

# Calculating cost-of-living deflators without data on prices: A simple non-parametric approach

Anastasiya Lisina

LISER & University of Luxembourg

Philippe Van Kerm

LISER & University of Luxembourg

Draft, please do not cite or circulate, version March 6, 2023.

## Abstract

When studying large countries, having access to infra-national, regional cost-of-living indicators is important (i) to be able to make (income- or expenditure-based) welfare comparisons across regions net of differences in prices, or (ii) to construct national welfare measures based on individual-level measures of welfare that reflect local variations in prices faced by households. Regional cost-of-living indices are however not always available or may not be updated as regularly as national CPI series. We therefore propose a simple approach for calculating regional cost-of-living deflators that do not require access to data on prices and that can be implemented directly from standard surveys on income and living conditions. To calculate the regional cost-of-living index, we apply methods proposed for the similar problem of estimating equivalence scale parameters: we use subjective satisfaction data to capture indirect utility and adopt non-parametric matching methods to construct regional cost-of-living indices. As an illustration, we construct price indices for Russian regions from the Russian Longitudinal Monitoring Survey for the period 2000–2015 and compare our cost-of-living deflators with those from the Statistical Office of Russia (available from 2009 only). Over the years covered by both series, the indices correlate highly but not perfectly. Application of regional price indices does not appear to alter inequality trends, but income levels.

*Keywords* — cost-of-living indices; regional deflators; income; subjective well-being; matching; Russia; RLMS

# 1 Introduction

To make meaningful welfare comparisons based on income (or expenditure) across individuals living in different places (or observed at different points in time), it is important to adjust for differences in prices faced by individuals. Determining how to make valid inter-temporal and international welfare comparisons is a whole field of research in itself (Almås, 2012). In large countries, prices may also vary substantially across different regions. Being able to account for infra-national (regional) cost-of-living differences is important in at least two situations. First, and most obviously, it is critical to correct for the local price differences if one is interested in making direct welfare comparisons of regions based on income or expenditure data. Not doing so would lead to an overestimation of the welfare level of people living in “expensive” regions and underestimate the welfare of people living in comparatively “cheaper” regions. Second, one also need to account for regional prices differences when constructing aggregate, national series of social welfare indicators—such as inequality or poverty series—since such series should reflect heterogeneity of individual welfare across the country. To adequately reflect such heterogeneity in individual welfare, individual incomes need to be adjusted for differences in cost-of-living.

Income deflators used to adjust for differences in the prices faced by different individuals are typically based on price indices constructed from data on prices and consumption over time and/or across countries, such as the consumer price index or purchasing power parity series. While these are almost universally applied by practitioners, the construction of such series is complex, data demanding and not without conceptual and empirical problems—such as how to reflect individual differences in consumption baskets, the change in consumer behaviour, the introduction of new products and outlets, changes in the quality of goods others (Diewert, 1998). The development of those series is therefore typically entirely conducted by specialist agencies and statistical offices.

Beyond potential concerns about the measurement and construction of price indices, a more pragmatic issue for practitioners is that regional price indices are not always available, or not released as frequently as national series by statistical offices in charge. This may practically limit the possibility to do adequate adjustments for regional cost-of-living differences in welfare analysis.

We therefore propose a simple approach for constructing regional cost-of-living deflators that does not require access to data on prices but that can be implemented directly from standard surveys on income and living conditions. This allows practitioners to generate their own regional income deflators in cases where regional prices indices are not readily available. Moreover, even

when regional price indices do exist, it can be of interest to assess the robustness of one’s results to the use of an alternative series of cost-of-living measures, especially in the light of the debates about the best way to construct price deflators based on price data.

The proposed approach for the calculation of regional cost-of-living indices directly draws upon a methodology that has been advocated for the similar problem of estimating equivalence scale parameters. In a nutshell, we use subjective satisfaction data to capture indirect utility given income, needs and local prices and then adopt a non-parametric matching method to assess “counterfactual” indirect utility given income, needs and prices in a reference region. Finally, use those matched pairs to construct regional cost-of-living indices that equalize, in expectation, counterfactual indirect utility across regions for households of identical needs. The approach has two advantages. First, the data required for implementing these methods merely consist in income, household attributes that capture differences in needs (e.g., household composition, age of members, home ownership status, employment status, and possibly educational attainment, etc.) and, crucially, a variable that approximates household financial welfare. In the tradition of the “Leyden school” and of more recent applications to the estimation of equivalence scale parameters, the latter can be assessed using self-reported, subjective variables of satisfaction with one’s own financial situation (Biewen & Juhasz, 2017; Van Praag, 1968; Melenberg & van Soest, 1995; Bellemare et al., 2002).<sup>1</sup> Many standard surveys on income and living conditions collect such variables. Second, the methodology is simple and involves no parametric assumptions.

Two key conditions must be satisfied for the approach to be valid. First and most obviously, the subjective variable of household financial welfare must accurately reflect differences in financial situations across responding households. This requires comparability of subjective responses across households sharing similar observable characteristics. It also requires that there is sufficient variation in the ordinal responses of households in the survey. Second, conditional independence between the unobserved, idiosyncratic component of (subjective) financial welfare and region of residence must hold conditionally on observable household characteristics, income and regional prices. This second condition means that the financial welfare of a household with a given income and observable characteristics can only vary across regions as a function of the regional cost-of-living differences. No other regional factor that should *not* be reflected in the cost-of-living index influence subjective financial welfare.

---

<sup>1</sup>The use of a variable capturing financial welfare independently on income to construct price indices and equivalence scales also underlies early approaches based on the Engel curve where the food share in expenditure captures financial welfare.

We illustrate our approach on Russian regions using data from the Russian Longitudinal Monitoring Survey and compare our cost-of-living deflators with official cost-of-living indices provided by the Russian statistical office. Our cost-of-living indices are similar to the official indices in years where the two series co-exist. We also re-estimate inequality and poverty series using both deflators and provide evidence for a gap in inequality and poverty trends when not accounting for differences in prices. We find that Russia experienced an even larger decrease in inequality (driven by a decrease in between-region inequality estimates). The correction obtained with the two series are in the same direction but differ somewhat in magnitude. The overall trends are however reassuringly similar.

The remainder of the paper is organized as follows. In section 2, we briefly discuss the related literature, notably the Leyden school and the use of subjective equivalence scales. Section 3 introduces our approach and the estimation methods. Section 4 presents the data, estimates of cost-of-living indices for Russian regions and application to inequality and poverty analysis. Section 5 concludes.

## 2 Methods

Our approach to the construction of cost-of-living indices is “direct” in the sense that it does not rely on the specification of a structural model. It consists in evaluating the monetary value of living in every region covered by the data relative to a reference region and use these monetary evaluations as cost-of-living indices.

The basic argument is that average income differences between any two observationally identical households who both reach the same level of financial welfare but who are randomly drawn from two different regions reflect cost-of-living differences between these two regions. Cost-of-living indices are thereby obtained non-parametrically by estimating average income differences in a sample of matched pairs of households where matching is based on two conditions: (i) matched households reach identical “financial welfare” (defined shortly) and (ii) matched households have (approximately) identical characteristics on dimensions that reflect potential differences in needs (such as household size and composition, home-ownership status, etc.).

### Definition of the cost-of-living index

To formalize this, let us consider a country with  $R$  regions. Let  $S_{ir}$  denote the financial welfare of a household  $i$  residing in region  $r \in \{1, \dots, R\}$  and write  $S_{ir}$  as a function of income  $Y_i$ , observable

characteristics  $Z_i$ , prices  $P_r$  and an idiosyncratic term  $\epsilon_{ir}$

$$S_{ir} = V(Y_i, Z_i, P_r, \epsilon_{ir}). \quad (1)$$

In a standard utility maximization perspective,  $V$  is an indirect utility function for expenditure  $Y_i$  and prices  $P_r$ . In the absence of a structural model, we view it more directly as representing the level of subjective satisfaction that household  $i$  achieves given income, characteristics (that reflect needs) and prices in region  $r$ .

We define a cost-of-living index for region  $r$  to be a region-specific deflator  $d_r$

$$Y_{ir}^* = \frac{Y_i}{d_r} \quad (2)$$

where  $Y_{ir}^*$  is the income needed for household  $i$  to achieve the same level of financial welfare it achieves in region  $r$  if it was facing prices of a reference region 1:

$$V(Y_i, Z_i, P_r, \epsilon_{ir}) = V(Y_{ir}^*, Z_i, P_1, \epsilon_{ir}). \quad (3)$$

Classic structural approaches would proceed from here by using demand functions to estimate the expenditure level  $Y_{ir}^*$  satisfying (3) and would recover  $d_r$  accordingly as  $Y_i/Y_{ir}^*$ . Using a direct measure of financial well-being instead, the deflator in (2) determines the income that would be required to leave household  $i$  with the same financial well-being in region 1 it reaches in its region of residence  $r$ . The former,  $V(Y_{ir}^*, Z_i, P_1, \epsilon_{ir})$ , is however unobserved so  $d_r$  cannot be determined yet.

The matching solution we propose is to substitute the unobserved  $V(Y_{ir}^*, Z_i, P_1, \epsilon_{ir})$  in (3) by  $V(Y_j, Z_i, P_1, \epsilon_{j1})$  where  $j$  is a household that resides in region 1, has  $Z_i = Z_j$  and has the same level of financial welfare as household  $j$ ,  $S_{ir} = S_{j1}$ . The deflator then becomes the solution to

$$V(Y_i, Z_i, P_r, \epsilon_{ir}) = V(Y_{ir}^*, Z_i, P_1, \epsilon_{j1}) \quad (4)$$

$$= V(Y_j, Z_i, P_1, \epsilon_{j1}) \quad (5)$$

which holds by construction (since  $i$  and  $j$  are chosen such that  $S_{ir} = S_{j1}$ ), and therefore

$$d_r = \frac{Y_i}{Y_j}. \quad (6)$$

Such a solution is not unique: any pair of matched households  $i \in \{1, \dots, N^r\}$  and  $j \in \{1, \dots, N^1\}$  that satisfy the conditions  $Z_i = Z_j$  and  $S_{ir} = S_{j1}$  possibly determine a different deflator according to (6). Therefore, we redefine our cost-of-living index to be a region-specific constant such that (5) is satisfied in expectation over all  $i \in \{1, \dots, N^r\}$  and  $j \in \{1, \dots, N^1\}$  satisfying  $Z_i = Z_j$  and  $S_{ir} = S_{j1}$ . The cost-of-living index is therefore the solution to

$$E\left(\frac{Y_i}{d_r} - Y_j\right) = 0. \quad (7)$$

## Key assumptions

The approach rests on two key assumptions. The first key assumption for the matching argument to be valid—namely for unobserved  $V(Y_{ir}^*, Z_i, P_1, \epsilon_{ir})$  to be well captured by  $V(Y_j, Z_i, P_1, \epsilon_{j1})$  in expectation—is conditional mean independence of the idiosyncratic component in  $V$ :

$$E(\epsilon_{ir}|Y_i, Z_i, P_r) = E(\epsilon_{ir}|Y_i, Z_i).$$

This means that the idiosyncratic part of the financial satisfaction of household  $i$  in region  $r$  does not depend on regional characteristics. In another words, all regional characteristics that influence financial well-being are factored in the cost of living.

The second key assumption is independence of the idiosyncratic component on income  $Y$

$$\epsilon_{jr} \perp Y_j | Z_j \quad \forall j, r.$$

This second assumption means that the idiosyncratic components do not vary systematically with income (conditional on  $Z_j$ ) – so  $Y_j$  is a valid proxy for  $Y_{ir}^*$  when matching is done on  $Z$ , but not on  $\epsilon$ .

## Implementation

Practically, the matching estimation can be implemented as follows. Start with a sample of  $N$  households distributed across  $R$  regions where the data for each household consists of the tuples  $\{Z_i, Y_i, S_{ir}, r\}$ .  $Z_i$  is the vector of characteristics of household  $i$ : (i) demographic characteristics such as family size and composition, age of spouses and children; (ii) variables that may impact the “needs” and cost of living (such as home-ownership status, rural/urban area).  $Y_i$  is the household

income.  $S_{ir}$  is the variable of satisfaction with financial situation (e.g., “Are you satisfied with your financial situation?”, “how easy is it to make ends meet?”, etc.) and  $r$  denotes the region of residence of household  $i$ .

For each household in the sample, find a matched household in each of the other regions. Denote by  $m(i, s)$  the identifier of the household from region  $s$  matched to household  $i$ . The match is based on  $S_{ir}$  and  $Z_i$ . The match should be exact on financial welfare, so  $S_{ir} = S_{m(i,s)r}$  while the match can be approximate on household characteristics if the number of characteristics is large, so  $Z_i = Z_{m(i,s)}$  or  $Z_i \approx Z_{m(i,s)}$ . Differences in the logarithm of income ( $\ln Y_i - \ln Y_{m(i,s)}$ ) across all matched pairs should therefore reflect income differences for people of identical financial welfare and identical (or similar) characteristics, but living in different regions.

An estimate of the “pairwise” cost-of-living deflator for any pair of regions  $(r, s)$ ,  $\hat{d}_{r,s}$ , is then such that scaling  $Y_i$  by  $\hat{d}_{r,s}$  eliminates differences in average log income for the matched pairs between region  $r$  and  $s$ . A simple pairwise deflator could be obtained as

$$\frac{1}{N_r} \sum_{i=1}^{N_r} \left( \ln \frac{Y_i}{\hat{d}_{r,s}} - \ln Y_{m(i,s)} \right) = 0 \quad (8)$$

$$\hat{d}_{r,s} = \exp \left( \frac{1}{N_r} \sum_{i=1}^{N_r} (\ln Y_i - \ln Y_{m(i,s)}) \right) \quad (9)$$

If  $\hat{d}_{r,s}$  is about 1, then this indicates that there is no systematic cost of living difference between the regions  $r$  and  $s$ . This can be done for all pairs of regions. However, to obtain an overall cost-of-living deflator further normalization is required.

A simpler and more efficient estimation is feasible, however. First, stack all matched log income differences  $D_{is} = (\ln Y_i - \ln Y_{m(i,s)})$  for all households  $i$  and regions  $s$  in a long vector (of size  $N \times R$ ). Then, define two sets of dummy variables – one for region  $r$  in which household  $i$  resides (the origin set) and one for the region  $s$  of the matched value (the matched set). Reference region 1 is omitted from both sets of dummy variables. Finally, apply a constrained OLS regression of  $D_{is}$  on the two sets of dummy variables, constraining the coefficients on the dummy variables for the matched set to be equal to minus the coefficient of the dummy variable of the same regions in the origin set. Denoting by  $\hat{\beta}_r$  the resulting coefficients for region  $r$  and setting  $\hat{\beta}_1 = 0$  (for the omitted, reference, region), the final estimate of the cost-of-living deflator is

$$\hat{d}_r = \exp \left( \hat{\beta}_r \right).$$

where  $\hat{d}_1$  is equal to 1 for the reference region. The cost-of-living deflators  $\hat{d}_r$  can then be used to adjust household incomes or expenditures to adjust for regional cost-of-living differences. Note that confidence intervals for the estimated deflators can be derived from the regression coefficients by applying the standard delta method on exponentiated coefficients.

The procedure is easy and flexible. Modern statistical software have packaged routines for forming matched pairs and running constrained regression. It allows the cost-of-living deflators to be estimated for different geographical units (cities, regions, countries etc.) without imposing specific utility functions or other structural parameters and without data on prices.

There is no free lunch however: the validity of the approach hinges on the key assumptions described above to be satisfied. In particular such assumptions in the context of using subjective financial satisfaction variables imply the following. First, we must assume that every respondent is able to evaluate her financial situation (e.g., on a scale from one to five). Second, we must assume that verbal responses to subjective questions across respondents are comparable, implying that verbal labels signify the same welfare experience to every respondent – at least among respondents with similar observable characteristics. Such assumptions are similar however to those made in the derivation of subjective equivalence scales which are now routinely estimated.

In fact the procedure bears much similarity with the subjective approach to the construction of equivalence scales (see, e.g., Biewen & Juhasz (2017)). To see the similarity, one needs to change the label  $r$  for regions to denote instead different household composition classes and much of the procedure carries over. However, whereas we estimate separate unconstrained cost-of-living measures for each different region  $r$ , the literature on equivalence scale typically imposes a functional form onto the resulting  $d_r$ , such as, e.g,  $d_r = e(n_a, n_c) = (1 + \alpha(n_a - 1) + \beta n_c)^\theta$  where  $n_a$  and  $n_c$  would denote the number of adults and children in the household respectively. Non-linear regression methods are then be applied to estimate parameters  $\alpha$ ,  $\beta$  and  $\theta$ . No such second step is needed in the context of regional cost-of-living deflators.

### 3 Application to Regions in Russia

We illustrate the proposed framework on the real-world data. First, using the household survey data we estimate cost-of-living indices for Russian regions. Then, we compare these indices to the cost-of-living indices from the Russian Statistical Office.

### 3.1 Household Survey Data

The RLMS-HSE is an annual survey that collects information on socio-demographic characteristics, employment, satisfaction measures, and income sources. The survey is conducted in 32 out of 85 regions, covering 96% of the whole Russian population (Kozyreva et al., 2016).<sup>2</sup> Our dataset includes 16 waves covering 2000–2015. RLMS-HSE is the only household survey that collects information on subjective well-being and, therefore, we have chosen it as the source of evidence.

In order to estimate cost-of-living indices with the matching approach, we need information on income, material needs and welfare level. All of this information is provided by the survey. Income is defined as nominal household income, which is the sum of all private sources of income of every household member, including state transfers, minus taxes. We adjust it by the OECD equivalence scale. Material needs are captured by household and individual characteristics. These include: age and gender of respondents, household size, household composition, number of children, number of pensioners, number of employed household members, type of residence (owned, rented, or dormitory) and place of residence (urban or rural). Welfare is captured by satisfaction with income. We use answers to the question "*How satisfied are you with **your** economic conditions at present?*", on a scale from one to five. The phrasing of this question in Russian aims at capturing a respondent's satisfaction with his living conditions, material well-being and purchasing power, which is not reflected in translation to English. As our approach is based on equalizing the utility assigned to economic well-being, this subjective measure effectively captures the satisfaction of the respondent's material needs and is the best fit for our approach. Table 1 in the Appendix provides summary statistics of the pooled sample.

Given the fact that satisfaction with economic conditions is a subjective measure, we address potential obstacles related to its use (for a detailed discussion, see Bertrand & Mullainathan (2001)). Answers to subjective questions might be shaped by various factors including question ordering, question phrasing, scale design and survey framing. Furthermore, they might be subject to measurement error correlated to observable characteristics of respondents. First, despite the fact that this question includes a five and not commonly used ten-point Likert scale, we stress that this scale is appropriate for our approach as the five-point scale is used in the educational system, and

---

<sup>2</sup>The RLMS-HSE was designed as nationally, not regionally representative survey. We do not claim that the estimated with matching approach indices are the true indices, but we rather illustrate the matching approach with this survey data. This survey data, first, provides the required information, second, it is available for a country with large regional price differences, and, finally, this country provides official cost-of-living estimates for comparative reasons. The estimates based on small samples should be taken with caution. The comparison with the official survey indices shows that despite the regional unrepresentativeness the two indices are similar.

is thus a common scale of evaluation for Russians. Second, the question regarding well-being is simple and straightforward, and it is unlikely that respondents misunderstand or misinterpret it. Third, a question about satisfaction with financial well-being comes after a satisfaction with life question, which is preceded by a general section on employment. Thus, reported answers are not affected by non-related matters (for example, no questions on health or political attitudes are asked before a financial satisfaction question). Finally, Krueger & Schkade (2008) shows evidence for the reliability of subjective measures, especially if the question is part of a repeated sample, sufficiently specific, or if comparison between socio-demographic groups is the purpose of the study.

### **3.2 Federal State Statistics Data**

Since 2009 the Russian Federal Statistical Office (Rosstat) has published cost-of-living indices across cities in Russia (Federal State Statistical Agency of Russia, 2020). These indices show how much more or less expensive the same basket of goods and services is in different regions. The basket consists of 275 items which are consumed by a majority of the local population. This basket includes prices for goods and services, including prices for rent. Every item is weighted according to its share of total consumer expenditures from the yearly household budget survey conducted by Rosstat. Note that the Rosstat does not collect price data in rural areas. Therefore, the cost-of-living indices from the Rosstat reflect price changes in urban areas only.

? convert city-level indices to regional-level in the following way: a city-level index is weighted by its population share in its region and the sum of the weighted city-level indexes in each region is the region-level index. We use these converted to region-level cost-of-living indices prices in this paper. We take Moscow as the reference level, and thus normalize price indices accordingly.

### **3.3 Cost-of-living Estimates**

Using the household survey data, we estimate cost-of-living indices for 38 regions in Russia over the years 2000–2015. We compare these indices with the official indices from the Rosstat over the period 2009–2015. Cost-of-living indices are normalized to the level in Moscow. Cost-of-living estimates are shown in Figure 1 below and in Figure A in the Appendix. Additionally, the correlation between the two indices is shown in Table 2 in the Appendix .

We find that for some regions the estimates are almost identical (see, for example, Chuvashia region, Saratov region, Chelyabinsk region, Krasnoyarsk region), while others are more expensive according to the matching approach (see, for example, Komi region, Leningrad region, Saint-

Petersburg), and some are cheaper (see, for example, Tver region, Penza region, Rostov region). On average, the Rosstat indices are higher than the matching indices, indicating that the regions are more expensive according to the Rosstat approach. However, the two indices converge over the recent years. Importantly, even when the two indices differ, they follow similar trends. This indicates that the conceptual gap between the matching and the Rosstat approaches is constant. Correlation estimates reassures high similarity between the two indices.

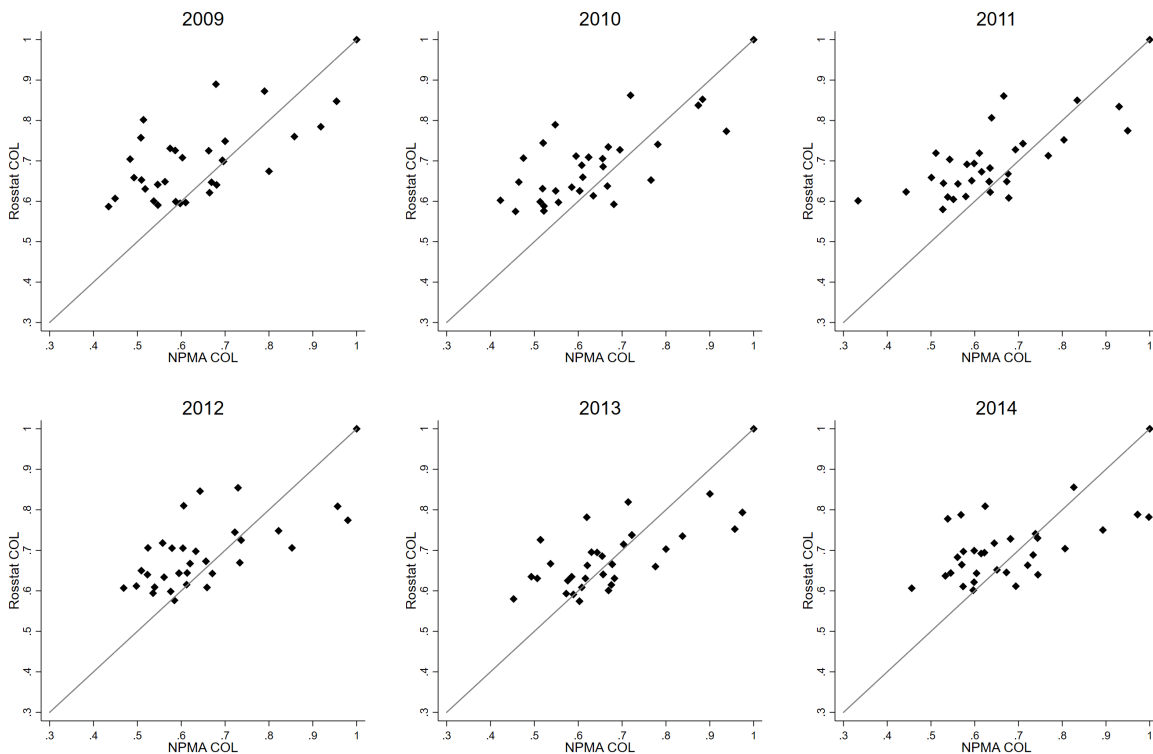
We also analyze cost-of-living indices in relative terms by ranking the regions from the most expensive to the cheapest over the years 2009–2015. The most expensive and the cheapest regions are depicted at the top and the bottom respectively. Over the years 2009–2015, Moscow is the most expensive region according to the two approaches. The majority of the regions remain in their relative positions according to the two approaches. For some regions we document a tendency for the relationship to be reversed: the more expensive regions from the Rosstat data tend to be cheaper in the matching approach, while the cheaper regions from the Rosstat data tend to be more expensive in the matching approach. In particular, all regions from the Far-East Federal district range from average-expensive to very expensive according to the matching approach, while they rank lower according to the Rosstat approach. Some regions from the Central, Southern and Volga Federal districts range from average-expensive to cheaper, while they rank higher according to the Rosstat data.

### **3.4 Implications for Inequality and Poverty**

As cost-of-living indices play an important role in inequality and poverty analysis, we investigate the impact of cost-of-living adjustments on inequality and poverty. Therefore, we compute the Gini index, average income and poverty rate with and without adjustments. Figure 2 shows the estimates under three scenarios: without adjustments (black line), with cost-of-living adjustments using the Rosstat indices (dashed gray line), and with adjustments using the matching indices (solid gray line).

Overall, there was a decrease in income inequality and poverty over the years 2000–2015, but the decrease is even larger once cost-of-living differences are accounted for. Inequality accounting for cost-of-living differences follows similar trends but different levels as inequality not accounting for differences. Income levels differ substantially: an increase from 28,000 Rubles in 2000 to 41,000 Rubles in 2015 compared to an increase from 20,000 Rubles in 2000 to 28,000 Rubles in 2015. We also document that incomes adjusted by the two approaches yield very similar trends in inequality,

Figure 1: Scatterplot of NPMA and Rosstat cost-of-living indices

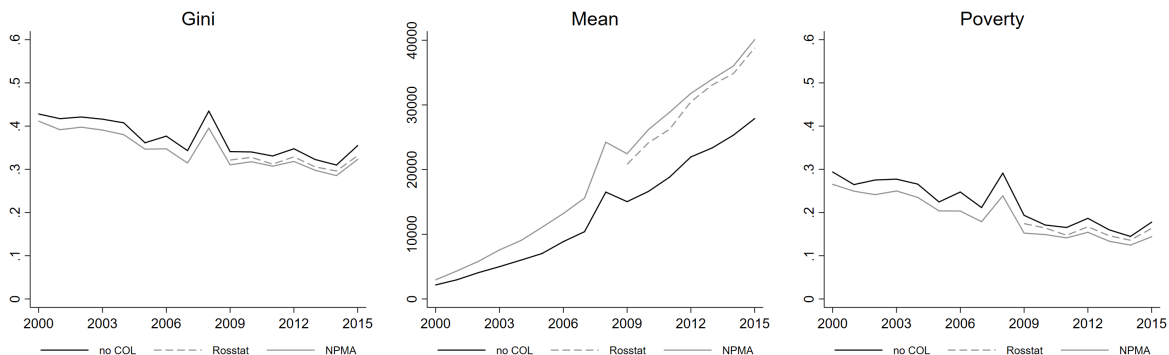


Source: RLMS-HSE (2020); Federal State Statistical Agency of Russia (2020), own calculations.

poverty and average income levels. Therefore, we conclude that inequality and poverty estimates differ in levels, but not in trends when adjusted for cost-of-living differences. More importantly, inequality and poverty estimates adjusted by cost-of-living indices from the Rosstat and the matching approaches show very similar results.

Additionally, we address the question of which drivers decrease inequality and poverty by decomposing inequality into two components: within-region and between-region. The results are given in Figure 2 in the Appendix. We find that accounting for cost-of-living differences decreases inequality between regions. Furthermore, between-region inequality accounting for cost-of-living differences from the matching approach is very close to zero, which means that the indices from the matching approach equalize regions in term of cost-of-living. The larger decrease in inequality is driven by a reduction in between-region inequality.

Figure 2: Inequality and Poverty Trends in Russia



Source: RLMS-HSE (2020); Federal State Statistical Agency of Russia (2020), own calculations. Note: NPMA is non-parametric matching approach estimate. Poverty line is defined as 50% of the average income.

### 3.5 Understanding the differences between NPMA and Rosstat cost-of-living indicators

Last, we address why the cost-of-living indicators from Rosstat and the matching approach are different or similar. In particular, can the differences between the Rosstat and the matching cost-of-living indicators be explained by observed regional characteristics? To answer this question, we run an ordinary least squared regression of the matching index, Rosstat index and the difference between the two on regional characteristics. We include all available regional statistics, which, we believe, are likely to be correlated with the cost-of-living in these regions. These statistics are budget revenue, gross regional product per capita, employment, number of enterprises, export and import statistics, investments, a region’s production, location, area in  $\text{km}^2$ , current population, housing index, and sample size. The results are presented in Table 1 below.

Table 1 shows that the two indices are positively related to investments in fixed capital, a region’s size, and population: the larger the investment, the region’s size and population, the more expensive the region is. The two indices are negatively correlated to the location: the further away the region is from Moscow, the cheaper the region is. At the same time, the export statistics are negatively associated with the matching index and positively with the Rosstat index. The last column of the table shows the relationship between the differences between the two indices and the regional statistics. The difference between the two indices is larger for the more populated regions with higher revenues and smaller for the regions with higher employment and more exports.

The matching approach based on subjective well-being shows that the costs-of-living are higher

in the more populated regions with higher revenues compared with the Rosstat indices. In the matching approach, a more expensive region is directly associated with a greater share of individuals who are dissatisfied with the economic conditions. According to the matching approach, individuals living in more populated regions with higher revenues need more income to feel that their material needs are being met.

In the Russian context, the regions with higher populations and and consequently more revenue are typically those regions with larger cities, better employment opportunities, higher living standards and better social and economic infrastructure. Living in larger regions might, for example, imply higher costs for commuting from home to work, higher costs on drinking clean water, but also higher costs associated with the established lifestyle as regards clothing, gadgets, leisure time etc.

The matching approach also shows that regions with higher employment and higher exports are less expensive in comparison to Rosstat estimates. On average, people are more satisfied with their economic conditions in these regions. Typically, regions with higher exports are also regions with big manufacturing companies, offering plenty of jobs for the local population. In this regard, regional statistics for exports and employment are highly interrelated. Manufacturing companies not only offer employment, they may also invest in building public roads, parks, and recreation centers. In addition, they may support their employees by providing non-monetary goods, including food products or a payed stay at a recreation center. All these elements appear to make citizens in these regions more satisfied with their economic conditions.<sup>3</sup>

---

<sup>3</sup>For example, the largest petrochemical company, Sibur, provides free milk for their employees believing it helps people while working in unhealthy conditions.

Table 1: Pooled regression of NPMA and Rosstat indices and regional statistics, 2009–2015

	NPMA	Rosstat	NPMA-Rosstat
Budget revenue	0.001*** (0.003)	-0.000 (0.158)	0.001*** (0.000)
Gross regional product	0.000 (0.243)	0.000 (0.572)	0.000 (0.385)
Employment	-0.138** (0.027)	-0.016 (0.595)	-0.122* (0.058)
Number of enterprises	0.000 (0.136)	0.000** (0.020)	0.000 (0.738)
Export	-0.003*** (0.000)	0.001*** (0.008)	-0.004*** (0.000)
Import	0.000 (0.906)	-0.001 (0.271)	0.001 (0.520)
Investments in fixed capital	0.000* (0.075)	0.000* (0.055)	0.000 (0.412)
Industrial production index	-1.044 (0.101)	-0.100 (0.748)	-0.944 (0.148)
Location	-21.733*** (0.000)	-16.537*** (0.000)	-5.196 (0.295)
Area in km2	0.000*** (0.000)	0.000*** (0.000)	-0.000 (0.407)
Population	0.062*** (0.008)	0.022* (0.052)	0.039* (0.098)
Housing index	0.098 (0.467)	0.014 (0.833)	0.084 (0.544)
r2	0.660	0.804	0.333
N	224.000	224.000	224.000

\*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

Source: RLMS-HSE (2020); ?; ?, own calculations.

Note: The difference between the cost-of-living indices is scaled by 1,000. Regional statistics are expressed in current prices. Budget revenues include corporate income tax, personal income tax, property tax, and donations. Gross regional product is a market value of all goods and services produced in a region. Employment includes individuals between 15 and 72 years old working at least one hour per week, or being temporarily absent with a reserved working place. Number of enterprises include all officially registered enterprises and companies in a region. Export and import statistics includes market values of all export or imported goods. Fixed-capital investments include construction and re-construction costs, acquisition of equipment and vehicles. Industrial production index is an index of a change in the production in mining, manufacturing, and production and redistribution of gas, electricity and water spheres. Location takes the following values if a region is located in (1) Central, (2) Northwestern, (3) Southern, (4) North Caucasian, (5) Volga, (6) Ural, (7) Siberian, (8) Far Eastern Federal districts. Housing index indicates a price change on the primary housing market.

## 4 Conclusion

Regional price indices are often unavailable, or may not be released as regularly as standard national CPI series. These indices are, however, important to account for differences in cost of

living across regions when income or expenditure data are used for welfare comparisons in large countries. We propose an approach that makes regional cost-of-living deflators available and that do not require access to data on prices and that can be implemented directly from standard household surveys. To do so, we apply methods that have been advocated for the similar problem of estimating equivalence scale parameters: we use subjective satisfaction data to capture financial welfare instead of relying on calculation of indirect utility and adopt non-parametric matching methods to construct regional cost-of-living indices.

The approach is applied to estimation of cost-of-living indices for regions in Russia. This application allows us to compare the matching indices with the official indices from the Russian Statistical Office for the overlapping years. The results show that the two indices are similar, though the matching indices are higher for the more populated and richer regions than the Rosstat indices. The application of the approach to Russian regions, nevertheless, reassures the feasibility of the approach to estimating cost-of-living indicators. Furthermore, the application to inequality analysis proves the importance of acknowledging existing regional differences in prices and their impact on inequality and poverty estimates.

The proposed approach can be applied universally, including to estimating cost-of-living indices within regions, across rural-urban areas, across cities, but also to estimating price indices across countries as well as over time if more traditional consumer price indices are not readily available. This approach also allows researchers to estimate cost-of-living (price) indices for household of different characteristics with respect to education, age, family composition etc. The matching approach only requires household data with information on income (expenditures) and satisfaction with financial conditions.

The paper proposed an approach to estimating cost-of-living indices without data and proves the feasibility of this approach on real data. Any application of the matching approach to real data must, however, recognize its potential limitations. The key underlying assumption of the approach is comparability of individual responses to subjective question about satisfaction with financial conditions. One may argue that satisfaction scores of those living in poorer neighborhoods may not be identical to satisfaction scores of those living in richer neighborhoods. We include individual and household controls in order to mitigate possible bias in the cost-of-living estimates. However, the application to the real data and comparison to estimates based on price data reassures the feasibility of the matching approach.

## References

- Almås, I. (2012). International Income Inequality: Measuring PPP Bias by Estimating Engel Curves for Food. *American Economic Review*, *102*(2), 1093–1117.
- Bellemare, C., Melenberg, B., & van Soest, A. (2002). Semi-parametric Models for Satisfaction with Income. *Portuguese Economic Journal*, *1*(2), 181–203.
- Bertrand, M., & Mullainathan, S. (2001, May). Do People Mean What They Say? Implications for Subjective Survey Data. *American Economic Review*, *91*(2), 67–72.
- Biewen, M., & Juhasz, A. (2017). Direct Estimation of Equivalence scales and more Evidence on Independence of Base. *Oxford Bulletin of Economics and Statistics*, *79*(5), 875–905.
- Diewert, E. (1998). Index Number Issues in the Consumer Price Index. *Journal of Economic Perspectives*, *12*(1), 47–58.
- Federal State Statistical Agency of Russia. (2020). *Statistics on Cost-of-Living in Russian Regions*. Retrieved from <https://gks.ru/folder/13397>
- Kozyreva, P., Kosolapov, M., & Popkin, B. (2016). Data Resource Profile: The Russia Longitudinal Monitoring Survey Higher School of Economics (RLMS-HSE) Phase ii: Monitoring the Economic and Health Situation in Russia, 1994–2013. *International Journal of Epidemiology*, *45*(2), 395–401.
- Krueger, A. B., & Schkade, D. A. (2008, August). The Reliability of Subjective Well-being Measures. *Journal of Public Economics*, *92*(8-9), 1833–1845.
- Melenberg, B., & van Soest, A. (1995). *Semiparametric Estimation of Equivalence Scales using Subjective Information* (Vol. 1995-71; WorkingPaper). Unknown Publisher. (Pagination: 24)
- RLMS-HSE. (2020). *The Russia Longitudinal Monitoring Survey - Higher School of Economics*. Retrieved from [dataretrievedfromhttp://www.cpc.unc.edu/projects/rlms-hse](http://www.cpc.unc.edu/projects/rlms-hse), <http://www.hse.ru/org/hse/rlms>
- Van Praag, B. M. S. (1968). *Individual Welfare Functions and Consumer Behavior: A Theory of Rational Irrationality* (Vol. 57). North-Holland Publishing Company.

## A Appendix

Table 1: Descriptive Statistics of the pooled sample, 2000–2015

	Mean	SD
Age	43.60	18.68
Female	0.57	0.50
Urban	0.73	0.44
Children	0.73	0.91
Employed	0.55	0.50
Retired	0.62	0.78
Family size	3.33	1.64
Family type 1	0.07	0.25
Family type 2	0.08	0.28
Family type 3	0.03	0.16
Family type 4	0.33	0.47
Family type 5	0.03	0.16
Family type 6	0.46	0.50
Home ownership	0.91	0.29
Renting	0.06	0.24
Living in dormitory	0.03	0.16
Income	15202.58	32882.55
Fully satisfied	0.03	0.17
Fully unsatisfied	0.26	0.44

Source: RLMS-HSE (2020), own calculations.

Note: Income is defined as nominal household disposable income adjusted by the OECD equivalence scale. We consider individuals to be employed if (a) they are currently working; or (b) they are on paid leave; or (c) they are on unpaid leave; or (d) they are self-employed; or (e) they are farmers. Household composition: type 1 - single pensioner, type 2 - multiple pensioners, type 3 - single adult without children, type 4 - multiple adults without children, type 5 - single adult with children, type 6 - multiple adults with children.

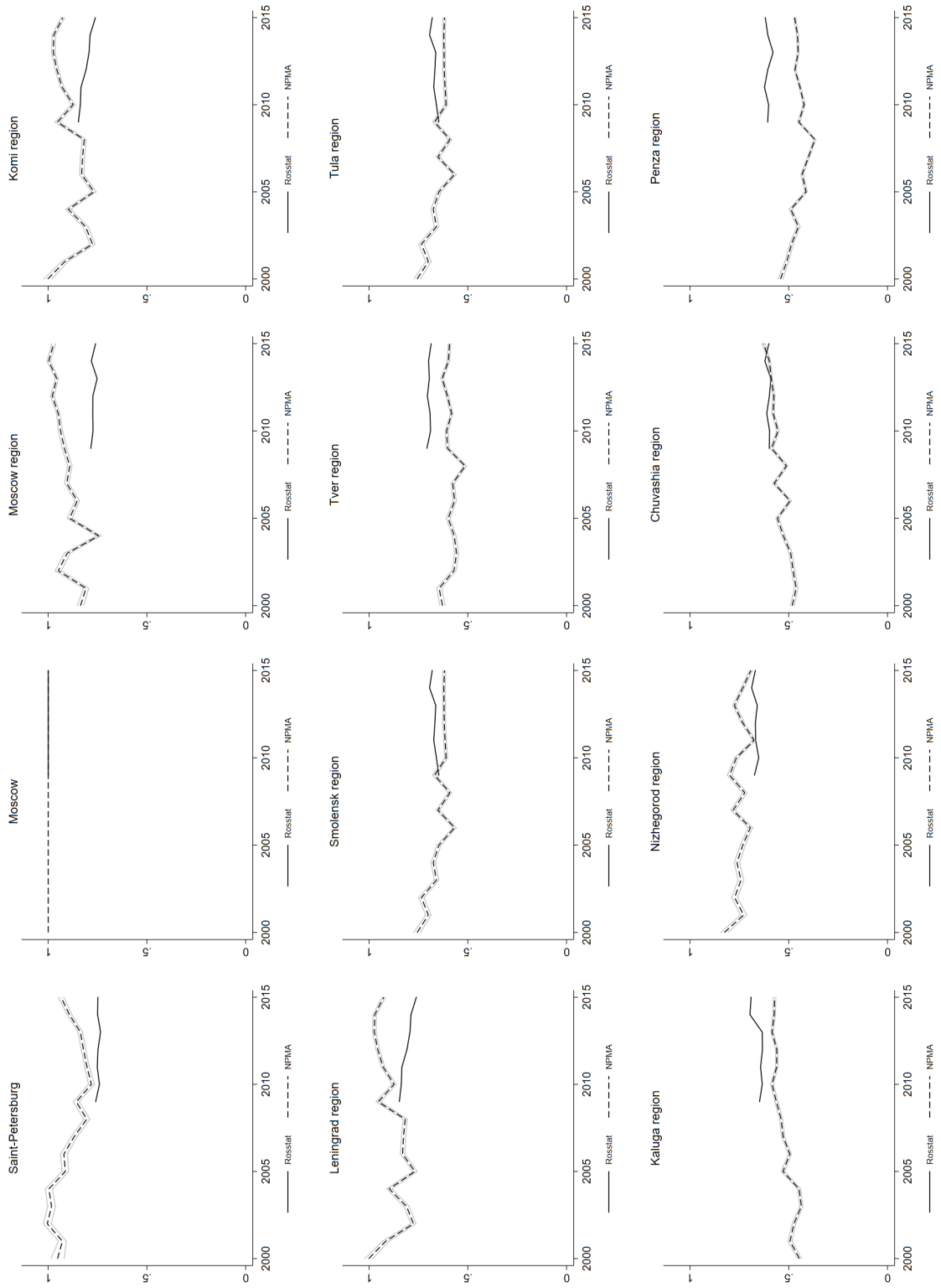
Table 2: Correlation between Rosstat and Non-parametric matching approach cost-of-living indices

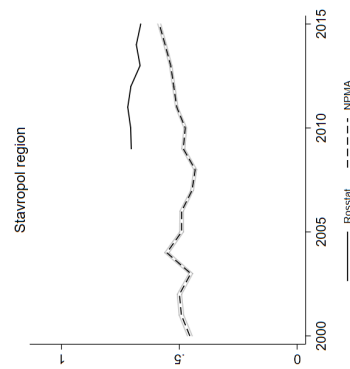
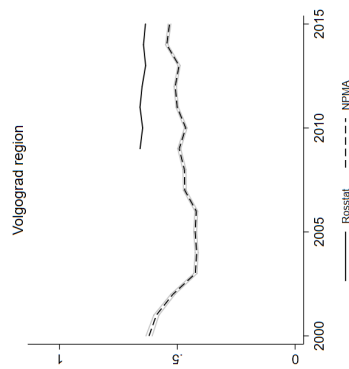
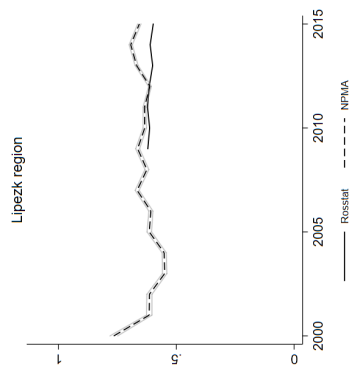
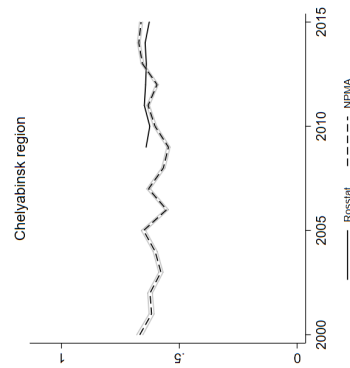
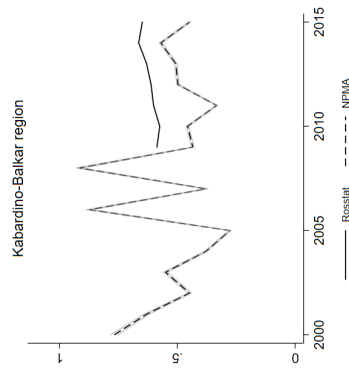
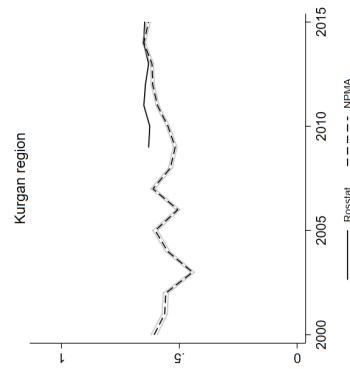
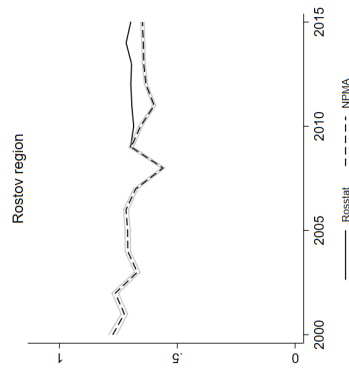
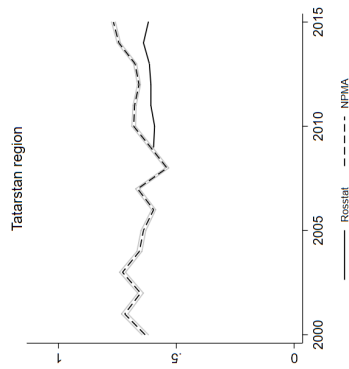
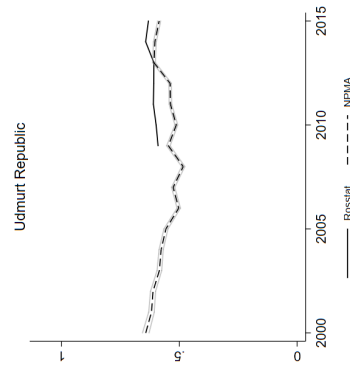
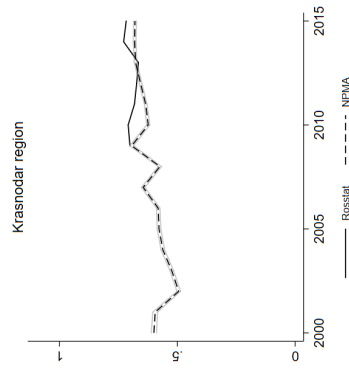
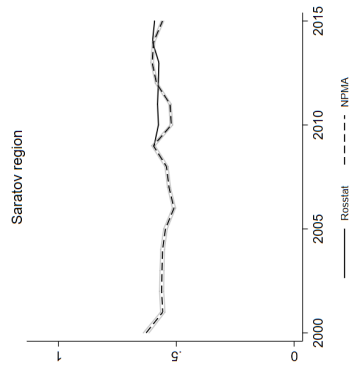
	2009	2010	2011	2012	2013	2014	2015
Spearman correlation	0.580***	0.584***	0.758***	0.727***	0.669***	0.554***	0.563***
Pearson correlation	0.748***	0.684***	0.817***	0.791***	0.780***	0.709***	0.706***

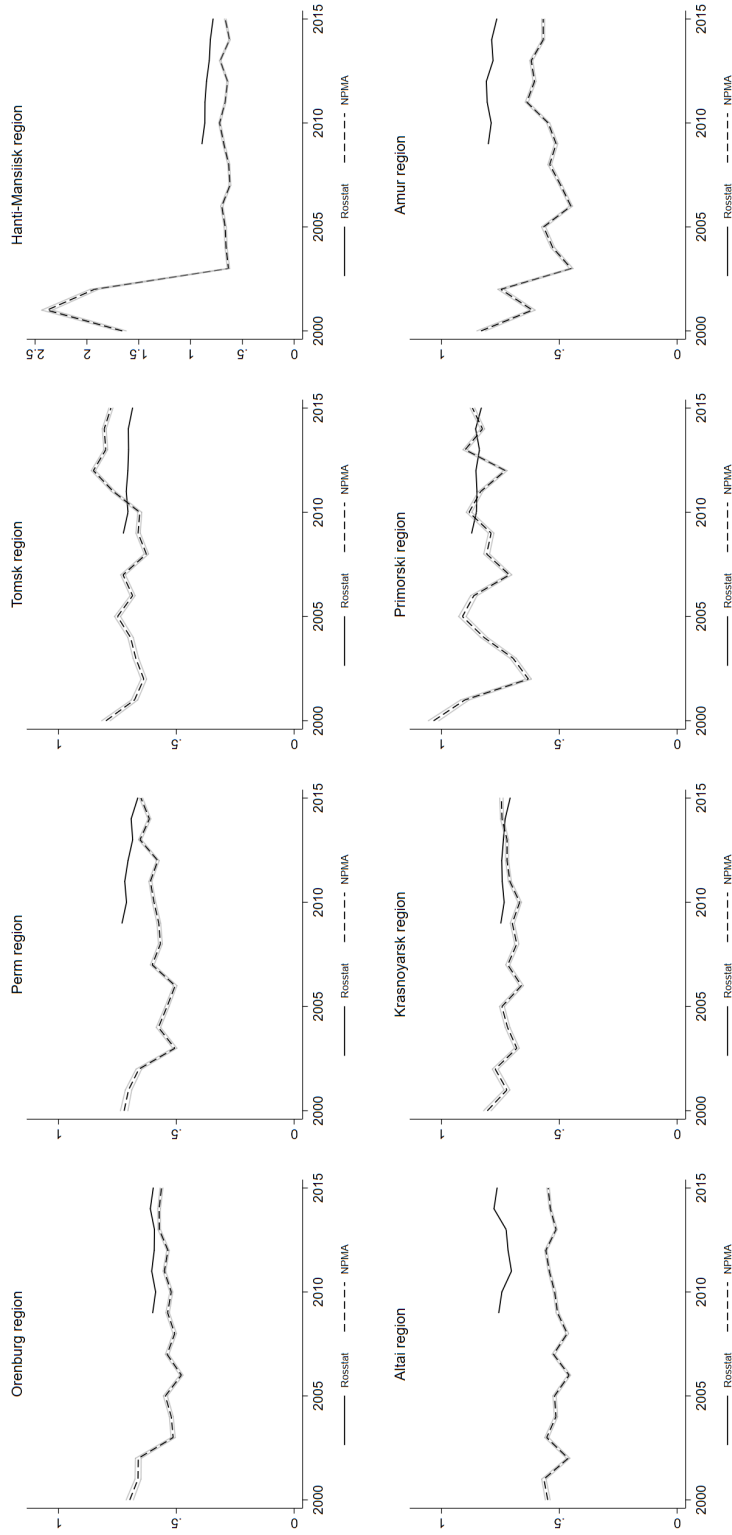
Source: RLMS-HSE (2020); Federal State Statistical Agency of Russia (2020), own calculations.

\*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

Figure 1: Price indices for Russian regions, 2000–2015



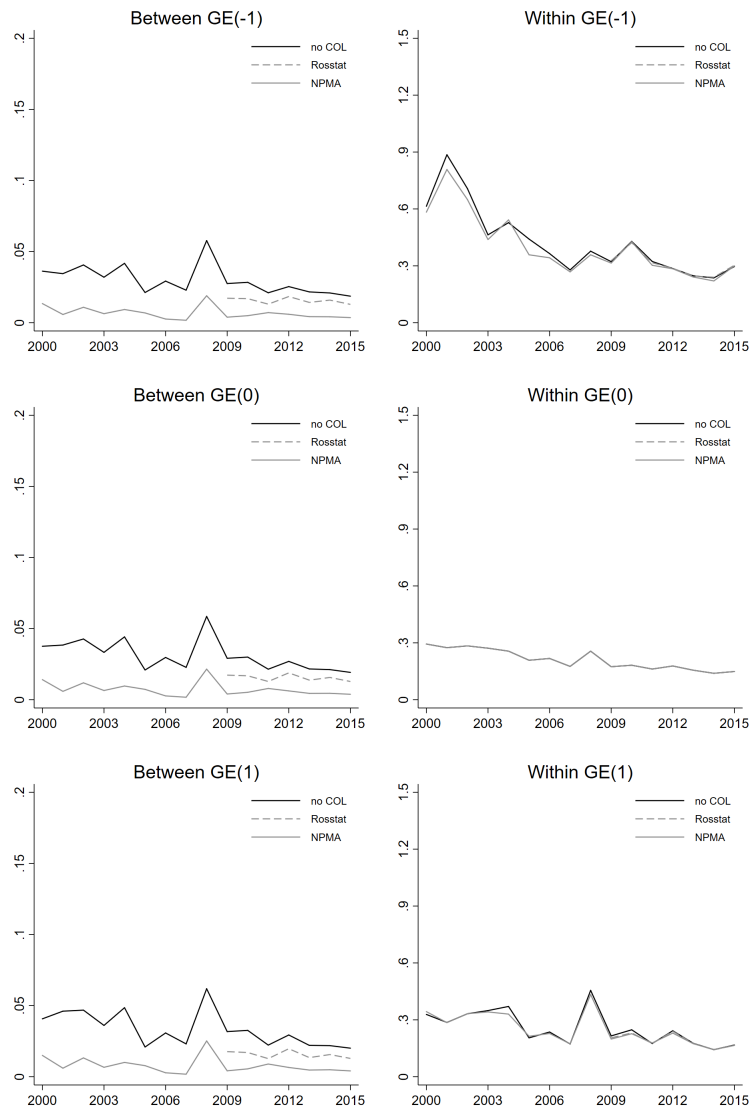




Source: RLM5-HSE (2020); Federal State Statistical Agency of Russia (2020), own calculations.

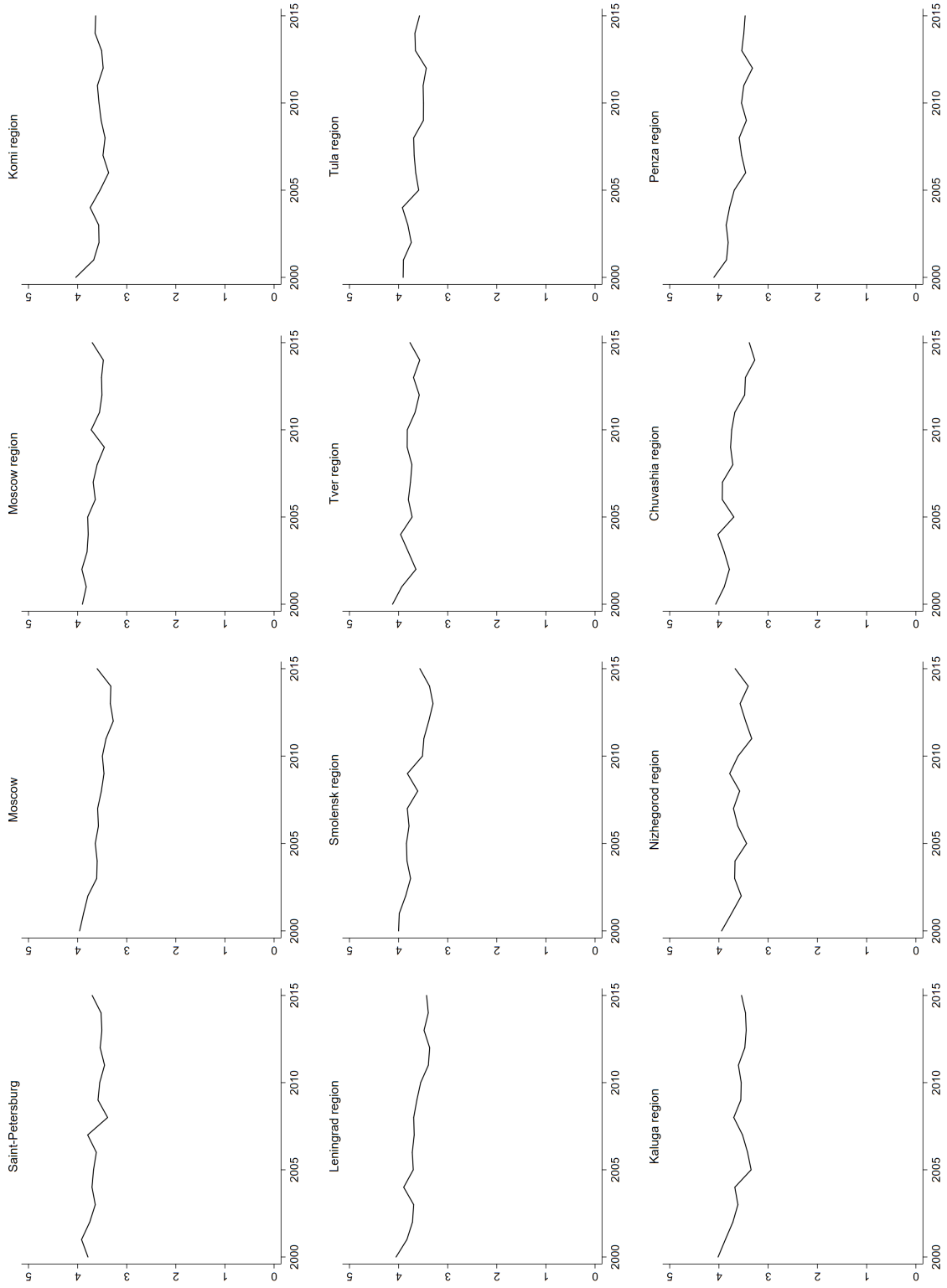
Note: Grey lines are confidence intervals. NPMA is non-parametric matching approach estimate.

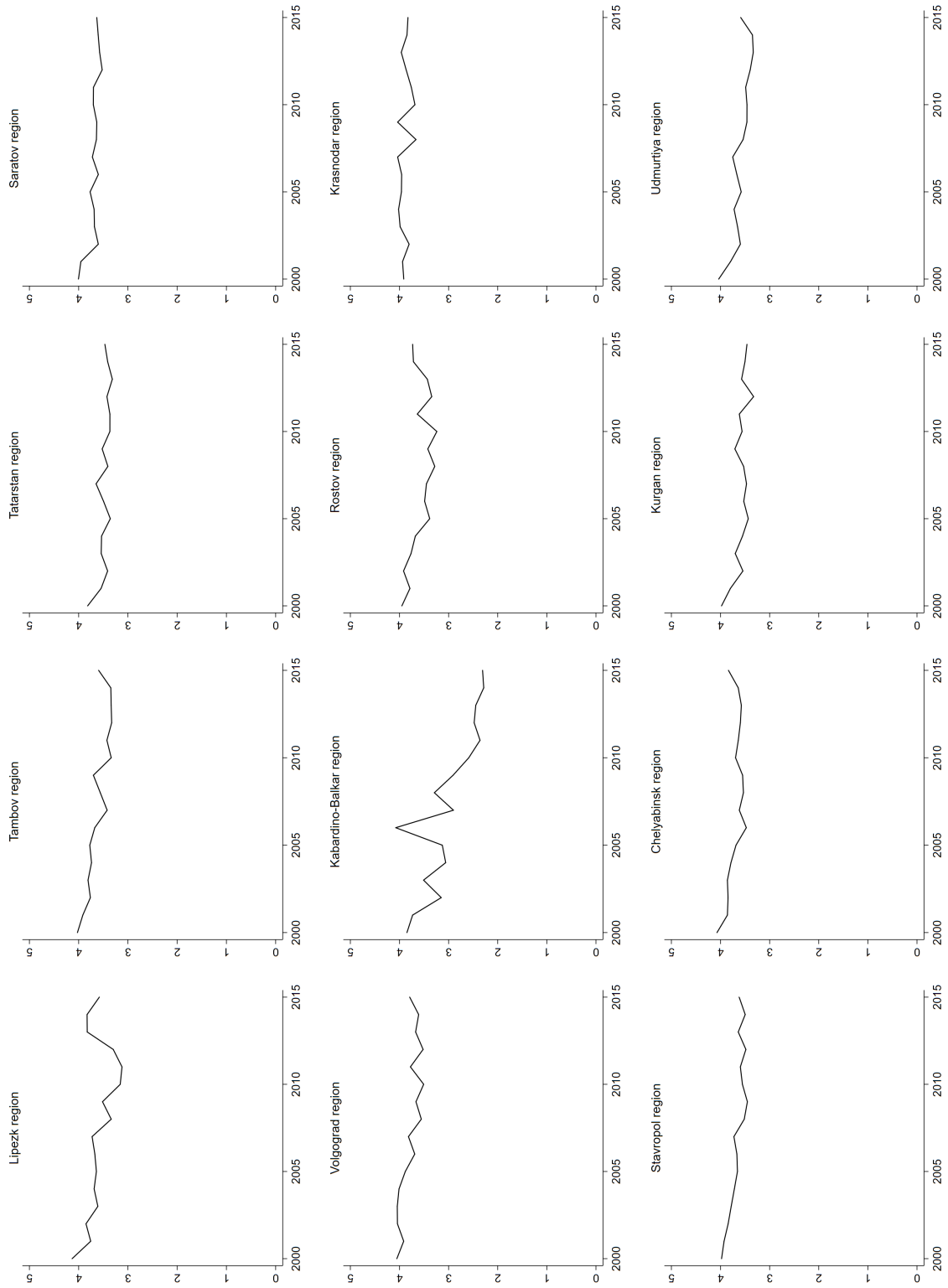
Figure 2: Between- and within-regions decomposition of inequality

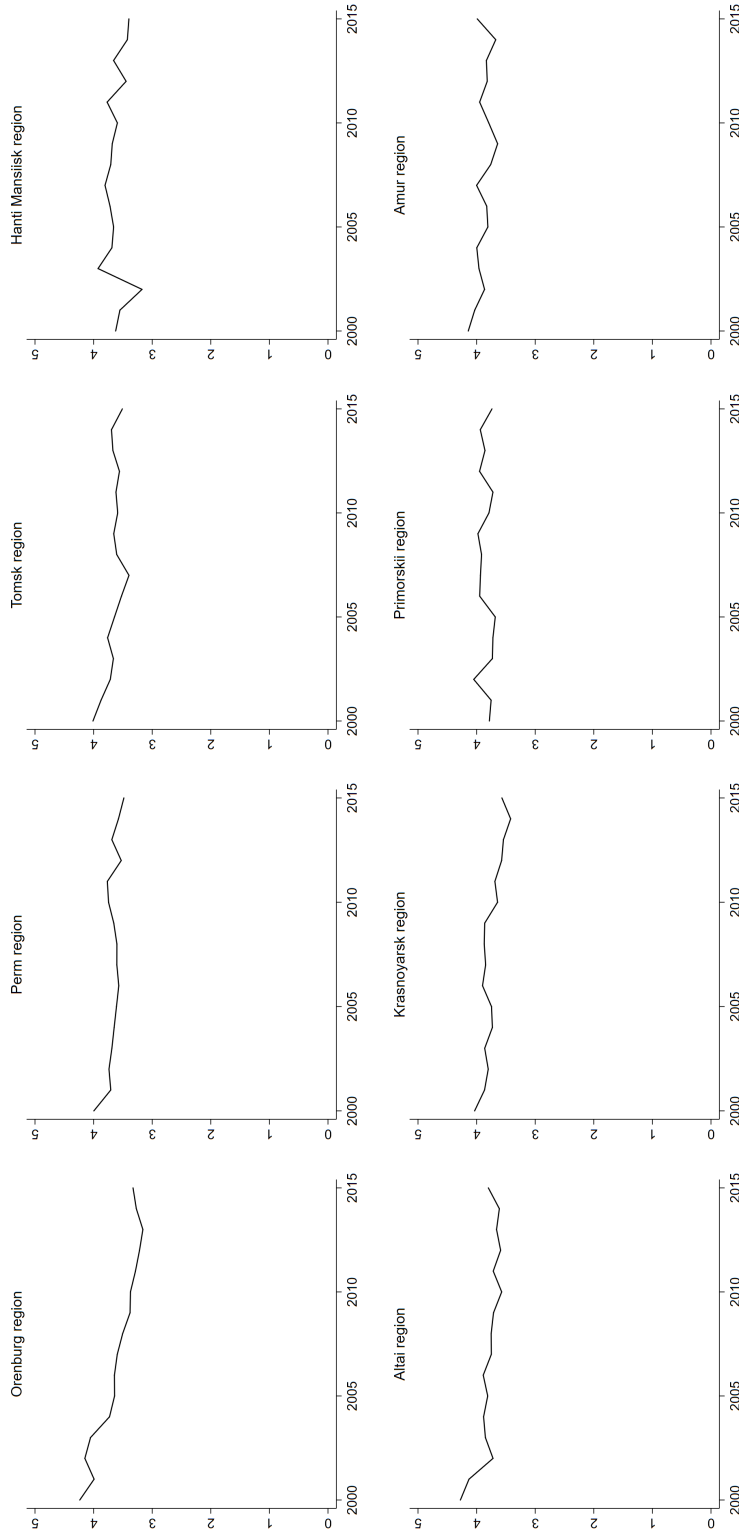


Source: RLMS-HSE (2020); Federal State Statistical Agency of Russia (2020), own calculations. Note:  $GE(\alpha)$  is a General Entropy Family measure. The larger the  $\alpha$  is, the more sensitive  $GE(\alpha)$  to changes in incomes at the top. NPMA is a non-parametric matching approach estimate.

Figure 3: Average satisfaction with economic conditions across Russian regions, 2000–2015







Source: RLMs-HSE (2020); Federal State Statistical Agency of Russia (2020), own calculations.