in Mitra and Yemtsov (2007)

But at a closer inspection the story ceases to be simple and straightforward. A significant fraction of the total change occurred in early transition, and should be treated with caution, as there are well known biases regarding officially published data for the pre-transition and early transition years (Atkinson and Micklewright 1992, Alexeev and Gaddy 1993 for the case of Russia). The raw data on which these indices are based in both China and Russia

were not available to researchers, and the assessments based on available data vary depending on ten source (Annex table 2).

Methods of aggregation and estimation used or introduced by the Russian and Chinese statistical offices *during* this period are not neutral to the measured level of inequality. These methods will be a discussed at length in this paper.

Even though there are uncertainties regarding the magnitude of the change, its direction seems unmistakable (see Flemming and Micklewright 1999). Knowing the exact level of inequality is important not only for historical reason. This is because inequality may get in the way of economic growth, and because it clearly does have consequences for poverty, instability and anxiety. What it takes to measure the exact magnitude of inequality in a country experiencing rapid transformation like Russia, is the main question of this paper.

The first step to answer such a question is an understanding the nature of inequality and its measurement data. The paper omits well known facts about measures of inequality (see e.g. Cowell 1995) and instead focuses on various steps in data processing behind published inequality statistics in Russia. Data on inequality in Russia, as elsewhere, come from a survey - Household Budget Survey (HBS) conducted by the State Statistical Agency of Russia (Rosststat).[2] This data remained unavailable to researchers for a long time, but since 2007 the first full open access datasets of two years (2003 and 2004) have been posted on the web,[3] to be followed by 2005-2006 data release. But, as will be demonstrated, there are many step which intervene between the survey data and published series. The secondary objective of the author, who though his work at the World Bank has been exposed to this data and contributed to opening access to them, aims at introducing this source to a wider audience. All sources used in the paper are in public domain.

It is always important to complement the analysis of one source – official survey - with the analysis of alternative nationally representative surveys, with comparable instruments. For Russia there are two alternative sources for the same year, 2003: NOBUS[4] and RLMS.[5] Of

[2] The HBS is a continuous survey, initially introduced in 1952. It was substantially improved and revised in 1997. The HBS sample consists of 49,000 households drawn from the whole country, with data collected for each quarter. The household expenditure data measurement is among the most rigorous and detailed internationally, given that it relies on each household's maintaining a diary for 2 weeks and a log book for 11 weeks per quarter.

[3] http://www.micro-data.ru/obdh/obdhmicr/Main.htm (microdata)

[4] Russian abbreviation of National Survey of Household Welfare and Program Participation. Data have been collected by the Federal State Statistics Service (Rosstat) in collaboration with the World Bank. It was carried out in April and May 2003 and covered a sample of approximately 45,000 households with 118,000 individuals. NOBUS is a cross-sectional survey that uses a three-stage stratified sample design and is representative on the national, and on the regional level for 46 selected regions. NOBUS instruments resemble closely the questionnaires of the standard LSMS-type survey.

[5] Russian Longitudinal Monitoring Survey (RLMS) started in 1992, and is carried out now by the Institute of Sociology of the Russian Academy of Science with technical assistance from the University of North Carolina (UNC). It is a comprehensive panel survey (with refreshed sample each round) of all aspects of living, based on the first random nationally-representative sample of several thousand households across the Russian Federation (sample size decline somewhat from 5500 in 1992-1993 around 4700 households since 1994) 15 rounds of the survey are available (1992-2006). Typically the survey is conducted during the Fall.

particular importance is that the first survey was conducted by Rosstat, using a new sample and internationally consistent questionnaire. These two alternative data sources are publicly available,[6] and their comparative analysis is of interest. These data sources have already been used for inequality analysis using internationally accepted methodologies (latest study of inequality using RLMS is by Gorodnichenko et al 2008, NOBUS data have been thoroughly analyzed by Tesliuc and Ovcharova 2006), and the author will simply use the results obtained by other authors as benchmarks.

Previous paper by the author on a related subject (Yemtsov 2002) revealed inconsistencies in the official methodology of regional aggregation of inequality indices. The paper recalculated inequality indices and argued that recalculated published national inequality indices based on regional data and has shown that these recalculated series follow strikingly different trends compared to the officially published data. This was not to suggest that the "true" level or trend in inequality has been purposefully distorted by the statistical agency, but rather to emphasize the degree of uncertainty inherent in the official statistics. The paper did not discuss the underlying inequality data at the regional level, nor did it try to compare it to other data sources. This paper attempts to cover both fronts.

I am is going to show that the unique processing techniques are used by Rosstat to deal with the issue of non-response and under-reporting in its main survey address the right problems, but do not solve them. These issues are common to household data collection. Russian statistical office developed its own way to deal with the problem amidst rapidly changing economic and social environment. These steps consisted of moving away from survey data and in significantly altering survey results to correct for the biases. The apparent consistency of inequality indices achieved in this way comes at a price. First, the indices of inequality produced by Rosstat cannot be directly compared with similar statistics form other countries, and are not strictly comparable over time. By presenting lessons from Russian experience, the author hopes to contribute to efforts for understanding the nature and dynamics of inequality when data quality is problematic. Second, a through analysis of different published sources, publicly available sets of survey data, and a review of new on-going research efforts, allows me to show the ranges for basic measures of inequality in Russia for 2003. They remain large, motivating further research efforts.

The paper is organized in five sections. First section assembles key indices of inequality for Russia available from various data sources. It also presents brief history of HBS and basic framework of inequality measurement in Russia. Second section describes the key issues related to official procedures to estimate inequality indices based on HBS. The third section shows how the methodologies used affect the levels of inequality and discusses the likely effects over time. Fourth section puts the results in the context of the on-going studies and concludes.

[6] NOBUS: http://www.nobus.worldbank.org.ru; RLMS: http://www.cpc.unc.edu/projects/rlms/rlms_home.html

Section I. Inequality in Russia over Time and across Sources: Is there a Story?

There are several officially published and alternative (unofficial) data series of inequality for Russia. It is worth comparing them.

1.1 Russia during the transition: trends in inequality from official data sources.

The Table 1 below is based on official reports. It reproduces all available data on inequality changes in Russia over 1989-2006. [7] Even though money incomes may seem somewhat an inferior indicator of living standards especially in the presence of significant in-kind components of consumption, this is the only welfare index over 1992-2000. Moreover, this is the only indicator used to officially assess the extent of poverty. Thus, the choices of researchers who are willing to study the evolution of inequality relying on official data are very limited.

Table 1: Official Data for 1989-2006: Summary of Income Distribution Statistics for Russia (Shares of quintiles in per capita money incomes, percent, and the Gini index)

Quintiles of per capita incomes	*1989*	*1991*	*1992*	*1993*	*1994*	*1995*	*1996*	*1997*	*1998*	*1999*	*2000*	*2001*	*2002*	*2003*	*2004*	*2005*	*2006*
First quintile	9.8	9.4	6.0	5.8	5.3	5.5	6.2	6.0	6.1	6.1	6.1	5.7	5.7	5.5	5.4	5.4	5.2
Second quintile	14.9	14	11.6	11.1	10.2	10.2	10.7	10.2	10.4	10.5	10.6	10.4	10.4	10.3	10.1	10.2	9.9
Third quintile	18.8	17.9	17.6	16.7	15.2	15.0	15.2	14.8	14.8	14.8	14.9	15.4	15.4	15.3	15.1	15.2	15.0
Fourth quintile	23.8	22.8	26.5	24.8	23.0	22.4	21.5	21.6	21.1	20.8	21.2	22.8	22.7	22.7	22.7	22.7	22.6
Fifth quintile (richest)	32.7	35.9	38.3	41.6	46.3	46.9	46.4	47.4	47.6	47.8	47.2	45.7	45.8	46.2	46.7	46.5	47.3
Gini coefficient	**0.227**	**0.256**	**0.289**	**0.398**	**0.409**	**0.381**	**0.375**	**0.381**	**0.398**	**0.399**	**0.394**	**0.397**	**0.397**	**0.403**	**0.409**	**0.406**	**0.416**

[1)] Only 1989 is directly based on income survey , other data points are estimates. For 1992 and 1990 only total income data are available, rest of the series are for money incomes (series for total incomes stop in 1993). In 1992 total income exceeded money income by 11%, in 1993- by 14%, but their distribution (Gini index) was similar.
Source: Official publications of the Rosstat (1994, 2002, 2007, 2008).

Table 1 exhibits rapid increases in inequality in early transition years, followed by some moderation and gradual increase in the most recent period of economic growth, with official Gini reaching maximum on record in 2006.[8] There are some interesting features in the data: for example, the financial crisis of 1998 provoked only a slight increase in inequality. A really turbulent year of shock therapy of 1992 also seem to have had produced only a mild deterioration of inequality.

The first difficulty in interpreting the published series arises when they are compared with direct estimates of the same inequality measures from the underlying official survey: HBS. The recent public availability of this survey (HBS) for two years – 2003 and 2004 – offers for the first time an opportunity to do a quick check. It produces very surprising results: for 2003 Gini index for survey per capita incomes is equal of 0.449, not 0.403 officially

[7] In addition to quintile shares and Gini index Rosstat also reports "the coefficient of funds" (or the ratio of income shares of the top and bottom decile). Most importantly, no other commonly used inequality indices (general entropy class, or Atkinson measures) have never been systematically applied to the official sources.
[8] Preliminary data for 2007 suggest even higher level of Gini for per capita incomes – 0.422.

published; for 2004 – 0.457, much higher than 0.406 reported in the table; quintile shares also produces values which are statistically different from those reported in Table 1.

Moving backwards, and using the data reported for 1997-2002 in the Poverty Assessment (World Bank 2005), one can see that 2003 and 2004 are not exceptions. Indeed, for this period Gini estimated directly for survey per capita money income was ranging from 0.47 to 0.44 and declining, while Table 1 reports significantly lower values and, if anything, increasing over time.

This is a puzzling outcome which forces the user to inspect methodologies and definitions supplied by the Rosstat with the data. These documents make the point clear: reported inequality statistics do not directly come from a survey, they represent model estimates, calibrated with the survey. Parameters of the model are not reported, but it can be broadly described as a tool correcting for survey under-coverage of the rich. It leaves a user with a puzzle: how the survey –based indices, which presumably do not capture the rich can be lower than the "corrected" official figures reported in table 1? I will return to this puzzle later.

Inequality in Russia during the transition: comparing official data with alternative sources.

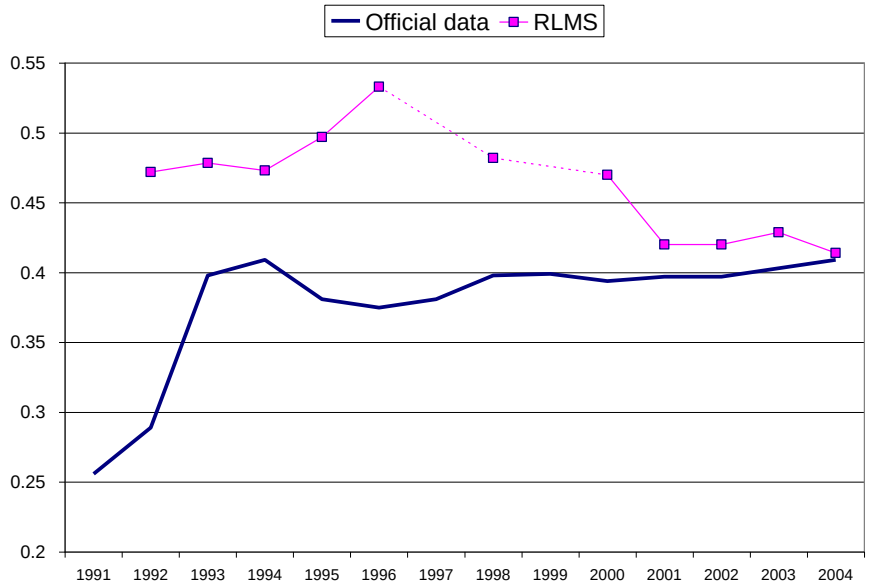
Given the fact that primary survey data on which official inequality measures are based were not available to researchers till very recently, it is not surprising that early empirical work was based on the **RLMS**. This survey which started in 1992 originally with participation of Rosstat (rounds 1-4) was the source of much of empirical research (Commander et al. 1999, Kislitsyna 2003, Gorodnichenko at al, 2008 etc.). De facto it was the only source of data on inequality in Russia which was open to researchers.

The particular significance of RLMS is that it can show changes in inequality over 1992-onwards, or most of the transition period, and its instruments allow comparability. RLMS has generally measured a wider range of household income and expenditure than the HBS. There is an acknowledged difference in the results from these two surveys, with the former suggesting, for example, a rapid rise in inequality in the mid-1990s, while the latter showed the opposite (Sheviakov and Kiruta, 2001, p. 7). There is also a well understood big discrepancy in means measured by the survey and coming from macroeconomic data (Popkin et al).

However, even acknowledging these differences in the means, researchers have not reached a consensus why estimates of trends and levels of inequality indices in RLMS and official data are so different. Figure 2, which puts time series of Gini index based on RLMS raw data and published series by Rosstat for the same variable side by side (per capita money incomes, no correction for regional differences in the cost of living) reveals striking differences. RLMS data shows high inequality already in its first round early in 1992, when official inequality measure was still low. RLMS based series peaked at the time when HBS –based series were on a declining trend – in a period preceding the 1998 financial crisis. It fell quite remarkably

since 1996 while official data show a slight increase. Both sources produce identical figure in 2004, but given differences for all other years, this is an inconsequential coincidence.

Figure 2. Gini index for per capita nominal money incomes in the official data and in the RLMS, 1992-2004.



Source: RLMS rounds 1-7 (1992-1996): Commander et al., rounds 8-13 – own estimates using primary data on nominal incomes. Official data - see Table 1.

Obviously, nominal money per capita incomes is an extremely poor indicator of living standards. RLMS produces much more accurate indices which mirror closely household well being and account for regional price differences. These series are widely used by researchers, instead of series reported on Figure 2. But this use of different, even superior definitions, leaves a question about what this discrepancy means. Should the RLMS be preferred over the official sources?

At a closer inspection, it is not obvious why RLMS based picture should be taken as more accurate when the inequality is described. First, RLMS is a small survey. It is particularly prone to data contamination issues and inequality indices based on it are very sensitive to high-leverage observations (see experiments and discussion in Cowell et al, 1999, and on the comparative assessment of noisiness of the RLMS data see Luttmer). Second, users of data which are forced to rely on RLMS because no other data were available often forget that the survey was not designed as a replacement of official statistics for broad distributional indices. It is a small panel survey (with sampling attrition issues) and its sampling strategy was to capture as much variance as possible (see sampling papers on the RLMS web site). This feature is bound to result in a higher observed inequality between survey respondents than a regular household survey. At the same time, the downward trend in inequality observed in RLMS may be attributed to some extent to "aging" of the panel and sample attrition (despite refreshments), as opposed to economic processes. Third, due to limited sampling size the survey does not capture to full extent the regional diversity of Russia and hence may not be fully accurate. Finally, for the same reason for which Rosstat is concerned with using the

raw survey data to assess the extent of inequality, one could be concerned with RLMS reported values. The gap between the survey mean and macroeconomic data in fact is persistently large – over 50% (Popkin et al).

NOBUS survey addressed one of the main problems of RLMS: small sampling size. Its design includes both consumption and income questions, and large sample based on the same sampling frame as HBS, and similar organization of field work make NOBUS-based directly computed inequality indices extremely informative for comparisons with official statistics. But it also has its key disadvantage: it is just one –time survey. Even though NOBUS was designed to be a periodic survey, it was conducted on a full scale only once (second quarter in 2003). Results of this comparison are reported in Table 2

Table 3. Gini index for per capita money incomes in HBS, RLMS, and NOBUS, 2003.

	HBS	**RLMS**	**NOBUS**
Money incomes, nominal	0.45	0.42	0.51
Money incomes, COL adjusted *	0.41	0.41	0.44

Data from HBS and NOBUS are for 2nd quarter of 2003, RLMS refers to the 4th quarter of the same year.
* COL- cost of living, NOBUS, RLMS and HBS are deflated using official regional subsistence minimum basket estimated in local prices
** with in-kind consumption of farm products, no imputed rent Source; Russia PA with authors' additions based on direct data estimates for NOBUS and published data for RLMS Source: own estimates.

HBS-based direct estimate of inequality in per capita money incomes without correcting for regional process differences (0.447) show a level of inequality below what is recorded in a parallel large NOBUS (0.512), and NOBUS estimate based on direct income questions is much higher than RLMS. There is more consistency across all three surveys in terms of Gini index for incomes once the adjustment for regional price differences is made. But these indices are still not identical to the published Gini.

Gini index for other welfare indices show even less coherence: per capita consumption expenditure inequality (as defined by Rosstat) is 0.36 in NOBUS as opposed to 0.40 in HBS and 0.45 in RLMS: a range of 10 percentage points of Gini represents a huge degree of uncertainty. There is a consensus among researchers that for theoretical and practical reasons survey based measures of consumption expenditures are more accurate than income-based in Russia (Tesliuc and Ovcharova, 2006). But the move to more "reliable" measures away from problematic income measures increases uncertainty in terms of how equal or unequal the distribution is!

1.2. Inequality in earnings.

Measuring dispersion of earnings involves smaller number of conventions than assessing household welfare. Nevertheless, since earnings represent a dominant share of household incomes the inequality of wages is the main contributing factor of levels and, most importantly, changes in inequality in transition (Mitra and Yemtsov, but because the concentration of earnings can also change rapidly there is no one-to one correspondence between inequality dynamics for wages and for total incomes). It can also rely on alternative

data sources- information on personal taxes, social contributions, or enterprise reports – and therefore is a useful check.

Table 3 reports main parameters of wage distribution reported by Rosstat. These are based on large periodic official survey of enterprises, covering over 200,000 employees. The pattern of increasing and then receding inequality is clearly pronounced. This evolution resembles the official series for the period of early to mid 1990s, but then RLMS –based trends in inequality for the late 1990s- early 2000s.

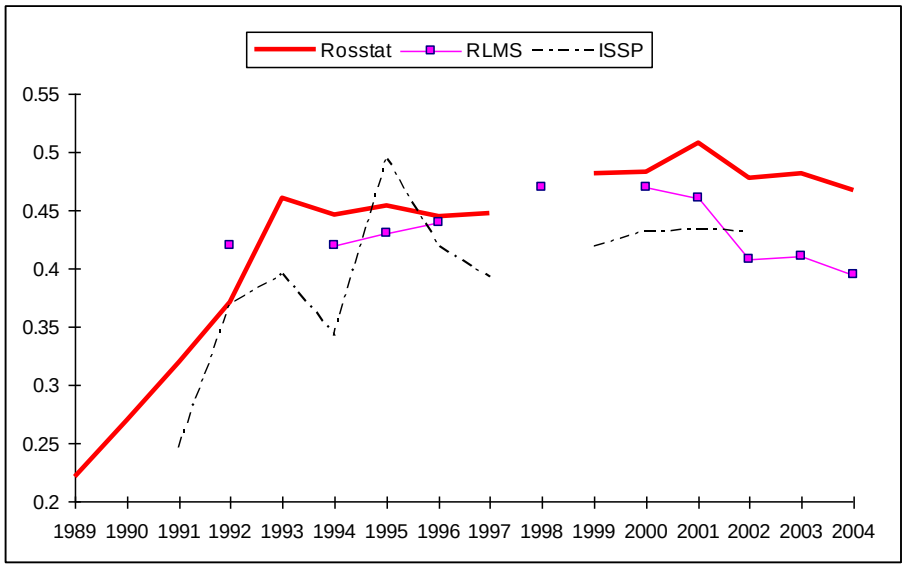
Table 3: Wage Distribution Statistics Published By Rosstat, 1990-2006
(Bases on enterprise data on wage bill, quintiles shares, percent, and the Gini index)

Quintiles of nominal wages	*1990*	*1993*	*1995*	*1999*	*2000*	*2001*	*2002*	*2003*	*2004*	*2005*	*2006*
First quintile (poorest)			3.9	3.5	3.4	3.0	3.7	3,6	3,9	4,1	4,1
Second quintile			8.6	8.1	8.0	7.5	8.4	8,2	8,6	8,9	8,8
Third quintile			14.2	13.5	13.4	12.7	13.6	13,5	13,9	14,1	14,1
Fourth quintile			22.6	22.0	21.8	20.7	21.3	21,7	21,6	21,8	21,7
Fifth quintile (richest)			50.7	52.9	53.4	56.1	53.0	53,0	52,0	51,1	51,3
Gini coefficient	0.269	0.461	0.471	0.482	0.483	0.508	0.477	0.481	0.467	0.456	0.459

Note: Wages are net of taxes for full-time employees, without annual bonuses and Source: Annual survey of wage bill, enterprises reporting to Rosstat (April or October), Official publications of the Rosstat (1994, 2002, 2007).

Figure 3 provides a further check by comparing trends in the Gini index for earning from official data and two alternative surveys on wages (RLMS and ISSP). Unlike in the case of inequality indices for per capita incomes, here the consistency across sources is clear. Wage data from whichever source follow an inverse U-shape trajectory.

Figure 3. Inequality (Gini index) for monthly wages



Sources: Russian Statistical office (Rosstat, 1999, 2001, 2003, 2006) "Trud i zaniatost' v Rossii"), ISSP: International Social Survey Program, Paternostro and Tiongson and RLMS – Russian Longitudinal Monitoring Survey, Gini index for monthly nominal wages for full time workers, calculated by Lukianova (2006) based on their contractual data. Note: Rosstat survey of wage distribution was not held in 1998.

Even though wage data are useful to confirm basic stylized facts regarding the degree of inequality, they are not sufficient to restore the full picture of distribution and its changes over time. Next section turns back to the household surveys to start disentangling the effects of data quality versus data processing and presentation in the official data.

1.3. Brief history of HBS and introduction to its concepts.

The key source of information on household welfare in Russia is the Household Budget Survey (HBS), conducted since 1952 as a regular survey by Goskomstat (later Rosstat). Its shortcomings are well known (see Atkinson and Micklewright, Chapter 3). This survey contained two clear sources of bias which led to under-estimation of inequality. The first bias can be traced back to the sampling frame, while the second source of bias lay with the under-measurement of unofficial economic activity.[9]

In the original survey design money incomes questions had to be collected in the survey alongside the expenditures. Goskomstat went to great lengths to "correct" reported information by survey respondents to keep an identity between what is reported at any given point in time in a survey and what comes from other ("macro") sources of information. For example, interviewers went to work places of a respondent to check the accuracy of reported wages, - and quota sampling based on the list of employees, plus high degree of command nature in the economy, would enable to do it.

Over time and with introduction of more radical market reforms it became impossible to trace the incomes reported by households to the "origins". In addition, the quota sampling based on employment in large and medium enterprises has become irrelevant source of information on the population at large.

Due to problems of "forcing" respondents to declare the true incomes, statistical agency conducted the last income survey in 1989, and it was not repeated since then (in 2006 there was a pilot experiment with income questionnaire). Direct questions about wage levels, revenue from business operations and property or gross household cash incomes[10] were excluded from the questionnaire and replaced with questions aimed at obtaining data that would characterize all the cash expenditures and savings of households. To obtain inequality and poverty statistics based on money income indicator Goskomstat and subsequently Rosstat started to rely on combining survey measures of expenditures and savings with non-survey sources of information on income referred as *macroeconomic money balances of*

[9] The sampling was concentrated among households employed in the state sector and in collective farms an under-represented private or cooperative activity, as well as pensioners. Somewhere between 1985 and 1995 Statistical agency was not able to revise its old sampling frame, and it became "better" over time through attrition, "aging" (thus increasing the number of retirees) and movement of respondents to new economic activities. But it remained completely at odds with the general practice of random sampling adopted in other countries all the way until the reform of 1997. No data were collected to reflect private business activities. As greater private activity was tolerated in the 1980s, under-estimation of inequality likely increased over time.

[10] The questionnaire nevertheless contains a question about the approximate level of total household cash income over 14 preset income intervals (up to 500 rubles, 500 to 1000 rubles, etc). The collected information is not used in practice.

population: (regionally disaggregated) data based on enterprise records, banking and retail trade statistics, and expert estimates.[11]

It is to note that in the FBS and HBS later all the way till 1995 the raw survey data stayed at the regional level and were processed to produce tables which then were communicated to the central office, for a model-based aggregations of income distribution. Out of necessity driven by computing power limitations and limited data transfer within the Goskomstat even direct survey observations have been processed as tables of summary statistics and aggregated using parametric forms. This necessity suggested the way to deal with the reporting issues in its surveys in the new period.

Thus, concluding this section it is important to note that:

1. For the entire period 1989-1996, which shows a dramatic increased in inequality, the sample quality, the data collection protocols and the data processing methods makes resulting data very dubious.
2. The lack of preserved primary raw survey data for 1980s-1995 mean that it will never be possible to restore "official survey-based" view of both the level and the change of inequality in early transition.

1.4. New design, 1997-2004, and recent changes.

Rosstat has responded to the new challenges in generally adequate way, by introducing major adjustments. The adjustment to the HBS design in 1997 was substantial and happened on many levels. It included change in the main concepts[12], questionnaire design[13], protocol of the survey to improve data quality, data management set up and sampling[14].

New random stratified sample has been drawn based on the primary records from Microcensus of 1994 (covering 2% of the population, listing all households for each census unit covered in all regions of Russia). Designing the sample was done in two stages, and at

[11] A number of income and expenditure concepts were used in Russia to monitor changes in the welfare of population. These concepts by and large have not remained stable over time and their changes reflected both changes in the economic environment and reforms in the data collection instruments. Despite these changes two welfare indices survived the modifications and are available on a regular basis: Money Incomes and Money Expenditures (and their per capita transformations). Table 1 in the annex gives details of expenditure definitions. Money incomes in Rosstat definition adopted since 1994 include both incomes received from formal sources (such as wages and salaries, social benefits, such as pensions and stipends, property income in the form of interests and dividends, insurance payments to households, loans to the population, and incomes from sales of any financial instruments) plus estimates of the incomes obtained outside officially registered economic activity (such as incomes of self-employed, supplementary income from sales of agricultural produce, private personal services etc.).

[12] From "family" to a standard definition of a household as primary unit.

[13] Each household in the survey is kept in a sample for one year, during which it is interviewed 5 times: each quarter plus one annual interview. The new survey contains a two-weeks diary of expenditures in cash and in-kind (conducted twice in each quarter) and a quarterly questionnaire of expenditures and in-kind consumption (prior to households kept a monthly diary). The annual interview includes collecting data about the households' living conditions, availability of consumer durables in the households, the turnover of livestock on private farms, the household members' educational level.

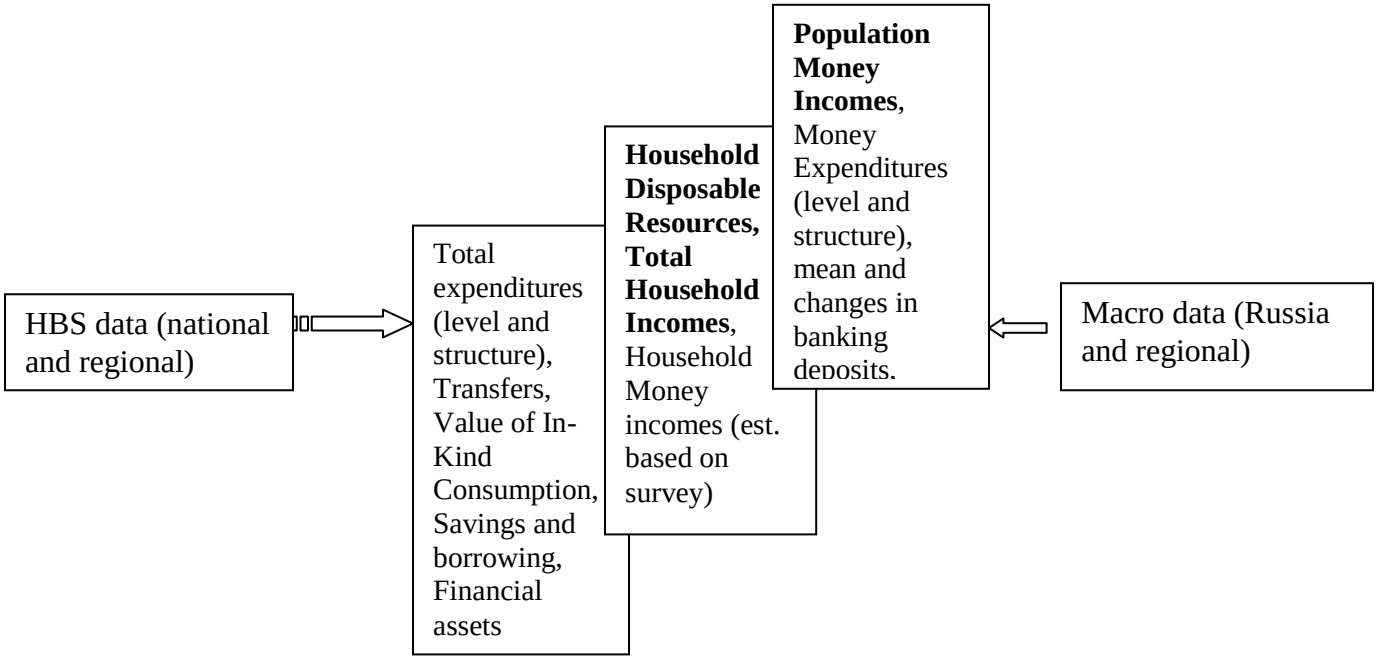
[14] Starting with 1995 all raw data started to be assembled by the central office in Moscow..

each of the stages random selection was made from stratified sample frames, designed as a combination of the following eight clustering variables: household size, type and ownership of dwelling, main income source, access to land and ethnicity, education, age and gender of household members. This stratification of the sampling frame is judged overly complex (Kalton and Chernysheva), but it is qualitatively superior to anything that was available prior to 1997. The revamped sample covers around 49 thousand families each quarter and is representative of each region of Russia for urban/rural subgroups.

While originally well conceived in principle the survey sample contained a lot of departures from good practice of random sampling, which later started to undermine the data quality. First, as documented by Kalton and Chernysheva, the sample provided extremely elaborated stratified long lists for "replacing" households who refused to be interviewed. These replacement were not counted in the statistics of refusals, and allowed the interviewers to fill in their quotas for field work. But they distorted the true picture of non-response, producing overly positive assessments, and most importantly, missing the opportunity to understand the process of survey compliance better. Second, the aim to produce quarterly representative statistics for originally 80+ regions of Russia resulted in a very inefficient sample (Gibson and Poduzov) with many small regions having to have relatively large sub-samples, and quickly exhausting their sampling frame. To continue data collection without refreshing the sample, it was decided to allow to interview the same household year after year. Third, in 2000 Rosstsat had to expand the survey coverage of regions (adding about a dozen of newly delineated regions) without additional funding, which resulted in cutting many outlying and remote PSUs and compromising the sample diversity.

A number of new welfare indicators were introduced based on household level data, which provided as increasingly compelling alternative to the money incomes. To summarize it a simple scheme below shows the key types of welfare indices depending on their sources (Figure 4)

Figure 4: Sources of main welfare indices in Russia, 1990-2008



Probably the most striking feature of this data, compared to other HBS elsewhere is that the Russian HBS does not collect information on income of household members. The *income* variables are **constructed** based on *expenditure* and flow of funds information (see Annex Table 1 for details). This indirectly estimated variable in a cornerstone of published inequality statistics. [15]

Note that the consumption expenditures, in-kind income and expenditures, total income and expenditures contain imputed value for some free-of charge social services (e.g. free transport) or in-kind transfers (since 1997), but they neither include any imputations for owner-occupied housing, nor any assessment of the flow-of-services from durables. These omissions may have impact on inequality (see Buckley and Gurenko). Many social services provided for free (i.e. health care and education) are also ignored.

The new wave of changes to produce an extended set of quality of life indicators occurred in 2005 – 2008. They were combined with a transition to the new sampling frame based on the new 2002 population Census. The new sample design is more transparent, without overly complicated stratifications and criteria for selection. But the replacement lists remain.

Thus, concluding this sub section it is important to note that:

1. The HBS survey was considerably revamped in 1997, and then again in 2005 and in 2008. These changes remain largely unnoticed by data users. Most importantly Rosstat main household survey up to 1996 was based on different principles of sample selection, different questionnaires and therefore are not strictly comparable to 1997-onwards data. There was a thorough revision of the questionnaires in 2005 and there is an imminent change in the sampling frame in 2008.
2. Key variable used for measuring inequality in Russia is not collected by the survey questionnaire, but is estimated indirectly using other variables collected in the survey.
3. The availability to researchers of raw survey data for 2003 and 2004 is a path-breaking change and hopefully will be continued with new years of data, leading to potential for re-estimating inequality indices in a more consistent way. But relying on ready-made expenditure variables for measuring inequality may be insufficient. The expenditure variable created by Rosstat contains a number of elements which make it a noisy indicator of living standards (expenditure on durables), but excludes some fairly standard imputations which enhance the comparability of data (e.g. imputations for owner-occupied housing).
4. Correction for regional or within-year price differences is not part of the data aggregation and computation of published inequality indices.

[15] E.g. Consumer expenditure, net of the value of the foodstuffs, given gratuitously by the household to others, together with the household's income in kind (estimated as the market value of consumed own production plus respondents' estimate of the value of benefits in-kind) make up the household's final consumption expenditures. Household disposable resources are represented by the sum of the household cash expenditures, withdrawal of savings and borrowings during the recall period, and income in kind.

Section 2. Main Issues behind Official Series of Data on Inequality in Russia.

Till very recently (early 2008) most official survey data remained off-limit to researchers it was impossible to see how the official inequality statistics are compiled and how close they depend on survey data. Now the public availability of data permits an assessment.

2.1 Calculation of distribution statistics.

There is a big discrepancy between survey mean and corresponding estimate based on macroeconomic data. The estimate household incomes in the survey are based on indirect questions, which presumably take into account sensitivities. But it does not solve the problem of comprehensive coverage. For example in 2003 HBS – dataset that will be used extensively in this paper, the mean per capita monthly income was 3179 rubles; data on wages, cash flows and banking statistics allowed Rosstat to estimate that in fact on average the monthly income for the same year was 5170 rubles. There is also an observation that this gap comes from richer households refusing to participate in the survey, and therefore distributional statistics based on a survey are systematically "biased".

To address this "bias" Rosstat adopted the procedure of restoring "true" unobserved parts of the distribution indirectly, by combining survey data, macroeconomic mean, and fitting the lognormal model of the distribution. The finalization of the methodology came in 1994 as an official methodology (even though the idea was discussed in the specialized statistical literature in Russia for some time (first publications are going back to 1986 - TSEMI). Once adopting it, Rosstat revised some previously published data (1985-2003) accordingly (Velikanova et al. 1996, Volkova et al., 1997).

Three assumptions are made for restoring "true" distribution if incomes:

- "Macro" estimate income is correct;
- The "true" distribution is invisible in the survey, because some social strata do not participate, but this invisible true distribution has a lognormal form;
- The survey mode, on the other hand, is not biased (lower tail of distribution is compliant, the rich are not)

The key hypothesis that was adopted by Goskomstat was that the unobserved "actual" distribution of welfare measure in the population in each region follows a log-normal distribution on the income variable, which has a well known form (see Aitchison and Brown):

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma x} e^{-\frac{\ln^2(x/m)}{2\sigma^2}} \quad (1)$$

which is a two-parametric functional form: σ, a measure of variance and the median, *m* (or other moment). Both parameters which have to be estimated (because of the fact that sample is biased) **cannot** be directly obtained from the survey data. This was a decision based on a large periodic special income survey module of HBS (170,000 households in the Russian federation) conducted by Goskomstat every 5 years, last time, however, conducted in a distant March, 1989.

The second condition is that the macro mean represents the "true" mean of the underlying distribution (not observed in the survey), mean from macro data *M* has to be equal to true mean, μ. [16] This gives therefore the first equation to estimate the parameters of lognormal distribution in the following way:

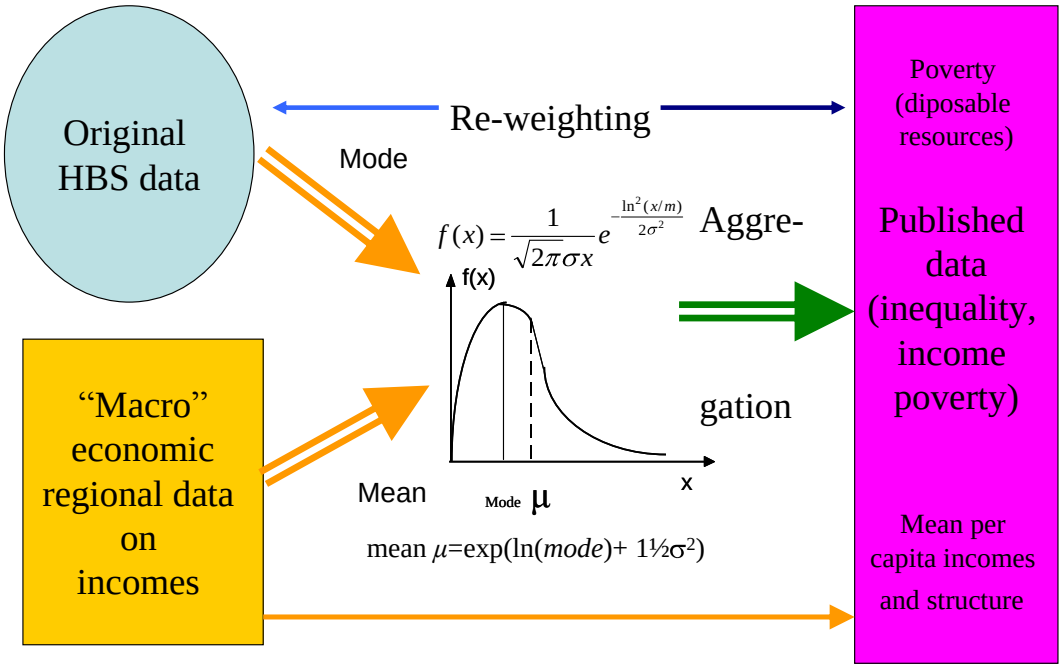
$$\text{mean } \mu = \exp(\ln(mode) + 1\tfrac{1}{2}\sigma^2) \quad (2)$$

From (2) one can easily obtain σ, a measure of variance. This is done at both the regional level and separately at the national level. Note that the Gini index can be directly estimated from the parameters of lognormal distribution once its form is established

$$\text{Gini} = 2\Phi(\sigma/2) - 1 = 2F(e^{m+\sigma^2/\sqrt{2}}) - 1 \quad (3)$$

where Φ(.) is the standard normal cumulative distribution - see Dikhanov for derivation. But in practice, the parameters of the lognormal distribution are used also to establish the boundaries of the deciles or ventiles which are then used for computing the Gini geometrically. This procedure is summarized on Figure 5.

Figure 5: Steps between survey and published official inequality stastics.



[16] Evidently, real survey data may contain multiples modes. To avoid the ambiguity, Rosstat is also using as a check a direct sample estimation of the variance of logarithms of per capita income (truncating 2.5 % at the top and the bottom of the sample). Sometime this estimate is used to "find" a correct mode. Moreover, the incomes are divided into standard intervals (between 10 and 20 groups), which also helps to select a unique value.

This is just a rough description of actual methodology (see Velikanova et al and Volkova et al) methodology, which calls for a use of a mixture of lognormal distributions for different social strata. But for most regions this description is an accurate summary of the estimation procedure. The adjustments for more complex assumptions affect only selected regions, where the survey and macro estimates are particularly far away, e.g. Moscow, resource rich regions, North Caucasus republics with large informal sector. Such adjustments consist of using the survey data to establish a proportion of the population which relied on 'entrepreneurial" income. This group is believed to have its own distribution (also lognormal). Since it is invisible in the survey, expert judgments are called to establish first, its mean and variance. Then the "visible" or survey part is calibrated using the "macro" data.

From 2002 onwards Rosstat started to publish distribution statistics by regions, and example is reported in Annex Table 3 (column 1) for 2002, and Annex Table 4 for 2003. Comparing published results with direct survey data estimates show that the published series on inequality are cannot be derived from the raw survey data.

Using the published data and survey raw data one can illustrate in detail how the methodology works for a simpler case of region where no special adjustment is done for the special stratas. We take a case of a region of Lipetskin central Russia, which is no different from 90% of regions. We take the parameters from the survey and published data and will trey to see whether it is possible to obtain with this information the published inequality index. I am also going to compare it to survey data.[17]

First, from macro data we know that the average money incomes in 2003 in this region was 3563 rubles per capita (Table annex 4). The survey mode is estimated at around 1750 (and it is based on diving the survey into 50-rubles intervals and picking the one with maximum frequency in the neighborhood suggested by the trimmed variance of log incomes as described in footnote 16). Solving a simple equation (2) we obtain "true " σ of 0.68846. Now we can restore the "true" distribution using the intervals of incomes (using (1)) and compute Gini [18]. Results are presented in Table 4. Since the actual parameters used by Rosstat are not known and it is not clear whether further adjustment to a simple two-paramater model are done, the fit between a simple estimate and the published series is good. The discrepancy in Gini can also come from a procedure used to compute Gini geometrically using points on the Lorenz curve rather than analytically.[19]

It is also remarkable how different is the published distribution form what is observed in the survey – the upper tail is much thicker in the estimated distribution. Despite that, the Gini index for this region is not much affected by the procedure.

---

[17] In that I follow the adjustment proposed in section 2.3 and used in section 3.2 for restoring the original sample weights.

[18] Using a useful approximation $$Gini = \frac{2}{\sqrt{\pi}} \int_0^{\sigma/2} e^{-x^2} dx \approx \frac{\sigma}{\sqrt{\pi}} \left(1 - \frac{\sigma^2}{12} + \frac{\sigma^4}{160} \right) \quad (4)$$

[19] It is interesting that using generalized Lorenz curve and software developed by Datt we can't obtain exactly the published Gini index using the intervals or deciles, the discrepancy is small with GC or with beta Lorenz curve, but it is always not zero.

Table 4: Illustration of the procedure to obtain inequality indices (Lipetsk oblast), 2003.

Monthly per capita incomes	Published	Estimated, mode=1750, σ =0.688	Survey original
Under 1000	6.5	6.7	18.0
1000.01-1500	11.5	11.4	18.4
1500.01-2000	13.1	13.0	18.9
2000.01-3000	22.9	22.7	22.0
3000.01-4000	15.9	15.8	11.5
4000.01-5000	10.3	10.3	5.3
5000.01-7000	10.8	10.9	4.4
Above 7000.01	9.0	9.3	1.7
Gini	0.371	0.374	0.367
Mean	3563.00	3563.00	2296.67

Source: own estimates and Annex Table 4 (see sources)

The inspection of Annex Table 4 tells that in most cases "true" Gini index as estimated by Rostat does not differ much from the survey based Gini, but the shape of the distribution apparently is. This is somewhat a paradoxical result, because the adjustment is meant to account for high-level incomes, not covered in the survey, and hence stretch the distribution. Instead, what is observed is hardly any different from direct survey estimate. As evident in Table Annex 4 this is a case for most regions, with notable exceptions. In Moscow, for example direct estimate of Gini based on the survey is 0.25, but the lognormal model (in this case a mixture of lognormal distributions for "visible" and "invisible" part with a high weight for invisible) produces a Gini of 0.615. Similar, albeit weaker effect is observed in Samara, Tumen, Irkutsk and St Petersburg.

It a first glance it is hard to explain a paradox of unchanged inequality as measured by Gini by the use of lognormal model. Indeed, among all possible parametric forms of the distribution analyzed by Bandourian et al (2002) it is lognormal model which other things being equal produces the highest measures of Gini. But at a closer inspection, this is not surprising. What researchers typically do is that they fit functional form empirical survey data, attempting to get as close match as possible. In the procedure adopted by Rosstat the logic is reverse: survey data are fit into the functional form, and the discrepancy between the fitted distribution and survey is not minimized. In fact, by taking the survey mode and adjusting the distribution to an exogenously given (and higher) mean, the procedure simply reflects the distance between the survey and the mean, but not the inequality. Moreover, the index obtained this way is not linked to survey data which makes it not decomposable and not interesting for research.

Upward adjustments to survey-measured inequality rely on expert judgments. It is hard to validate them. It is difficult to argue for the use of such indices to trace the changes over time. On the other hand a simple comparison to NOBUS data which contained direct income

questions reveals that the inequality estimated for exactly these regions is high (Gini over 0.5). Thus, even without modeling a well design survey can get over to the "invisible" part of the distribution in a very transparent way.

2.2 Definition of main aggregate for assessing the inequality.

HBS uses the following approach to derive an estimate of household cash incomes:

Cash expenditures
+ Net savings (increment in financial assets)

= Money income

However, the two components added together to form the income estimate have great differences in the likely reliability of their information. The cash expenditure estimates are derived from a very detailed and demanding data collection, with each household maintaining a diary for two weeks and a log book for several weeks per quarter. In contrast, the information on the increment in financial assets is obtained from extremely crude recall questions:

> "approximately what sum have you borrowed, taken as credit, or spent from your savings?"
> "what was the total sum you and your household members were able to save last month"

These recall questions are unlikely to accurately capture the change in household's financial assets (Gibson and Poduzov). It contaminates detailed expenditure data, when these data are added to the crude net savings estimates. Why given these uncertainties, "money income" is still used as the main welfare indicator? The answer is simple: since this is the only one for which "macroeconomic" series are available, and these are essential to "restore" the true distribution in the current methodologies. But how reliable are the macro data? The literature agrees that both survey data and macroeconomic data are prone to problems, and there is no presumption over which represent the "truth" (Ravallion, see also Deaton on India).

Table 5 Macro balance of incomes, 1992-2003, official data

	Total Money Incomes	Of which: Declared Salaries	Hidden salaries	Transfers	Entrepreneurial, property and mixed income etc.	*Memo: deflator and per capita average monthly money incomes*
			Structure			CPI Deflator92=1
1992	100%	70%	0%	14%	16%	1
1995	100%	38%	18%	13%	32%	119
1998	100%	45%	19%	14%	22%	258
2003	100%	39%	25%	14%	22%	925
			Real 1992 Rubles bln.			
1992	7,100	4,961	0	994	1,145	3,990
1995	7,657	2,894	1,345	1,005	2,413	4,323
1998	6,594	2,957	1,237	923	1,477	3,756
2003	9,605	3,755	2,382	1,354	2,094	5,471

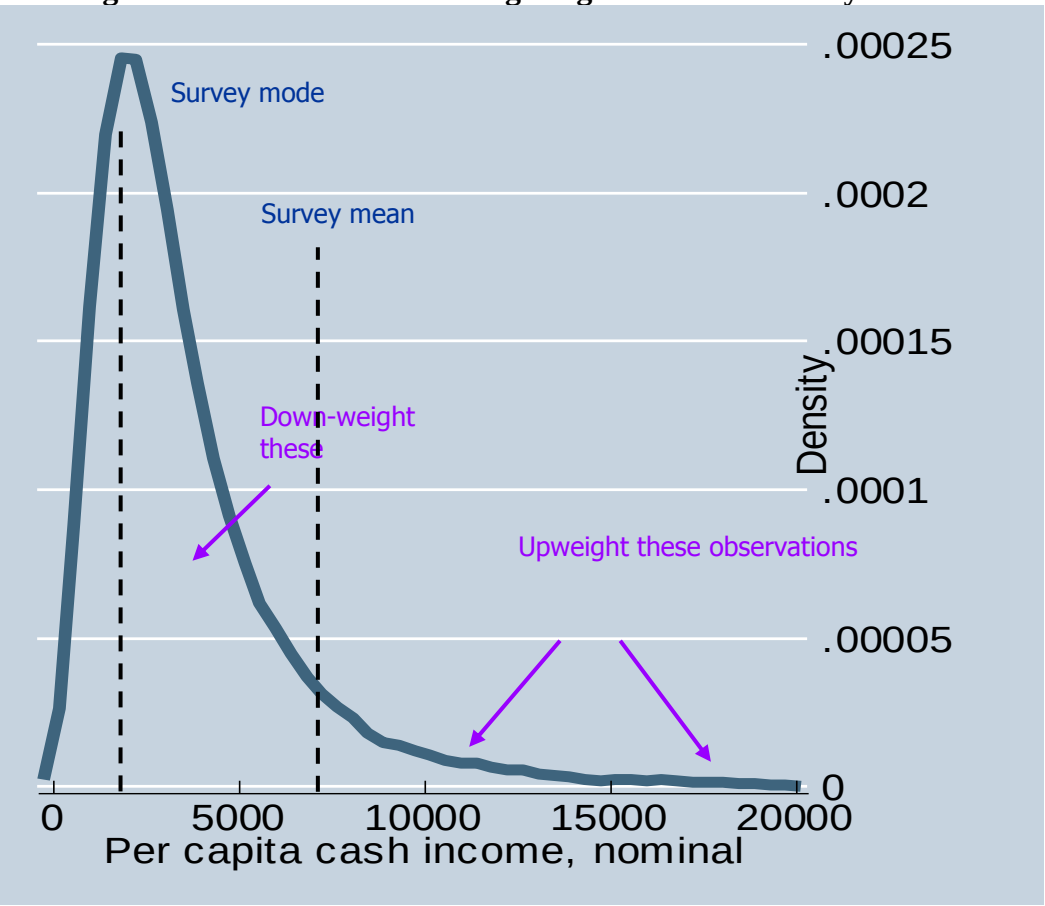
Sources: Rosstat 1997, 2003.

Table 5 shows that for Russia a significant part of incomes is accounted for by "hidden wages" –a product of itself of many assumptions and expert judgments. It is even less reliable at the regional level. Since methodologies for constructing money balances are not widely known, and underlining data are not in a public domain the decision could benefit from a dialogue between Rosstat and researchers around this tool.

2.3 Re-weighting of survey data.

The effect of adjusting for "missing" data does not stop at the level of "modeling" which is a source of published inequality and official poverty data. It also affects back the survey data through the adjustment of weights, or re-weighting to correct for non-response. As any complex multi-stage sampling survey, HBS is characterized by an unequal probability for households to be selected in the sample. The inverse of this probability or sampling weight is further corrected by Rosstat in a two stage procedure. **First** step "sample weighting" is standard part of any survey with unequal probability of households to be covered by the survey. There is nothing unusual about this stage. Most surveys in the world contain these sampling weights corrected for actual differences in response rates. The basic idea of really unusual **second step** of re-weighting is to "capture" the household decision to participate in the survey. A simple diagram can explain the major feature (Figure 6).

Figure 6: Features of the re-weighting Procedure Used by Rosstat

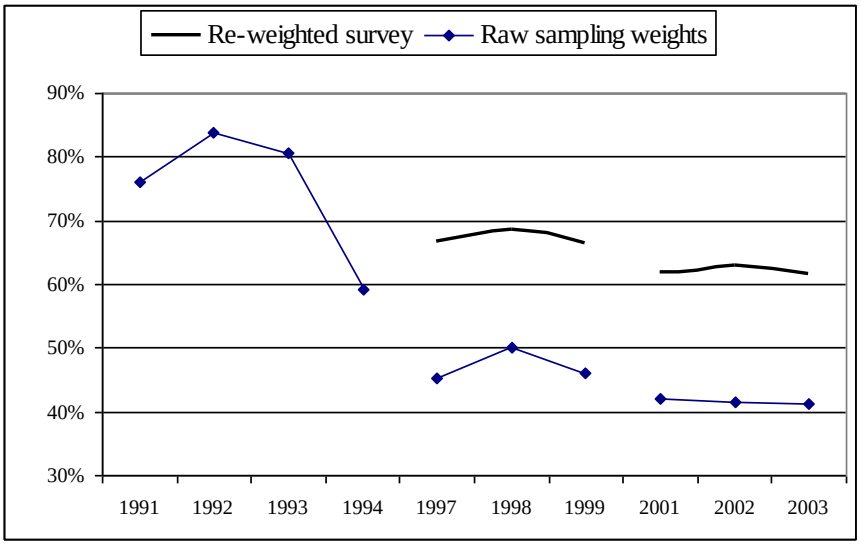


Source: Gibson and Poduzov based on Rosstat

The survey distribution of estimated per capita cash income often has a central tendency that lies a long way below the regional mean of macroeconomic variables that should reflect the same income processes. Assuming that the cause of this discrepancy is under-representation of the rich households, Rosstat raises the weights attached to the richest households in the sample, and lowers the weights attached to the poorest households. Importantly, the weight used to adjust for lower survey income is combined with the original sampling weight, so there is just a single weight attached to each household. This procedure is done for each quarter, region, territory, urban/rural location and standardized household types. Thus for each of such clusters there are **three values** of weights.[20]

This "correction of bias" is aimed at addressing the issue of discrepancy between survey and national accounts. Figure 7 shows how the resulting mean corresponds to the national mean income. The adjustment to survey weights which does produce very substantial changes in both the average welfare indices and measures of inequality is not eliminating the gap with the national accounts which are supposedly "accurate". This is due to a very crude nature of the adjustment practiced by Rosstat (just 3 intervals).

Figure 7: Russia: ratio of per capita money monthly incomes in the HBS (with raw and adjusted weights) to per capita incomes from macroeconomic data, 1991-2003 (%)



Source: Roskomstat statistical bulletins, No. 1,7,11, 1999.; data on 1991-94 from Voprosy statiskiki, 7/1995, and for 2001-2003 from Sotisial'no ekonomicheskoe poloizhenie Rossii (2006) and authors direct estimates.
Note: it is assumed that survey means published prior to 1995 had no adjustment to sampling weights (step two re-weighting).

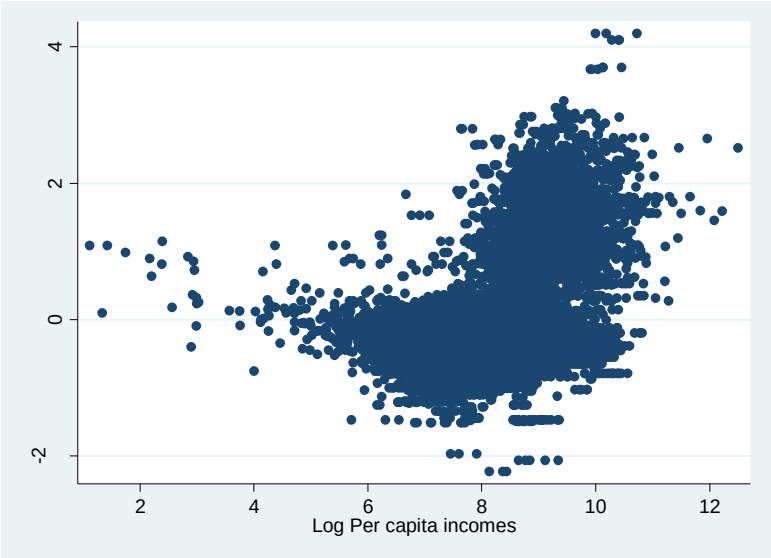
As weights are not uniform (they are not just simply scaling up the entire distribution), it is clear that the effect of re weighting will depend on the empirical form of data distribution. I illustrate this with an example of 2003 HBS (for Russia as a whole and for Moscow) on the Figure 8.

[20] In 2003 Rosstat modified slightly the procedure, now more than 3 weights are allowed to reflect separate weighting for the households which do have entrepreneurs among their members to the distribution, but still 95% of households fall into one of the three weights category.

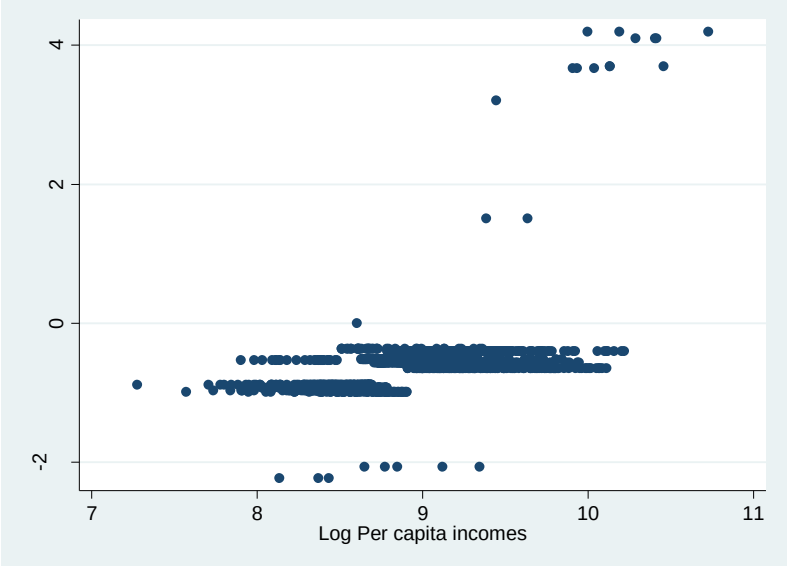
On this graph (log scale for x axis), each dot shows the degree to which are household sampling weight is adjusted. The graph shows a characteristic U-shape for weighting variable (which one may interpret as a reflection of the survey not capturing adequately both the rich and the very poor). Weights which have to be assigned to the relatively well-off households are indeed significantly above 1, while households around the median and mode are significantly below 1 (zero in log scale). The number of observations on the left side is very small, and once proper Goskomstat income brackets are used to define weights, the downward sloping portion almost disappears, and weights become represented by a simple upward-sloping line

Figure 8. Re-weighting of data in HBS 2003, 2nd quarter

Russia



Moscow



Source: own estimates

The effect of such re-weighting on inequality is obvious. It is illustrated for each region by Annex Table 4, which reports values of Gini for both original weights and re-weighted data. Not surprisingly, recalculating Gini index for "original" sampling weights one obtains a value with is 3-4 points lower. Interestingly, the direct survey estimate exceeds the published inequality indices (which hare based on model) for most of the time. Thus, the survey based statistics are affected by the matching to the macro mean in a very particular way. The published inequality indices are not consistent with direct survey estimates event though they are processed based on the same model.

2.4 No correction for regional and for within- year price indices.

The main purpose of the measurement of welfare is to be able to compare households to each other. The two key adjustments that must be made to allow valid comparisons of households are for spatial price differences and for household composition. These two adjustments enable a proper ranking of households and individuals to be made.[21] There is not official practice for both (see Sheviakov and Kiruta). This factor is bound to overestimate inequality for high-inflation years.

RLMS introduced the use of poverty line as regional deflators. In Russia there is an official national poverty line based on a subsistence minimum defined as basket of goods (referred to as the Minimum Subsistence Income or MSI) in use since 1992[22]. The data on minimal costs-of-living or poverty lines (which Rosstat calculates monthly, quarterly and yearly for each region and at the national level) represent average prices per region and do not take into account prices differences between urban in rural areas within regions.

Issue of equivalence scale. The official poverty line contains implicit equivalence scales to different age and gender groups (children under 18, male 18-59, female 18-54, male 60 and above, and female 55 and above), all based on nutritional needs and ignoring completely any economies of scale. On average application of such equivalence scale yields results similar to per capita calculations. There has been a debate over the implications and legitimacy of such an approach (see World Bank, 2005a). However, officially available data are based on per capita scale and we are going to use it throughout the paper, keeping in mind that such assumption yields generally upper-bound levels of inequality when compared to the inequality measured with some conventional equivalence scales.

[21] Note that, in the case where data are collected over a long period of time, it would also be necessary to adjust for changes in prices over time.

[22] The basis for establishing poverty basket or MSI was a President's Decree of March, 2, 1992 N 210; this decree allowed preparation of the official guidelines for regionally specific MSI baskets by the Ministry of Labor published on November, 10, 1992. These guidelines remained unchanged until the first quarter of 2000, when new methodology was introduced. This methodology itself takes its origin in the Federal Law of October, 24, 1997 № 134 and the corresponding guidelines issued by the Ministry of Labor and Goskomstat on April, 28 2000 (№ 36/34). The most profound aspect of this change is that the new Guidelines specify only the national level MSI basket; all regional methodologies had to be developed based on the Federal law framework by the regional parliaments.

2.5 Conclusions.

There are four issues associated with the existing official series on inequality: (i) inequality indices are calculated for indicators which are not collected by the survey in a way that is not standard; moreover the set of published indices is extremely constrained, (ii) the main welfare variable chosen for monitoring inequality is the least reliable survey variable , (iii) there is a contamination of survey data with the macroeconomic statistics by re-weighting procedures; (iv) there is no proper accounting for price differences across regions and within the year. These issues also represent possible areas for improving the current practices. Next section analyses their influence on inequality statistics.

Section 3. How Methodological Issues Affect the Accuracy of Published Inequality Indices?

Using raw HBS data for 2003, ways to adjust for its deficiencies described above, and comparisons to other surveys (RLMS and NOBUS), this section attempts to estimate the effects of methodological choices of Rosstat on published indices of inequality.

3.1 Definition of welfare aggregates and measured level of inequality.

The fist confusing aspect of data with which each user has to deal with are inconsistencies between "macro" or model level estimates and survey based assessments, even when they correspond to seemingly identical variables. Table 6 reveals these differences using both the "official" survey and RLMS, and given the deficiencies of this measure described above- for other welfare measures.

Table 6. Russia: Gini indices per capita for different definition of welfare: Official series versus direct estimates based on HBS and RLMS

	1997	1998	1999	2000	2001	2002	2003
Money incomes: official series	0.38	0.40	0.40	0.39	0.40	0.40	0.40
HBS (direct estimate from survey data)							
Money income, nominal	0.47	0.46	0.46	0.45	0.46	0.45	0.45
Money incomes, COL adjusted*	0.45	0.43	0.42	0.41	0.42	0.40	0.41
Cash Expenditures, nominal	0.45	0.46	0.45	0.44	0.45	0.45	0.45
Cash Expenditures, COL adjusted *	0.43	0.42	0.41	0.41	0.41	0.41	0.41
Consumption, nominal	0.37	0.39	0.37	0.36	0.37	0.37	0.41
Consumption, COL adjusted *	0.35	0.36	0.34	0.33	0.33	0.33	0.37
RLMS							
Total incomes, nominal**		0.48		0.47	0.42	0.42	0.42
Total incomes, COL adjusted **,*		0.45		0.41	0.40	0.40	0.41
Total expenditures, nominal		0.48		0.46	0.45	0.44	0.45
Total expenditures, COL adjusted *		0.47		0.45	0.44	0.42	0.44
Consumption, nominal		0.42		0.40	0.39	0.39	0.40
Consumption, COL adjusted *		0.41		0.39	0.38	0.37	0.39

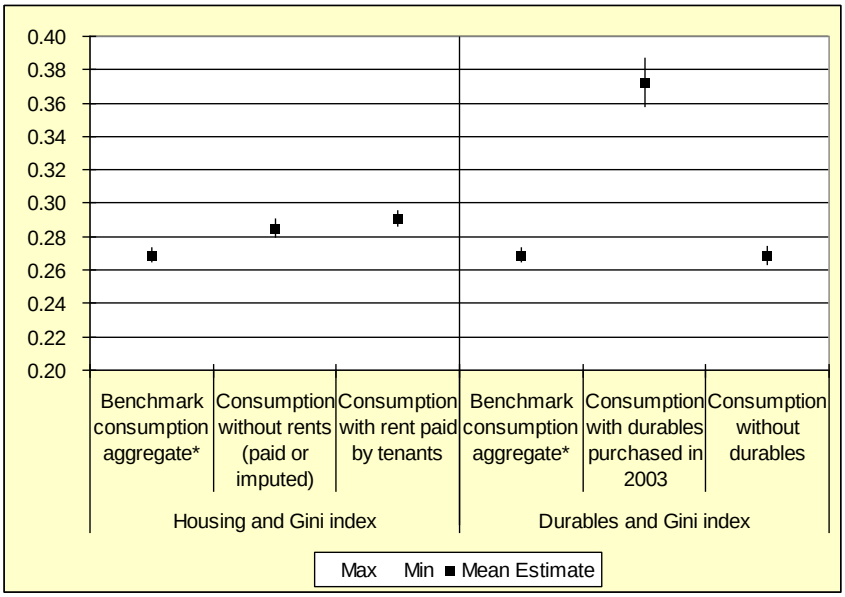
* deflated using regional subsistence minimum basket estimated in local prices ** with in-kind consumption of farm products, no imputed rent Source; Russia PA with authors' additions based on direct data estimates for NOBUS and published data for RLMS

The discrepancy between survey-based estimates and "modeled" inequality statistics does not follow a clear pattern. Both HBS and RLMS measures, once consumption is adjusted for price differences, shows a decline of inequality over time, unlike official series which show no change.

The set of Gini indices in Table 6 clearly indicates that the choice of welfare measures has similar effects on the level of inequality and its trends for both data sources. Constructing comprehensive consumption aggregate using NOBUS data further illustrates this point.

Tesliuc and Ovcharova by using to full extent the potential of NOBUS for doing all necessary adjustments for measuring real household consumption arrive at very surprising results, which qualify the assessment of inequality in Russia as extremely high. The just of their results is reproduced on a figure 9 below.

Figure 9 Impact of the treatment of durable goods and owner occupied housing imputations on inequality (Gini index)



* With imputed rents for owner occupied housing, imputed flow of services from durables and regional price adjustment.
Source: Tesliuc and Ovcharova

Only by including the durable purchases into the welfare aggregate is NOBUS data reproducing the level of Gini close to 0.4 so persistent in the official series. Once welfare index is constructed carefully in a way that has become common across countries, main "stylized" fact on inequality in Russia is destroyed: with inequality as measured by Gini of about 0.27 it is hard to argue that Russia is a deeply unequal country. This is in sharp contrast with a Gini index of 0.5 for money incomes for the same survey. Annex Table 4 shows why: the money incomes for a majority of regions are poorly reported. Mean incomes are well below macroeconomic aggregates, and well below what is reported by the same households as their consumption. Hence, survey-based income inequality is so high because it contains a lot of noise.

It is of particular importance that NOBUS being "outside" of official statistical sources was no subject to such adjustment of weights. These original undistorted weights and a careful sampling strategy make the estimate of consumption inequality based on NOBUS particularly plausible. But this single estimate shows that there are serious data problems with tracing inequality in Russia.

3.2 Weighting scheme: assessment and simulations.

There are several assumptions going into the official estimation procedures: assumption that the "true" distribution is lognormal, assumption about the equality of modes, and assumption about the true mean being represented by macro data.

Assumption 1: Is mean from "macro" data correct? As discussed a lot of guessing goes into estimating "hidden wages". In addition regional means estimates depend on cash flow statistics which do have gaps in the flow of funds across regions. Methods for estimating these can improve over time. Under the current methodology this will require revising all series for inequality as well.

Assumption 2: lognormal is not the best functional form approximating real distributions. The work by Bandourian and others (2002) used a set of LIS surveys (for 23 countries, including Russia, and various years – a total of 82 datasets) to assess how closely actual data correspond to parametric forms for 11 multi-parameter functions used in the literature to approximate the distribution. Log normal performs worst of all in practically all datasets.

Actual methods of implementing re-weighting on survey data remain crude. The identity between surveys and macro means is not achieved (and should it be?) at a price of changing sampling weights irreversibly. Increased variance of weights overstates inequality for any survey-based data – by 3-4 points of Gini.

The Russian statistical literature so far has concentrated on debating those assumptions and proposing alternatives. Rosstat estimations were criticized as sensitive to the sample bias, because the sample **mode** estimation of distribution density were found to be incorrect, or difficult to identify unambiguously. At the same time, in Russian scientific literature arose a discussion on income distribution form, and opinions that income distribution has a non-parametric form had been stated (see Sheviakov and Kiruta, this view is not shared by Aivazian who proposes a model of mixture of lognormal distributions). The aim of this section is not to discuss the validity of these assumptions, but try to see how they affect the results.

Aivazian and Kolesnikov, Sheviakov and Kiruta, Kolmakov and Velikanova proposed alternative measures to eliminate the gap between survey and macro means. In addition they produce striking discrepancies in estimations of inequality in per capita income, based on the same Goskomstat initial data, but obtained with the use of different weighting techniques.

For example, I. Kolmakov and T. Velikanova[23] (unpublished report to Goskomstat, 1995) estimated the ratio between incomes of 10% richest and 10% poorest population in 1994 as 27.1, and V. Ivanov and A. Suvorov (1997) estimated[24] that ratio in 1996 as 65, while corresponding official Goskomstat estimates were as 15.4 and 13. Aivazian and Kolenikov

[23] Kolmakov and T. Velikanova used as a distribution model a mixture of two lognormal distributions, the first for distribution of low and "low middle" incomes and the second for distribution of "top middle" and high incomes, estimating the mode of the first distribution as the sample mode.

[24] V. Ivanov and A. Suvorov used lognormal distribution model with the same estimation of its mode

apply their procedure to RLMS and find a Gini for Russia moving from 0.40 to 0.61. These illustrations show how profoundly the results are affected by the re-weighting, but they are not specific enough to tell whether this is leading to "correct" estimates of inequality.

Re-weighting for non-response has a long history in household survey methods. CALMAR procedure (Deville, LeGuenec) or the work by Korinek et al. developed models for survey compliance to adjust the survey data. However, so far the features of the procedure proposed by the researchers mentioned above amplify then problems in Rostat methodology.

First, this weight procedure creates extraordinary variation in the final weights applied to each household. For example, in the last quarter of year 2000 Gibson and Poduzov show that the five households with the largest sample weights have the same effect on calculated statistics as did approximately 5,400 households with the smallest weights. These extreme weights suggest both statistical and budgetary inefficiency. On the statistical side, it makes the calculated statistics more sensitive to the values taken by a very few observations than would be expected for a sample of 49,000.[25] From the budgetary side, all of the expense of collecting data from 5,400 households with the very low weights is hardly reflected at all in the final statistical calculations.

Table 7. Household Weights in Russian HBS, NOBUS and some international comparisons

	National Sample			Capital City Sample		
	Largest weight	Smallest weight	Ratio	Largest weight	Smallest weight	Ratio
Russia						
HBS, 2000, 4th Q	145,800	3	45,700:1	145,800	671	217:1
HBS, 2003, 2nd Q	154,862	12	12,905:1	147,000	28	5250:1
NOBUS, 2003	14,065	15	937:1	7,134	804	9:1
Other countries						
Poland, HBS 2002	588	302	1.9:1	475	316	1.5:1
Indonesia, SUSENAS	5172	44	118:1	484	173	3:1
Vietnam, LSMS	4723	815	6:1	n.a.	n.a.	n.a.
Egypt, HICES 2005	917	118	8:1	492	263	1.9:1

Source: Gibson and Poduzov for 2000 and Indonesia and Vietnam; 2003, NOBUS and other data- author's estimates

To a certain extent, the pattern of the weights reflects the unequal selection probabilities into the sample for households from different regions of Russia, which itself is driven by the need to have sufficiently large samples at the region level. So, for example, Moscow provides only 2.8 percent of the households in the sample, but about 6 percent of the national population, because regions with smaller populations are over-weighted in the sample. Thus the weights are naturally higher for households in Moscow, which contributes to the contrast between largest and smallest weights. But this is a similar situation in other countries, yet the weights are nowhere near as extreme as in the HBS. Moreover, this sampling effect can be removed by just examining the pattern of weights within the capital city. In the surveys from other countries, the ratio of largest to smallest weights within the capital city is never higher

[25] For example, removing the five households with the largest weights can reduce the survey estimate of mean per capita income by almost five percent.

than 10:1, whereas in Moscow it is about 220:1. Moreover, while there is almost no correlation between the household weight and per capita income in Jakarta (r=0.03), or Cairo (r=0.11), or Warsaw (r=-0.09), in Moscow there is a very strong correlation (r=0.47) because of the up-weighting of richer households.

These extreme weights in the HBS affect *all* of the potential welfare indicators because the single weight for each household is applied to all of the variables available from the survey. Because these weights are a response to the comparison of survey cash incomes, which is a derived measure, with macro data, any error in the method used to obtain the cash income estimates will affect all of the welfare indicators.

The second feature of the weighting used by Goskomstat is that it is based on the assumption that the discrepancy between macro variables and the survey measure of cash income reflects *unit non-response*, whereby rich households do not participate in the survey. However, as argued by Gibson and Poduzov, and later by Aivazian, another plausible cause of the discrepancy is that survey cash income is understated because of *item non-response*, whereby households refuse to answer income-related questions but they may answer expenditure-related questions quite cooperatively. There is a statistical evidence which supports this view. For example in 2002 For example, the probability of including in a Russian sample households with monthly average per capita income exceeding 60,000 rubles (about $2,200) is virtually zero, while the real share of such households in Russia ranges in different estimates from 2 to 3 percent. This is not a gap which would explain 50% gap between survey and macro mean.

It is important to note that the changes in the method of deriving parameters of lognormal distribution from survey data which occurred after 1994 do have effects on the relationship between the survey primary data and obtained inequality estimates. From the procedures that tend usually to "underestimate" inequality compared to primary data, Rosstat moved to a procedure which tends to "overestimate" it. As far as I am informed, there was no effort in revising the series on inequality backward with the new methodology. Therefore, all comparisons of inequality based on published data to pre-1996 (and for sure pre-1994) are not robust.

How large is the effect of weights on inequality indices? As most countries in the world do not practice adjustments to weights to account for survey non-response, international comparison of inequality which include Russian official survey data should rely on restored weights. Table 4 in the Annex shows that the effect does not exceed 4 points of Gini – much less than an affect of combined account for regional price differences and use of consumption as a robust welfare measures as opposed to problematic income index. These issues are secondary compared to the primary issue of defining welfare and collecting accurate data on non-response.

Section 4. What Do We Know about Inequality in Russia? Conclusions

Comparisons of inequality over time and across countries is a notoriously difficult task. But consistent use of definitions and robustness checks do help to establish facts, and rely on large national surveys for such comparisons (e.g. Cowell, Litchfield and Mercader-Prats).

Most surveys in the world face a compliance issue. Taking this issue very seriously, the Russian statistical agency proposed a solution which consisted on combining survey data with other sources. It is basing all its published series on inequality on this solution. The paper identifies six issues associated with the data on inequality coming from the official Russian sources:

(i) inequality indices are only tangentially based on data collected by the survey and are calculated in a way that is not standard;
(ii) the adjustments to survey data are based on assumptions which are hard to defend;
(iii) the set of published indices is extremely constrained;
(iv) the main welfare variable chosen for monitoring inequality is the least reliable survey variable ,
(v) there is an artificial contamination of survey data by re-weighting procedures;
(vi) there is no proper accounting for price differences across regions and within the year.

There are other ways to assess the "true" extent of inequality by non-distorting means. For example, Guriev and Rachinsky use grouped personal income tax records for Moscow city to assess the full extent of the inequality. They show that the NOBUS survey for Moscow city has missed the upper tail of the distribution. They argue why tax data (after the flat tax reform) can be relied upon to assess the distribution. By combining survey (NOBUS) and tax data they produce a synthetic distribution which is characterized by Gini of 0.625. Even though it resembles the Gini for Moscow estimated by Rosstat for the same year (0.605), this procedure is based on transparent aggregation of data and can be validated. Such efforts are increasingly common across countries in the world, and doing this for Russia as a whole over time will provide a major step forward in assessing the true extent of inequality.

Ersado uses simple Pareto distribution for the upper tail missed by the survey in Azerbaijan and relying good reporting of response rates, obtaining Gini of over 0.5 as opposed to less than 0.2 measured in the survey. Korinek et al. proposed for models which take into consideration survey response behaviour and can be calibrated with data from outside surveys. There are other non-intrusive ways to dealing with problems which all depend on the good survey practices. These all researchers share the view that it is simply impossible to ask from a survey to provide ultimate answers to the question on inequality levels and changes. The large literature on fitting the distribution to parametric forms (se review in Bandourian et al)) has stayed away from suggesting the "correct" models.

That are the lessons from this review?

1. Complex procedure do not solve data quality issues. Better work in the field, better data documentation, better sampling cannot be substituted by data processing.
2. Definition of welfare matters for the measured level of inequality more than processing of data.

The paper argues that the solution to the root problem, which motivates Rosstat to elaborate its methods, consists in collecting better survey data, greater transparency at all stages of survey cycle and use of other sources. The development of tools to "repair" deficient surveys can only be of limited use. In particular, matching survey indicators to macroeconomic data is not the way to overcome survey deficiencies. Use of specific lognormal distribution model for assessing level and changes in inequality may distort the actual signal. The use of consumption for measuring inequality consistently over time in Russia is possible, and will be superior to reliance on incomes. And finally, the set of indices for tracing inequality in the official publications can be usefully extended.

Concluding this discussion of available evidence on inequality provided by the official and independent sources, one clearly sees that inequality did grow rapidly in the 1990s in Russia. There are a variety of ways in which this has been measured, but the trend over time comes through clearly. However the ability of survey techniques to capture the upper end of the income and wealth distribution has been limited and we cannot find a clear picture of the rich from this general work on inequality.

This paper described how between 1997 and 2005 Rosstat moved away from using survey data for monitoring inequality through introducing a procedure which on the one hand made artificial distortions to survey data by re-weighing them to "match" data collected in the macroeconomic statistics, while on the other relying entirely on calibrated models of distribution for inequality indices. While *mean*s for welfare indices are still collected in a similar way (i.e. through "macro" sources) as they were in 1980s, the approach and the data used to assess the ***distribution*** with respect to these welfare measures went through a number of substantial revisions.

Despite these advances, the true extent and even direction of change in inequality in Russia is not clear. The statement "most survey data suggest that Gini index is somewhere around 0.4 and is going up or down" is not adequate in a country where these issues are a subject of debate. Russia as other countries where there are issues of survey compliance and reporting may use other sources of available data which address the issue of non-response (by models of survey compliance or using tax records). Application of such models suggests that inequality is much higher (up to 20 points!) than survey data would suggest. But they do not cover all Russia or provide comparable data over time. Moreover, they are best developed outside of the official statistical practice. As raw data become available it will be only desirable to extend a set of indices and to introduce more nuances in the picture. The availability of raw survey data for 2003 and 2004 is a welcome sign and may herald the beginning of serious empirical work on inequality.

References

Aitchison J, and J. Brown. 1957. *The Lognormal Distribution*. Cambridge: Cambridge University Press.

Aivazian and S. Kolenikov, 2001 Poverty and Expenditure Differentiation of the Russian Population. — Working Paper N0 01/01, EERC, 2001

Alexeev, Michael, and Clifford Gaddy. 1993. "Income Distribution in the U.S.S.R. in the 1980s." Review of Income and Wealth, Series 39, pp 23–36.

Atkinson, Antony. 2003. "Income Inequality in OECD Countries: Data and Explanations." Revised version of paper prepared for the CESifo conference, "Globalization, Inequality and Well-Being," Munich, November 8–9, 2002.

Atkinson, Antony and John Micklewright. 1995. *Economic Transformation in Eastern Europe and the Distribution of Income*. Cambridge: Cambridge University Press.

Bandourian Ripsy, James McDonald and Robert Turkey 2002. A Comparison of Parametric Models of Income Distribution across Countries and Over Time. LIS paper.

Buckley, Robert, and Eugene Gurenko. 1997. [[check year/1998 in text]] "Housing and Income Distribution in Russia: Zhivago's Legacy." World Bank Research Observer 12 (1):19–32.

Commander, Simon, Andrei Tolstopiatenko, and Ruslan Yemtsov. 1999. "Channels of Redistribution: Inequality and Poverty in the Russian Transition*." Economics of Transition* 7, No. 2, pp. 411–47.

Cowell, Frank. 1995. *"Measuring Inequality*". Second Edition. London: Prentice Hall/Harvester Wheatsheaf.

Cowell, Frank, Julie Litchfield and Magda Mercader-Prats 1999. *"Income inequality comparisons with dirty data: The UK and Spain during the 1980s* ". Discussion paper DARP no 45, LSE, London.

Deville, J.-C. (1998). La correction de la non-réponse par calage ou par échantillonage équilibré. Papier présenté au colloque de la Société Statistique du Canada, Sherbrooke.

Dikhanov, Yuri. 1996. "Decomposition of Inequality Based on Incomplete Information*". IARIW 24th General Conference Paper*, Lillehammer.

Ersado, Lire 2006 "Azerbaijan's Household Survey Data: Explaining Why Inequality is So Low". World Bank Policy Research Working Paper 4009, September 2006

Flemming, J., and J. Micklewright. 1999. "Income Distribution, Economic Systems and Transition." In Handbook of Income Distribution , eds. A. B. Atkinson and F. Bourguignon, ch.14., Elsevier Science BV.

Gibson, John and Poduzov Alexander (2003). Assessing Welfare Indicators for Poverty Measurement in Russia. Background paper of the World Bank 2005a.

Gorodnicheno, Yuri, Klara Sabirianova Peter and Dmitry Stolyarov. 2008. A Bumpy Ride along the Kuznets Curve: Consumption and Income Inequality Dynamics in Russia. IZA (forthcoming)

Guriev S. and A. Rachinsky (2006) 'The Evolution of Personal Wealth in the Former Soviet Union and Central and Eastern Europe', UNU-WIDER Research Paper No 2006/120

Gustafsson, Bjorn, and Shi Li. 2001. The Anatomy of Rising Earnings Inequality in Urban China. Journal of Comparative Economics 29: 118–35.

Jovanovich, Branko (2001) 'Russian roller coaster: expenditure inequality and stability in Russia, 1994-98' Review of Income and Wealth, series 47, No 2, pp. 251-271.

Kalton, Graham and Tatiana Chernysheva (2005) Developing a Sampling Model for the HBS Based on 2002 National Census Data Background paper in World bank 2005a

Kislitsyna, O (2003) 'Income inequality in Russia during the transition. How can it be explained?' Economics Education and Research Consortium, working paper series, no 03/08.

Khan, Azizur R., and Carl Riskin. 2004. "China's Household Income and its Distribution, 1995 and 2002." http://www.economics.ucr.edu/seminars/spring04/05-28- 04AzizurKhan.pdf.

Kolenikov, Stanislav and Serguei Aivazian 'Expenditure modeling with a mixture of lognormal distributions"

Lehmann H., Wadsworth J. 2000 "Wage Arrears and the Distribution of Earnings: What can we learn from Russia?" IZA….

Li, Shi. 2000. "Changes in Income Inequality in China Transition." Chinese Academy of Social Sciences,

Lukianova, Anna. 2005. “Wage Differentiation and Phenomenon of Working Poor in Russia.” World Bank, Washington, DC.

Luttmer, Erzo. 2000. “Inequality and Poverty Dynamics in Transition Economies: Disentangling Real Effects from Data.” *Background Papers: Making Transition Work for Everyone: Poverty and Inequality in Europe and Central Asia*. World Bank. Washington, D.C.

Maleva T. M. et al. 2000, “The Middle Class in Russia: Quantitative and Qualitative Assessments”, Moscow: Bureau of Economic Analysis Foundation (in Russian).

Meng, Xin. 2000. “Economic Restructuring and Income Inequality in Urban China.” Research paper, Australian National University, .

Mitra, P., and R. Yemtsov (2007). ‘Increasing Inequality in Transition Economies: Is There More to Come?’, in Proceedings of the Annual World Bank Bank Conference on Development Economics “Beyond Transition”, 2006. Ed. by F. Bourgingnon and B. Pleskovic, 2007. Also available as World Bank Policy Research Working Paper 4007, World Bank: Washington DC

Popkin, Barry et. al. 2005, “Monitoring Economic Conditions in the Russian Federation: Russian Longitudinal Monitoring Survey 1992-2004,” University of North Carolina at Chapel Hill, April.

Ravallion M., and S. Chen. 1996. “Data in Transition: Assessing Rural Living Standards in Southern China.” China Economic Review 7(L).

---------. 2004. “China’s (Uneven) Progress against Poverty.” Policy Research Working Paper [[number?]], World Bank, Washington, DC.

Rosstat. 1998. *“Methodology of Aggregating Household Budget Survey Measures*”. Moscow: Goskomstat.

Rosstat. 1999b, 2003, 2005, 2006a and 2007. *“The Social Situation and the Standards of Living of the Russian Population (Sotsial;noe polozhenie I uroven’ zhizni naselenia Rossii)*”. Moscow: Goskomstat.

Rosstat. 1999c, 2000b, 2006c. *“Russian Statistical Yearbook*”. Moscow: Goskomstat.

Sheviakov, Alexey and Kiruta Alexander, 2001. “*Economic Inequality, Standards of Living and Poverty of Population in Russia (Measurement Methods and Analysis of Causal Dependencies)”*Center for Socioeconomic Measurement. Russian Academy of Sciences and Economics Education and Research Consortium. Moscow.

Sheviakov, Alexey “Possibilities and prospects of using statistical data in measuring the extent of concealed income,” Voprosy Statistiki, 6, 12-22 (2003)

Soptsov, V. (2006) Conducting Simulation Experiments with a View to Testing Alternative Methods of Weighting and Calibration of HBS Data. Background paper for World bank 2005 a

Tesliuc E. and Ovcharova L. 2006 Sensitivity of Poverty and Inequality Statistics to Alternative Definitions of Household Welfare Illustration Using the NOBUS Survey. World Bank , published in Russian as a book, Mir 2007.

Velikanova T., Kolmakov I, and Frolova E. 1996 “Improving the modeling methods for assessing the distribution of population by per cpiata incomes”/ Vorosy Statistiki, No 5, 1996

Volkova G, L. Migranova and I. Rimashevskaya. 1997 “Issues of methods for assessing the income inequality” / Voprosy Statistiki, No 2 1997.

World Bank. 2000c. *“Making Transition Work for Everyone: Poverty and Inequality in Europe and Central Asia”.* Washington, DC.

World Bank 2005a, “Russian Federation: Reducing poverty through growth and social policy reform”, The World Bank

World Bank 2005b, “Growth, Poverty and Inequality in Eastern Europe and Former Soviet Union”, The World Bank

Yemtsov Ruslan. 2003. “Quo Vadis Inequality and Poverty Dynamics across Russian Regions. ” Discussion Paper 2003/67, United Nations University/World Institute for Development Economics Research (UNU/WIDER).

Yemtsov, R., S. Cnobloch, and C. Mete. 2006. “Evolution of the Predictions of Earning during Transition.” World Bank, Washington, DC.

Annex 1 Components and definitions of household welfare indicators based on HBS (Rosstsat)

No.	Welfare indicator	Calculation procedure
1	Food cash expenditures	Total expenditures on: bakery products and grains, meat and meat products, fish and other seafood, dairy products and eggs, oils and fats, fruit and berries, vegetables, including potatoes, sugar, honey and confectionery, other food products, non-alcoholic beverages
2	Expenditures on dining out	Total expenditures on: restaurants, cafes, diners
3	Expenditures on alcoholic beverages	Total expenditures on: hard liquor, wine,beer
4	Expenditures on non-food goods	Total expenditures on: fabrics, clothing and underwear, headwear and accessories, shoes, building materials for current repairs, household fuel, furniture, rugs and carpets, household textiles, household appliances, kitchenware and utensils, tools and equipment for home and garden, household chemicals and other household supplies and accessories, drugs, medical equipment, vehicles, spares and accessories, fuel and lubricants, cultural goods, games, hobby supplies and sports gear, books, newspapers and stationery, perfumes, cosmetics and personal care products, watches, jewelry and other personal effects, tobacco products and narcotics
5	Expenditures on services	Total expenditures on: professional cleaning and repair of clothing and shoes, rent, home repairs, utilities, repairs of household appliances, furniture and other household goods, other household services, medical services, sanatorium and health-building services, vehicle maintenance and repair, transportation services, communication services, cultural products repair, veterinary services, recreational and cultural activities services, gambling, recreational centers' services, tourism, educational services, housing provision services, personal services, commercial sex services, social protection services, child care services, insurance services, other services
6	**Consumer expenditures**	1 + 2 + 3 + 4 + 5
7	Inputs into private subsidiary plots and other production	Total expenditures on: farming technology and equipment, seeds, fodder, fertilizer, raw materials, fuel, livestock and poultry, services received in the course of production, taxes and mandatory payments
8	Expenditures on buying real estate	Total expenditures on acquiring real estate
9	Expenditures on buying jewelry	Total expenditures on acquiring jewelry and antiques
10	Expenditures on construction and capital repair services	Total expenditures on construction
11	Expenditures on intermediate consumption and gross accumulation	7 + 8 + 9 + 10
12	Taxes, levies and payments	Total payments for: personal income taxes, other taxes and mandatory payments
13	Other expenditures	Total expenditures on: alimony, gifts, advance payments, other expenditures
14	Household cash expenditures	6 + 11 + 12 + 13
15	Total amount of savings	Bank deposits, invested in securities, exchanged for foreign currency, loaned, put aside
16	Amount of loan and expended savings	Derived amount
17	Increment in financial assets	15 – 16

18	**Household cash income**	**14 + 17**
19	Value of food products received in kind	Total of estimates based on average prices of food products received in kind
20	Value of non-food goods and services received as remuneration for labor	Total of estimated values of the provided services and used non-food goods
21	Value of in-kind benefits provided by the employer	Total of benefits: transportation benefits, housing benefits, construction benefits, meals benefits, holiday benefits, medical benefits, child care benefits, educational benefits, other benefits
22	Value of in-kind benefits provided by the state	Total of benefits: transportation benefits, housing benefits, construction benefits, meals benefits, holiday benefits, medical benefits, child care benefits, educational benefits, other benefits
23	Value of in-kind benefits	21 + 22
24	Value of consumed in-kind receipts	19 + 20 + 23
25	Gross household income	18 + 24
26	Value of food products given away by household	Given away for processing, fed to the livestock, given away as gifts
27	Expenditures on end consumption	6 + 19 + 20 + 21 – 26
28	**Disposable household resources**	**14 + 15 *or* 25 + 16**

Note: main indicators used for production of inequality indices are highlighted in bold.
Source; Gibson and Poduzov, 2003, based on Rosstat, 1997

Annex Table 2 China: Gini Coefficients for different definitions of welfare indices and for various surveys

			Rural			Urban		
Welfare index	**Study**	**Data source**	**1988**	**1995**	**2001–2**	**1988**	**1995**	**2001–2**
Income, no deflation	Official	SSB	0.30	0.35	0.41	0.21	0.31	0.34
Consumption, no deflation	Official	SSB				0.24	0.32	0.35
Income , COL adjusted	Ravallion and Chen (2004)	SSB	0.297	0.334	0.365	0.211	0.283	0.323
Income, COL adjusted	Wu and Perloff (2004)	SSB	0.300	0.338	0.343	0.201	0.221	0.269
Consumption, COL adjusted	Benjamin et al. (2004)	SSB				0.22	0.28	0.32
Income. COL adjusted	Khan and Riskin (1998, 2004)	CASS	0.338	0.416	0.375	0.233	0.332	0.318
Income, no deflation	Benjamin et al. (2004)	CHNS	0.39 [a]	0.41 [a]	0.46 [a]	0.29 [a]	0.35 [a]	0.38 [a]

Note : SSB (NBS) = State Statisitical Bureau based on household budget survey; CASS = Economics Institute of the Chinese Academy of Social Sciences Survey; CHNS =China Health and Nutrition Study (CHNS) data,
a. 1989, 1997 and 2000.

Annex Table 3 Gini Coefficients by regions of Russia, 2002: official "model" and direct HBS-based; for various measures of household welfare

Regions	Per capita income, published ("modeled")	Per capita income, direct survey estimate	Per capita disposable resources, survey	Per capita consumption, survey
RUSSIA (real)*	**0.385**	**0.412**	**0.388**	**0.330**
Russia (nominal)	*0.403*	*0.452*	*0.419*	n.a.
Belgorod oblast	0.340	0.311	0.264	0.239
Briansk oblast	0.330	0.388	0.367	0.298
Vladimir oblast	0.300	0.318	0.297	0.260
Voronezh oblast	0.353	0.411	0.363	0.319
Ivanovo oblast	0.307	0.296	0.283	0.248
Kaluga oblast	0.313	0.317	0.282	0.257
Kostroma oblast	0.339	0.329	0.296	0.264
Kursk oblast	0.335	0.466	0.405	0.313
Lipetsk oblast	0.362	0.391	0.362	0.287
Moscow oblast	0.345	0.367	0.355	0.317
Orel oblast	0.359	0.401	0.351	0.293
Riazan oblast	0.331	0.362	0.328	0.304
Smolensk oblast	0.336	0.339	0.305	0.277
Tambov oblast	0.361	0.373	0.335	0.285
Tver oblast	0.305	0.409	0.375	0.303
Tula oblast	0.311	0.312	0.284	0.249
Yaroslavl oblast	0.340	0.361	0.335	0.287
Moscow city	***0.609***	***0.368***	***0.364***	***0.340***
Karelia Republic	0.331	0.421	0.407	0.336
Komi Republic	0.411	0.447	0.420	0.378
Arkhangelsk oblast	0.334	0.394	0.376	0.292
Vologda oblast	0.331	0.403	0.403	0.329
Kaliningrad oblast	0.321	0.405	0.377	0.342
Leningrad oblast	0.304	0.298	0.284	0.241
Murmansk oblast	0.364	0.395	0.373	0.282
Novgorod oblast	0.352	0.419	0.380	0.324
Pskov oblast	0.315	0.453	0.410	0.328
St. Petersburg city	0.347	0.303	0.291	0.261
Adygeia Republic	0.333	0.391	0.422	0.388
Dagestan Republic	0.357	0.359	0.312	0.283
Ingushetiya Republic	0.317	0.355	0.306	0.292
Kabardino-Balkariya Republic	0.329	0.432	0.363	0.314
Kalmykiya Republic	0.373	0.400	0.379	0.322
Karachaevo-Cherkessiya Rep	0.351	0.419	0.360	0.294
Severnaya Osetiya Republic	0.329	0.362	0.327	0.285
Krasnodar territory	0.377	0.415	0.377	0.324
Stavropol territory	0.341	0.496	0.436	0.320
Astrakhan oblast	0.338	0.473	0.465	0.384
Volgograd oblast	0.328	0.407	0.373	0.324
Rostov oblast	0.364	0.379	0.339	0.313

Regions	Per capita income, published ("modeled")	Per capita income, direct survey estimate	Per capita disposable resources, survey	Per capita consumption, survey
Bashkortostan Republic	0.368	0.479	0.414	0.346
Mariy El Republic	0.346	0.419	0.351	0.314
Mordoviya Republic	0.332	0.396	0.339	0.299
Tatarstan Republic	0.374	0.435	0.398	0.320
Udmurtiya Republic	0.308	0.384	0.377	0.300
Chuvashiya Republic	0.308	0.366	0.323	0.288
Kirov oblast	0.306	0.389	0.334	0.291
Nizhniy Novgorod oblast	0.343	0.395	0.366	0.338
Orenburg oblast	0.319	0.408	0.348	0.291
Penza oblast	0.309	0.384	0.335	0.284
Perm oblast	0.397	0.472	0.450	0.366
Samara oblast	0.424	0.461	0.450	0.326
Saratov oblast	0.321	0.394	0.351	0.327
Ulianovsk oblast	0.360	0.377	0.335	0.313
Kurgan oblast	0.375	0.441	0.373	0.309
Sverdlovsk oblast	0.356	0.360	0.338	0.315
Tumen oblast	***0.448***	***0.479***	***0.463***	***0.363***
Chelyabinsk oblast	0.350	0.402	0.380	0.301
Altay Republic	0.319	0.397	0.353	0.309
Buriatiya Republic	0.399	0.415	0.362	0.328
Tyva Republic	0.347	0.447	0.422	0.386
Khakassiya Republic	0.346	0.413	0.367	0.313
Altay territory	0.361	0.441	0.362	0.295
Krasnoyarsk territory	***0.397***	***0.474***	***0.409***	***0.324***
Irkutsk oblast	***0.399***	***0.427***	***0.396***	***0.345***
Kemerovo oblast	0.367	0.390	0.360	0.314
Novosibirsk oblast	0.349	0.397	0.340	0.307
Omsk oblast	0.362	0.467	0.434	0.375
Tomsk oblast	0.368	0.456	0.419	0.310
Chita oblast	0.320	0.441	0.383	0.333
Sakha (Yakutia) Republic	0.360	0.452	0.421	0.361
Primorie territory	0.338	0.377	0.364	0.313
Khabarovsk territory	0.359	0.458	0.420	0.330
Amur oblast	0.342	0.395	0.357	0.295
Kamchatka oblast	0.350	0.387	0.366	0.297
Magadan oblast	0.326	0.440	0.424	0.399
Sakhalin oblast	0.355	0.445	0.415	0.355
Evreiskaya autonomous oblast	0.349	0.424	0.372	0.288
Chukotka autonomous region	0.373	0.381	0.353	0.333

*Poverty lines used as deflators; capital city and resource rich regions are highlighted

Sources: Published offical data – Rostat 203, survey based estimates – World bank 2005a.

Annex Table 4 Data on means an Gini indices by regions of Russia, 2003: official "model" , HBS direct survey-based and NOBUS- based.

	Macro data		HBS-based income data				HBS o.	NOBUS-based welfare indices			
description	Mean PC Income	Gini for PC inc.	PCincome original weights	PC income Re-weighted	*Gini* ind. Original weights	*Gini* re-weighted	Disp. Resources PC Gini	Mean PC income	Gini index for PC inc.	Mean PC consumption	Gini iindex for PC cons.
Russia (nominal)	***5170***	***0.400***	***2122***	***3180***	***0.374***	***0.450***	***0.422***	***2001***	***0.512***	***2921***	***0.282***
Belgorod oblast	3357	0.348	1770	2417	0.282	0.340	0.325	977	0.420	3140	0.238
Briansk oblast	3135	0.350	1500	2043	0.325	0.379	0.328				
Vladimir oblast	2842	0.317	1617	2060	0.267	0.323	0.306	1870	0.421	3610	0.286
Voronezh oblast	3391	0.376	1560	2293	0.357	0.415	0.425				
Ivanovo oblast	2293	0.325	1657	1887	0.267	0.291	0.278				
Kaluga oblast	3340	0.327	2340	2743	0.292	0.321	0.297	1730	0.378	2770	0.238
Kostroma oblast	3089	0.346	1870	2453	0.268	0.337	0.330	1420	0.408	2680	0.289
Kursk oblast	3371	0.349	1283	1850	0.307	0.363	0.311	1240	0.317	2710	0.272
Lipetsk oblast	3563	0.371	2297	2927	0.367	0.415	0.400	1410	0.418	3240	0.249
Moscow oblast	4425	0.360	2160	3293	0.306	0.382	0.373	2840	0.528	3200	0.240
Orel oblast	3225	0.365	2067	2657	0.354	0.393	0.360				
Riazan oblast	3309	0.344	1753	2270	0.355	0.397	0.359				
Smolensk oblast	3724	0.347	1170	1770	0.286	0.334	0.304	1660	0.375	3090	0.219
Tambov oblast	3416	0.375	1653	2500	0.357	0.459	0.409				
Tver oblast	3021	0.311	1890	2330	0.348	0.387	0.373				
Tula oblast	3383	0.318	1850	2397	0.249	0.306	0.283	1450	0.308	3330	0.241
Yaroslavl oblast	4267	0.360	2227	2997	0.314	0.372	0.358	2010	0.381	3100	0.254
Moscow city	***16827***	***0.615***	***3313***	***7167***	***0.254***	***0.361***	***0.358***	***3840***	***0.553***	***2740***	***0.243***
Karelia Republic	4937	0.341	2457	3247	0.332	0.383	0.365	3190	0.429	3810	0.300
Komi Republic	7477	0.427	2613	4000	0.389	0.475	0.439				
Arkhangelsk oblast	4834	0.353	2560	3400	0.345	0.404	0.385				
Vologda oblast	4412	0.361	2497	3333	0.353	0.401	0.373	2170	0.443	2850	0.286
Kaliningrad oblast	3807	0.325	1937	2643	0.335	0.395	0.374	2350	0.354	3210	0.300
Leningrad oblast	3052	0.313	2123	2493	0.258	0.295	0.285	2750	0.532	3210	0.212
Murmansk oblast	7135	0.373	4233	5100	0.331	0.357	0.356	3710	0.336	3570	0.232
Novgorod oblast	3715	0.360	1850	2690	0.364	0.437	0.416				
Pskov oblast	3556	0.336	1770	2567	0.343	0.433	0.393				
St. Petersburg city	6851	0.388	3157	4667	0.261	0.345	0.338	3410	0.502	3430	0.226
Adygeia Republic	2549	0.339	1557	1977	0.304	0.353	0.336				
Dagestan Republic	2125	0.355	847	1200	0.331	0.349	0.309	770	0.530	2050	0.295
Ingushetiya Republic	1402	0.317	1117	1207	0.214	0.235	0.241	307	0.598	1800	0.292
Kabardino-Balkariya	2571	0.328	1153	1663	0.360	0.393	0.346	1180	0.441	2630	0.291
Kalmykiya Republic	2100	0.373	1247	1607	0.353	0.387	0.368	629	0.369	1840	0.220
Karachaevo-Cherkess	2619	0.353	1430	1960	0.360	0.403	0.356	813	0.514	2660	0.222
Severnaya Osetiya	2596	0.365	1353	2020	0.334	0.388	0.358	1250	0.396	3470	0.256
Krasnodar territory	3662	0.385	1560	2380	0.324	0.397	0.361	1530	0.433	2650	0.279
Stavropol territory	3072	0.353	1547	2677	0.368	0.533	0.488	1110	0.457	2630	0.256
Astrakhan oblast	3864	0.361	2113	2950	0.381	0.451	0.444	1550	0.486	2800	0.285
Volgograd oblast	3803	0.355	1627	2460	0.328	0.408	0.367				
Rostov oblast	4024	0.373	1663	2563	0.306	0.377	0.350	1620	0.351	2590	0.285

	Macro data		HBS-based income data				HBS o.	NOBUS-based welfare indices			
description	Mean PC Income	Gini for PC inc.	PCincome original weights	PC income Re-weighted	*Gini* ind. Original weights	*Gini* re-weighted	Disp. Resources PC Gini	Mean PC income	Gini index for PC inc.	Mean PC consumption	Gini iindex for PC cons.
Bashkortostan	4153	0.391	1763	2923	0.404	0.495	0.443	1530	0.500	2870	0.309
Mariy El Republic	2189	0.363	1463	1777	0.372	0.406	0.354	1130	0.421	2570	0.299
Mordoviya Republic	2720	0.338	1487	1993	0.362	0.404	0.360	1030	0.442	2830	0.258
Tatarstan Republic	4273	0.388	1603	2617	0.319	0.400	0.376	2567	0.457	2840	0.281
Udmurtiya Republic	3098	0.313	2020	2413	0.324	0.351	0.342	1590	0.470	2720	0.295
Chuvashiya Republic	2749	0.329	1380	2297	0.335	0.349	0.322	1420	0.511	3060	0.310
Kirov oblast	3094	0.321	1783	2260	0.334	0.379	0.339				
Nizhniy Novgorod o	4000	0.344	1687	2477	0.301	0.363	0.342	1770	0.474	2990	0.271
Orenburg oblast	3135	0.335	1817	2323	0.364	0.411	0.357	1230	0.407	2660	0.263
Penza oblast	2765	0.325	1633	2060	0.302	0.356	0.328	1110	0.400	3240	0.215
Perm oblast	5257	0.412	2010	3467	0.364	0.471	0.441	1950	0.425	2970	0.290
Samara oblast	5788	0.430	2423	3600	0.330	0.399	0.388	1870	0.483	2910	0.256
Saratov oblast	3337	0.341	1637	2363	0.357	0.405	0.371	1450	0.454	2780	0.272
Ulianovsk oblast	3063	0.366	1530	2080	0.339	0.398	0.382	1490	0.319	3270	0.237
Kurgan oblast	2867	0.387	1623	2110	0.388	0.426	0.376	1390	0.466	2610	0.310
Sverdlovsk oblast	5278	0.380	2207	3367	0.344	0.411	0.398				
Tumen oblast	***10556***	***0.450***	***2537***	***4000***	***0.380***	***0.470***	***0.445***	***3970***	***0.534***	***3160***	***0.333***
Chelyabinsk oblast	3998	0.363	2287	2877	0.339	0.396	0.380	3590	0.444	3030	0.291
Altay Republic	2876	0.324	1660	2247	0.375	0.430	0.400	1650	0.434	3090	0.340
Buriatiya Republic	3857	0.401	1397	2080	0.362	0.412	0.403				
Tyva Republic	2873	0.352	1497	1977	0.398	0.466	0.417	978	0.523	1650	0.370
Khakassiya Republic	3765	0.358	2100	2630	0.356	0.380	0.350	1770	0.322	3080	0.284
Altay territory	2895	0.367	1953	2433	0.390	0.452	0.398	1510	0.446	3020	0.303
Krasnoyarsk terr.	***5509***	***0.401***	***2647***	***3867***	***0.359***	***0.420***	***0.366***	***2560***	***0.507***	***3480***	***0.337***
Irkutsk oblast	***4550***	***0.406***	***2660***	***3400***	***0.398***	***0.443***	***0.422***			***2940***	***0.295***
Kemerovo oblast	4907	0.378	1820	2927	0.343	0.417	0.401	2080	0.418	3150	0.284
Novosibirsk oblast	3893	0.367	1847	2677	0.358	0.416	0.371				
Omsk oblast	4513	0.385	1750	2880	0.368	0.461	0.422				
Tomsk oblast	5407	0.374	2140	3043	0.359	0.450	0.406	3150	0.632	3190	0.281
Chita oblast	4016	0.368	2263	2767	0.422	0.448	0.430			2150	0.313
Sakha (Yakutia) Rep	8240	0.385	3400	5233	0.399	0.467	0.462				
Primorie territory	4246	0.341	2457	3177	0.339	0.370	0.338	2340	0.450	2840	0.294
Khabarovsk territory	6205	0.379	3293	4667	0.366	0.449	0.421	2700	0.550	2900	0.272
Amur oblast	3852	0.345	1983	2663	0.320	0.381	0.341				
Kamchatka oblast	7101	0.364	3933	5367	0.314	0.391	0.374				
Magadan oblast	8185	0.382	2393	3833	0.343	0.429	0.424	2920	0.506	2780	0.249
Sakhalin oblast	7676	0.376	3153	4500	0.345	0.394	0.366				
Evreiskaya AO	4062	0.352	2433	3027	0.333	0.380	0.337	1450	0.463	2530	0.302
Chukotka AR	13664	0.390	3400	5067	0.378	0.446	0.424	3840	0.548	1940	0.339

Sources: Published (official data) Rostat 2004; other – own estimates based on raw HBA and NOBUS data.
NOBUS regions with too small sample size to derive regionally representative data are blank; these data are used in national aggregates.
PC- per capita