

**UNITED NATIONS  
ECONOMIC COMMISSION FOR EUROPE  
CONFERENCE OF EUROPEAN STATISTICIANS**

**Expert meeting on measuring poverty and inequality  
29-30 November 2018, Vienna, Austria  
Topic B: Improving response rate and sampling precision in surveys**

**Empirical challenges comparing inequality across countries: the case of middle-income countries from the LIS Database**

Prepared by: Daniele Checchi <sup>a</sup>, Andrej Cupak <sup>b</sup>, Janet Gornick <sup>c</sup>, and Teresa Munzi <sup>d</sup>  
(Preliminary version: 11/10/2018)

**Abstract**

This study presents new empirical results, using microdata from the LIS database, on development patterns in economic inequality for a set of countries that are less covered in the empirical literature, mostly due to the lack of appropriate data. After discussing the main challenges when harmonizing income and consumption microdata from middle-income countries, we focus on Brazil, China, India, Russia and South Africa, in a comparative perspective, and we compare them with a selection of benchmark middle- and high-income countries. We also run country-level regressions to correlate the inequality measures with selected macroeconomic indicators.

**Keywords:** inequality, economic development, survey data, Luxembourg Income Study, middle-income countries

**JEL codes:** D31, D63, O1

---

<sup>a</sup> Corresponding author. LIS Cross-National Data Center in Luxembourg, Esch-Belval, Luxembourg. E-mail: [checchi@lisdatacenter.org](mailto:checchi@lisdatacenter.org).

<sup>b</sup> LIS Cross-National Data Center in Luxembourg, Esch-Belval, Luxembourg. E-mail: [cupak@lisdatacenter.org](mailto:cupak@lisdatacenter.org).

<sup>c</sup> Stone Center on Socio-Economic Inequality, The City University of New York, New York, USA. E-mail: [igornick@gc.cuny.edu](mailto:igornick@gc.cuny.edu).

<sup>d</sup> LIS Cross-National Data Center in Luxembourg, Esch-Belval, Luxembourg. E-mail: [munzi@lisdatacenter.org](mailto:munzi@lisdatacenter.org).

## Table of contents

1 Introduction.....	3
2 The challenges of harmonizing data from middle-income countries.....	4
3. Data and variables.....	9
3.1 Dataset.....	9
3.2 Variables.....	10
4 Methodology .....	10
4.1 Inequality measures.....	10
4.2 Regression analysis .....	11
5 Results.....	11
5.1 Descriptive analysis.....	11
5.2 Regression results.....	14
6 Concluding remarks .....	16
References.....	17
Tables and Figures .....	19
Appendix: Additional tables and figures .....	35

# 1 Introduction

During the past few decades, the focus in research of income inequality has been predominantly on the high-income OECD countries (see, e.g. Atkinson and Brandolini, 2001). More recently, the research has turned to an analysis of top-income shares and its historical development (e.g. Atkinson et al., 2011), again, covering mostly high-income Anglo-Saxon and Nordic countries. Yet, the research analysing income inequality in middle-income and developing countries from the Middle East and North Africa (MENA) region, Asia, Sub-Saharan Africa, and Latin America is still evolving (Assaad et al. 2017).

Some of the recent examples, and perhaps the closest studies to ours in terms of the (partial) focus on the middle-income countries, are studies of Alvaredo and Gasparini (2015) and Lakner and Milanovic (2016).

Lakner and Milanovic (2016) analyse global income inequality (at the individual level) with international panel data (generated as country-level averages obtained from different household surveys) covering the period 1988-2008. The authors show that the global Gini index reached 70.5% in 2008 and report that the income inequality levels were quite stable over the analysed time period. However, according to their analysis, the inequality levels differ for different sub-regions. The lowest income inequality was observed for India with the Gini spanning from 31.1% in 1988 to 33.1% in 2008. Mature economies experienced a growth in the Gini from 38.2% to 41.9% during the same time period. The highest inequality levels were observed in Sub-Saharan Africa increasing from 53.5% in 1993 to 58.3% in 2008. According to the authors, the biggest increase in inequality during the analysed time period was observed in China, with Gini index rising from 32% in 1988 to 42.7% in 2008, overall representing a jump by 33.5%.

Alvaredo and Gasparini (2015) in their exhaustive analysis show that national income inequality for developing countries first increased during the 1980s and 1990s and dropped during the 2000s. The authors conclude that despite caveats related to consistency and comparability of microdata from low-income countries, a tremendous progress has been done in measuring and monitoring income inequality and poverty.

Another example of studies analysing inequality in the middle-income countries is the work of Adams and Page (2003) exploring trends in poverty, inequality and growth in the MENA region countries. More recently, trends in inequality and poverty based on microdata have been explored by Assaad et al., (2016, 2017) for the MENA countries, and by Piketty and Qian (2009) for China and India. Gornick et al. (2009) have been among the first to report inequality trends in the Latin American countries based on the harmonized LIS microdata.

The main goal of this study is to update the existing evidence on income and expenditure/consumption inequality, focusing on a set of middle- and high-income countries from Asia (East and South), MENA region and Latin America. To present the results in a comparative perspective, we also add high-income countries from neighbouring countries to our analysis.

The main contribution of this study is twofold. First, we extract household- and individual-level income and consumption aggregates for which we compute various inequality measures. Then we merge the computed indicators (country-level averages) with macroeconomic characteristics obtained from the World Bank Indicators Database. Our final database covers almost 40 years (from 1976 to 2016) with an unbalanced panel of 20 countries, summing up to 150 observations.

By using descriptive and regression analysis, we aim to uncover possible correlations between recent trends in income and expenditure/consumption inequality measures and compositional population statistics related to educational attainment and gender participation to the labour market. We also explore the different magnitudes correlations between our inequality measures and some institutional indicators capturing countries' economic development. Our empirical analysis updates the findings of Alvaredo and Gasparini (2015) and Lakner and Milanovic (2016) on inequality trends in middle-income countries by analysing more recent years. In the second part, we contribute to the macro-level analysis of Roine et al. (2009) who analyse macroeconomic determinants of economic inequality in a panel of 16 high-income countries.

The paper unfolds as follows. Section 2 discusses the main challenges when harmonizing microdata from developing and middle-income countries. In Section 3 data and variables are presented, while in Section 4 describes the applied methodology. Empirical results are presented and discussed in Section 5. Finally, Section 6 concludes and offers policy implications.

## 2 The challenges of harmonizing data from middle-income countries

From its inception in the 1980s, LIS has been historically focused on high-income countries. A pilot project was carried out in 2007 with the collaboration of a team at the World Bank in order to study the feasibility of including middle-income countries into the LIS Database. Following the decision to go ahead with this expansion, LIS has made some conceptual adjustments and changes to its list of harmonised variables in order to accommodate more diverse labour market characteristics, social benefit structures, consumption patterns, transnational income flows and within-country variability.

Among the main changes achieved with the major template revision which took effect in 2011, the following were mostly aimed at maximising its applicability at the same time to datasets from both high- and middle-income countries:

- Adjustment of the disposable household income (DHI) concept, such that it also includes non-monetary income from labour and from public and private third parties. The main reason for this enlargement of the DHI concept stemmed from the fact that in many middle-income countries the proportion of non-monetary incomes from own consumption and social and/or private assistance-based transfers was too important to be left out, and in fact these amounts are much more often available in middle-income countries than in high-income ones, where many data providers do not even collect them given their irrelevance (see Figure 1**Error! Reference source not found.**, where only the LIS countries that collect non-cash incomes are included).
- Adjustment of the concept of household member to ensure that persons who are physically present in the household but whose incomes do not contribute to the household income (namely live-in domestic servants, boarders and lodgers) are not accounted for in the creation of total household income or the calculation of the equivalence scale.
- Inclusion of a number of living arrangement variables allowing for a better analysis of multi-unit/multi-generation households, so that, if available, information on partnership and parenthood of adults outside the nuclear family is retained.
- Inclusion of variables containing information on an array of new topics, including rural/urban indicator, farming activity indicator, type of dwelling, involvement in marginal/informal work, characteristics of second job.

After ten years of harmonising data from middle-income countries along-side the high-income ones, LIS has acquired some experience with the main challenges which are typically found when dealing with income micro data from these sources, as discussed below.

Because of the diversity of *rural versus urban areas*, in many middle-income countries, income surveys are either only covering urban areas,<sup>1</sup> where it is easier to capture incomes, or use very different instruments (including a different sampling and questionnaires) for the urban and rural areas. This issue proved particularly challenging with the Chinese survey, where the integration of the three different samples (urban, rural and rural to urban migrants) into a unique national sample required some adjustments to the weights and to the variables themselves that risked putting the quality of the resulting file at stake (to the point that for the year 2007, where the issue was particularly severe, LIS decided not to make public the Chinese data).<sup>2</sup>

The definition of *household membership* (and ensuing treatment of individual incomes when creating household level incomes) is of particular relevance in middle-income countries. Live-in domestic servants are much more common, and are either treated as part of the household (but then there is a risk of double-counting the incomes that they receive from their employer – the household), or are not considered as part of the household (and there is the risk that they are simply not represented in the sample). Family members temporarily absent are sometimes treated as household members, sometimes not, and it is often difficult to distinguish the two situations – this is particularly challenging when adults are temporarily absent to work elsewhere (the case of absent heads or spouses) as it considerably changes the way their incomes should be accounted for (include their total income as labour incomes, or only the part that they send back to the family as remittances). In some cases, individual level information is also collected about enlarged family members, and it can become difficult to decide whether they are part of the household or not (again with obvious implications for the treatment of their incomes).

Multigenerational households, and more in general large or *complex households*, are much more common; depending on who is defined as head (older or middle generation in case of three-generational households), the characteristics of the household – often based on the head and its nuclear family – will differ.<sup>3</sup> In case of polygamy, many of the usual indicators typically based on the head and a single spouse become much more difficult to create.

[Figure 1 about here]

When analysing the educational achievement in high-income countries, the information used typically refers to *highest educational qualification* achieved (in terms of highest diploma

---

<sup>1</sup> For example, the Continuous Household Survey (ECH) of Uruguay for years antecedents to 2005 only covers the urban population.

<sup>2</sup> Another example is the Korean data, which is based on two different surveys with different structures and questionnaires. The first one is the Household Income and Expenditure Survey (HIES), covering all non-institutional households residing within the territory of South Korea but excluding 1) farmers households, 2) fishermen households, 3) households whose ordinary incomes and expenditures are difficult to separate from business incomes/expenditure (such as households running restaurants, inns, or boarding houses in their dwellings, and households with two or more live-in-employees), 4) foreigners households. And the second one is Farm Household Income and Expenditure Survey (FHES), which covers the farming households (with some exceptions, e.g. single person households).

<sup>3</sup> So that surveys where the head belongs to the older generation will have a disproportionate amount of elderly households, and vice-versa, with obvious implications also for representing (average) educational attainment, employment condition and so on.

obtained). However, this is not always easy to obtain in middle-income countries, as the information on education typically refers to the highest grade (or level) ever attended rather than the highest level completed; information on literacy is also often relevant; the transformation of this information into highest attained qualification is often very challenging.

When turning to the labour market information, and especially to the *employment definition*, it should be noted that in many surveys of middle-income countries, the labour market module follows the method of the “catch-all” question on employment: after responding that he does not have a job or does not work, an interviewee is asked a series of questions to determine whether he or she has done any activity in order to help the family (such as cultivating fruits and vegetables, selling products in the street, carrying out services for other persons, helping out in a household business, etc.). Given the large extent these activities are performed by women and children in many middle-income countries, considering these activities as ILO employment (under the argument that they fall either under the category of paid work for at least one hour, or as unpaid family work) increases considerably the employment rates, creating large gaps between population with positive earnings and population employed, as well as potentially creating a bias versus those countries that do not include such questions.<sup>4</sup>

A related point concerns child labour. In most of middle-income countries information about labour market is collected for children as well, in order to analyse its diffusion. Some surveys have a special section asked to the children only; in other cases, (some of) the same questions asked to adults are also asked to children. In both cases, the creation of a fully comparable labour market participation rate (or employment rate) between those different countries becomes very tricky. In addition, there are typically many questions on not official work, non-regular activities, household production, illegal labour (not registered, not covered by social insurance, not taxed) in order to capture some measure of the *informal labour*. These questions are typically very different from survey to survey (often referring to the institutional set-up of the country), and are almost impossible to harmonise. As it is extremely difficult to capture and compare the extent of informal employment among the different surveys, one way of proxying it is to calculate employment rates for the elderly (under the assumption that elderly labour is most often related to informal labour).

[Figure 2 about here]

Finally, the wide extent of persons having *multiple jobs* makes the harmonisation (and hence ensuing comparison) of job characteristics challenging especially when the questionnaires ask about different types of work in different sections of the questionnaire (e.g. work in a family business separately from work on the farm separately from other jobs); this is troublesome because: i) it is often not clear if the persons report the same jobs in several sections of the questionnaire, hence incurring in the risk of double counting the jobs; and ii) it becomes hard to determine which is the main job.

---

<sup>4</sup> A very clear example of this arises in Peru (with data from the National Household Survey - ENAHO), which, together with Switzerland, is the country that exhibits the highest employment rate of all LIS countries. Like in the surveys of most other Latin American countries, individuals are first asked if they have a job, and if they respond that they do not, then they are asked if they have carried out any activity to help out the family, and the question includes a long list of possible marginal activities. The very high number of persons who answer negatively to the first question and positively to the second implies that the employment is mostly consisting of marginal employment; it is however very likely that the way the question is formulated invites many people who would have not answered positively with a different question, to answer positively.

[Figure 3 about here]

When looking at the income variables, several issues become at stake when considering middle-income countries. First and foremost, indicators of inequality, poverty and well-being are still prevalently based on *consumption rather than income* data, which often implies that income micro data are either non existing or insufficient for the purpose of calculating robust income indicators (not collected, collected but not provided, collected but not exhaustive to capture the totality of household income).<sup>5</sup>

As already mentioned above, the enlargement to the middle income countries has been followed by a necessary adjustment of the concept of total disposable income to include also *non-monetary incomes* from labour (notably own-consumption of products stemming from farming activities) and public and private transfers (notably public and private assistance as benefits in kind).

[Figure 4 about here]

Whereas the adjustment was necessary in order to get a more unbiased picture of the households' standards of living in those countries, the inclusion of those incomes in the data has often proven to be particularly tricky.

- The very first problem is due to the fact that the coverage of the non-monetary incomes collected by the different surveys widely differs across countries, hence implying a situation where the comparability is at stake. For example, in surveys that are mostly focused on consumption, the value of most goods and services consumed but not paid for (either because own-produced or because received from the employer, the government, charitable institutions or other private households) is collected with great detail and precision, whereas in other types of surveys the availability of those goods becomes much more scarce.

- Another problem arises with the non-monetisation of quantities of goods and services; as of this stage, LIS has taken the approach to only include those incomes that have been monetised by the data provider, thus increasing the potential bias due to the fact that in some countries, for a purely practical rather than conceptual reason, the final income concept includes more non-monetary incomes than in others.

- Somewhat arbitrary assumptions are to be made in case non-monetary incomes are collected in different sections of the questionnaire (among the consumption variables, among the household level incomes from household activities, and among individual level labour incomes); it becomes clear that those amounts will certainly overlap to some extent, and the creation of a final amount that does not include any under- or over-counting of some income sources proves extremely hard to obtain.

Independently from (but related to) the issue of the non-monetary incomes, another problematic area is that of the *self-employment incomes* in general – especially those from farming activities and informal activities. Being those incomes more irregular and difficult to measure by nature, the reliability of a total household income variable which is composed in large part by those types of incomes becomes naturally much more difficult. In addition, when it is collected at the household level only (as is often the case in middle income countries where surveys have specific sections

---

<sup>5</sup> The situation with respect to the availability of complete income data is often unclear until one looks at the microdata themselves, and this has often led to situations where a possibly lengthy process of acquisition of microdata from a new middle income country has ended with the disposal of the data. This has happened for example with older waves of data from Palestine, where income data was existing on paper, but was then not distributed due to the low level of quality, and with older waves of Egyptian data, where income from self-employment was simply not collected.

about the household activities), the creation of a comprehensive measure of total individual labour income becomes impossible, hence restricting the possibility of using such an important variable in many analyses.

[Figure 5 about here]

Other than the measurement itself of the income, its classification into the different income subcomponents can also become more problematic in middle-income countries. One particular issue refers to the classification of *employer-provided pensions and benefits into labour income versus social security*: while benefits provided by the employer (such as allowances and subsidies paid together with the basic wage income) have typically been considered as labour income in high-income countries, when moving to middle income countries it becomes clear that some of those benefits were actually replacing an almost non-existing social security and were thus getting much closer to social security benefits than labour income. Similarly, the usual distinction between *social insurance, assistance and universal benefits* has proven often irrelevant in middle-income countries where the employment-related benefits are purely stemming from the willingness of the employer and not from the benevolence of the government, and most of the strictly speaking public benefits are targeted to the very poor.

The *treatment of taxes and social security contributions* also differentiates middle- from high-income countries. The issue in high-income countries is centred on the difference between the countries/surveys that provide income data gross of taxes and contributions (so-called “gross datasets”) and those that provide the data after such deductions (“net datasets”). More specifically, in the first case all the incomes provided are gross, and the totality of taxes and contributions are deducted from total gross income to obtain the concept of disposable income; on the other hand, for countries that provide each income source already net of taxes and contributions, the sum of all income subcomponents is already net of taxes and contributions, and hence corresponds to the concept of disposable income. As a result, while at the level of total disposable income the variables are perfectly comparable, the comparability is reduced by the fact that at the subcomponent level some datasets provide gross incomes and others net incomes. For middle-income countries the challenge concerning the treatment of taxes is rather different. The very low reliance on direct taxes in most middle-income countries makes the above-mentioned issue almost irrelevant, as the difference between “gross” and “net” datasets is very tiny. It actually becomes problematic to even simply distinguish the surveys in between gross and net, as the situation is either a mixture of the two (in some cases with only wage income being gross of taxes and contributions and all others net), or simply not being defined at all by the data provider itself.<sup>6</sup> Several middle-income countries even provide the income data only in gross terms, without indication of the amount of taxes and contributions paid on them, which results in having to simulate taxes and contributions in order to obtain a measure of disposable income comparable to other countries.<sup>7</sup> In any case, even in the presence of full information on taxes and contributions, the low reliance on direct taxes relatively to the indirect ones in middle-income countries adds a bias to the comparability of well-being indicators based on disposable household income. If

---

<sup>6</sup> This is the case for example in India where, given the almost irrelevance of taxes, the data provider does not specify whereas the incomes should be reported before or after such deductions, so that the end result is a mix of the two depending on the observations, without any indication as to what the situation is for each observation.

<sup>7</sup> This is the case in the National Household Sample Survey (PNAD) of Brazil, the Colombian Great Integrated Household Survey (GEIH) and the Continuous Household Survey (ECH) of Panama.

indirect taxes were also taken into account the true difference between high- and middle-income countries inequality might even be more exacerbated than what the figures show.

[Figure 6 about here]

From a more technical point of view, an issue which can often become serious, especially in data from middle-income countries (but not necessarily confined to those), is the presence of a large amount of observations with *missing (or inconsistent) data*. When the percentage of households with missing (or zero) total disposable income goes beyond a certain threshold,<sup>8</sup> and especially when the data provider does not account for this in the calculation of the weights, the potential bias due to the non-random distribution of those households is large enough to put at risk any country-level analysis of the income distribution.<sup>9</sup>

[Figure 7 about here]

All in all, in spite of the efforts done at the various levels of the data production chain (survey conception, implementation, data editing and data harmonisation), there remain still some important gaps in order to ensure a perfect consistency of the income micro-datasets coming from high- and middle-income countries, and the question of whether those two sets of data can be analysed within the same framework or should be kept more separated remains an important one. LIS has adopted the view that a common framework is possible, but cannot but stress enough the importance of highlighting all the caveats that go with such an approach.

### 3. Data and variables

#### 3.1 Dataset

The present sample of countries is drawn from the Luxembourg Income Study (LIS) Database, the largest available income database of harmonized microdata collected from about 50 countries in Europe, North America, Latin America, Africa, Asia, and Australia spanning over five decades. The biggest advantage of the LIS microdata is the broad coverage of many countries across the world and a large set of standardized variables, making the results directly comparable. LIS datasets contain household- and individual-level data e.g. on the labour income, capital income, social security and private transfers, taxes and contributions, expenditures, as well as employment conditions and usual demographics.

In our empirical analysis we primarily focus on middle-income countries including Brazil, China, India, Mexico, Russia and South Africa. To present the results in a comparative perspective, we consider other middle-income countries (Colombia, Paraguay, Uruguay) and some high-income countries (Chile, Hungary, Israel, Mexico, Poland, Slovenia, South Korea, Taiwan and US) as potential benchmarks. In the regression analysis, we expand the country sample to include other high-income countries (Australia, Austria, Italy and Peru). The choice of the benchmark countries is mostly influenced by the country's sample length, reliability of the survey data, regional diversity, and finally by the country's economic environment with respect to the core countries

<sup>8</sup> LIS typically uses 10% as the threshold requiring some careful treatment, and 20% as the maximum acceptable threshold for reliable income estimates.

<sup>9</sup> For Tunisia for example, LIS obtained data from the only existing income microdata (the Tunisian Labor Market Panel Survey - TLMPS), but after careful consideration, decided not to include it in the LIS Database due to a very large portion of the sample with missing household income (about half of the households).

(i.e. matching GDP per capita, trade openness, financial development, etc.). A Comprehensive overview of our sample countries and years available is presented in Table A.1.

### 3.2 Variables

Our main outcome (household-level) variables are household pre-tax market income (consisting of labour and capital income), household disposable income and household monetary consumption, as well as the total individual-level labour income. Note that values in all outcome variables were top<sup>10</sup> coded and equivalised according to the OECD equivalisation scale. The covariates which are further used in the empirical part include basic information on gender, education, and employment status of individuals. Detailed definition of the employed micro variables is given in Table A.2.

In addition to variables used in the microeconomic inequality analysis, we consider a set of macroeconomic country-level characteristics which have been shown to be significant determinants of economic inequality (see, e.g. Roine et al., 2009, Alvaredo and Gasparini, 2015; Davies et al., 2017). Our considered macroeconomic characteristics include GDP per capita, share of agriculture in the GDP, share of urban households, life expectancy, age dependency ratio, and share of government spending in the GDP. Full list of country characteristics along with their definitions are presented in Table A.3.

## 4 Methodology

In our empirical analysis, we first compute a set of inequality indicators for the income and consumption aggregates, as well as for major population sub-groups. Then, we correlate the estimated inequality measures with country-level macroeconomic indicators which have been discussed in the previous section. Here we acknowledge that in all cases causality may go in both directions, and therefore these are to be considered as nothing more than descriptive statistics. In the next two subsections we briefly summarise our methodological framework.

### 4.1 Inequality measures

In this section, we describe the inequality measures which we apply to the main outcome variables: household-level disposable income and total consumption, and individual-level labour income. For the reader's convenience, we briefly recall the definition of the main inequality indicators we are going to use in the analysis. Following Cowell (2011), let us consider a population of households (individuals), indexed by  $i = 1, \dots, n$  with income (consumption)  $y_i$ ; the arithmetic mean of the income (consumption) for the population is given by  $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$ .

The main inequality indicator which we present through the paper is the Gini index. Having households (individuals) ranked in an ascending order of  $y_i$ , we can write the Gini as follows:

$$Gini = \frac{1}{2n^2\bar{y}} \sum_{i=1}^n \sum_{j=1}^n |y_i - y_j|.$$

A second inequality measure which we apply is the Atkinson index which is given by:

---

<sup>10</sup> In our case data for our main outcome variables were top-coded at 10 times the median of the considered non-equivalised variable.

$$A_\epsilon = 1 - \left[ \frac{1}{n} \sum_{i=1}^n \left[ \frac{y_i}{\bar{y}} \right]^{1-\epsilon} \right]^{\frac{1}{1-\epsilon}},$$

where  $\epsilon$  represents the weighting parameter measuring aversion to inequality. In our case, we compute the Atkinson index for  $\epsilon$  taking values of 0.5, 1, and 2. The higher is the parameter, the stronger is the expression of inequality aversion captured by the index.

Final inequality measure we consider is the Mean Log Deviation index which can be written as:

$$MLD = \frac{1}{n} \sum_{i=1}^n \log \left( \frac{\bar{y}}{y_i} \right)$$

that has the advantage of being exactly decomposable in a between-group and within-group component. In our case, we consider groups created out of gender and three educational categories.

In addition to the three inequality metrics discussed above, we also compute some percentile ratios and income shares. The percentile ratio exhibits the proportion of one income group over the other. Instead of analysing the distribution as a whole, it compares two points of the distribution. In our case we compare the average income of the richest 90% of the households (individuals) to the poorest 10%. Income shares measure gives an overview what share of the total income is held by a certain subpopulation group. In our analysis, we focus on the top 50%, 90%, and 95% of the households (earners).

## 4.2 Regression analysis

In the second-stage of our empirical framework, we run country-level regressions, where we correlate the computed (average) inequality measures to macroeconomic indicators capturing country's economic development. We estimate the relationships by the following linear regression:

$$Inequality_{it} = \beta_0 + \beta \mathbf{X}'_{it} + \delta_i + u_{it},$$

where the left-hand side variable is a specific inequality indicator estimated for country  $i$  in the period  $t$ ,  $\mathbf{X}'_{it}$  presents a vector of country-level characteristics including indicators such as GDP per capita, employment rate, educational attainment in the population, life expectancy, etc., along with the corresponding coefficients  $\beta$  to be estimated. We also control for country fixed effects  $\delta_i$  in the regressions; time fixed effect are more problematic, since the survey years are not coincident among countries (see table A.1 in the Appendix).

# 5 Results

## 5.1 Descriptive analysis

We start our analysis of inequality trends by showing the inequality levels for a set of BRICS middle-income countries (Brazil, China, India, Russia, and South Africa). We also add figures for the US as a benchmark country. To cross-check the picture emerging from the LIS database, we also include inequality measures and income shares from external sources: for inequality measures we collected additional data from the UNU-WIDER WIID Database

(<https://www.wider.unu.edu/project/wiid-world-income-inequality-database>) whereas for income shares, we considered data from the World Inequality Database (<https://wid.world>).

In Figure 9, the first aspect to be considered is that inequality in market incomes is always higher than inequality in disposable incomes, the average difference being 5 percentage points, with consumption inequality (when available) taking an intermediate value (but in India).<sup>11</sup> One might notice that household consumption is covered only for a subset of countries in the LIS Database. In Figure 8 we present a scatter plot between the Gini index of household disposable income and the Gini index of household monetary consumption. We can observe that the link between the two is quite strong with a Pearson's correlation coefficient of about 0.87. This might imply that whenever the information on consumption is not available, we could infer a trend for consumption inequality based on the income inequality data.

[Figure 8 about here]

Considering now inequality trends, from Figure 9 we observe that inequality is on a declining trend in the case of Brazil: for example, the Gini index of household disposable income declines from around 50% in 2006 to around 46% in 2013. A similar, declining trend can also be observed in the figures obtained from the WIID database. Consumption inequality also declines, albeit with smaller magnitude than the disposable income inequality. The decline in inequality in Latin America is a known phenomenon (see Cornia, 2014), which extends to most of Latin American countries available in our dataset (see Figure A.1 in the Appendix). Among the suggested explanations one may reckon a drop in the skill premium following an expansion of secondary education, the adoption of a new development model by a growing number of progressive governments which adopted prudent but more equitable macroeconomic, tax, social assistance and labour policies.<sup>12</sup> These changes raised the incomes of the poor (roughly defined as the bottom half of the distribution), at the expenses of the élites (again roughly identified with the top 5% or 10% in the distribution – see Figure 10).

Going back to Figure 9, for China and India we can hardly talk of any trend as there are only two data points available for each country in the LIS Database. With this caveat in mind, we observe that in the case of China inequality in market income is on the rise, while the (admittedly limited) redistributive activities of the public sector contained this trend, leading to a constant inequality in terms of disposable incomes: in facts, the Gini index of pre-tax market incomes rose from 41% to 46% between 2002 and 2013, while the Gini index of household disposable income spans around 40%.<sup>13</sup> The rising trend in pre-tax income inequality is confirmed in other studies, though it could hide an even higher peak reached around the year 2008 (Jain-Chandra et al. 2018). Income shares observed in Figure 10 suggest that the most of the gains from growth accrued to the richest 10% (though LIS data underestimate this share, due to inability to capture the top incomes, as detectable

<sup>11</sup> The average Gini indices for BRICS countries in LIS are 0.51, 0.46 and 0.47 for gross market incomes, disposable incomes and consumption respectively. The corresponding averages obtained from WIID are 0.49 and 0.44 for disposable incomes and consumption

<sup>12</sup> For a further, detailed discussion on possible causes of falling inequality in Latin America, see Cornia (2014).

<sup>13</sup> In the case of China, we are unable to assess whether inequality level and trend are consistent with other sources, as there is only one reference point available in the external WIID database: notice however that for the available year (2002) the figures are almost identical.

when comparing with WID top incomes data). Contrary to Latin America experience, differences in educational attainment at tertiary level and the skill premium are identified as drivers of the increase in income inequality (Jain-Chandra et al. 2018).

According to LIS data, India experienced a similar increase in market income inequality between 2004 and 2011 with Gini index of pre-tax household income rising from 0.49 to 0.52, while the corresponding figures for disposable incomes rose from 0.48 to 0.49 (confirming that in middle and low-income countries the distinction between the two income concepts is conceptually weak – see Section 2 above). In the case of India, we can only compare the consistency of our results against the WIID database for household consumption, showing that the trends for consumption inequality are identical between LIS and WIID (see again Figure 9). Income shares indicate a robust expansion of the élites, though it is weaker than what is recorded by corresponding WID data for the top 10% (see Figure 10). The rising trend would reverse a declining trend detected in the beginning of previous decade, as a consequence of trade liberalisation undertaken by local governments at the end of previous century (Krishna, and Sethupathy 2012). Over a comparable time interval, we do not find equivalent trend in income inequality for countries available in LIS database, since both Taiwan and South Korea exhibit rather stable inequality trends (see graphs presented in Figure A.2 in the Appendix).

Inequality trends for Russia presented in Figure 9 indicate a declining trend in income inequality, with the Gini index of household disposable income dropping from 41% in 2000 to 33% in 2013.<sup>14</sup> A possible reduction in inequality found in LIS data stands in sharp contrast with results from top incomes analysis: Novokmet et al. (2017) claim that official inequality estimates vastly underestimate the concentration of income in Russia. While income shares of the top 10% exhibits a declining trend in LIS survey data, tax records do indicate an opposite trend.<sup>15</sup> The Russian declining trend in inequality is partly in contrast with other Central and Eastern European countries (Hungary, Poland, and Slovenia), as detectable from Figure A.3.

Finally, we consider income inequality trends for South Africa (see again Figure 9). Among all the countries considered in our sample, income inequality in South Africa, measured by all three indices, is by far the highest. Based on the Gini index of disposable household income, the inequality was as high as 63% in 2008. It gradually dropped to 57% in 2015. As regards consistency with external data sources, we can only compare consumption inequality against the WIID figures. As there are only two points available, we cannot confirm, whether the trends are well captured. Nevertheless, one might notice that consumption inequality has opposite trends between 2008 and 2010 based on LIS and WIID figures. When cross-checking with national sources, the problem of data quality (coverage, weights, imputations) emerges immediately, since various data imputation of missing income values may produce alternative trends, though all of

---

<sup>14</sup> As in the case of India, we can only compare the validity of our computed inequality measures for household consumption: Figure 9**Error! Reference source not found.** shows that consumption inequality computed on LIS data is somewhat lower than corresponding figures from the WIID database.

<sup>15</sup> "...the Gini coefficient jumped from about 0.3-0.4 in self-reported survey data to over 0.6 using the leaked tax data, and the top 10% income share moved from about 30% to over 50% of total income." (Novokmet et al. (2017), pg.13).

them on the rise.<sup>16</sup> This is also confirmed by the trend in income shares of the top 10% from WID, while in the LIS data the corresponding top income share would have lost approximately ten percentage points. In the case of South Africa, there is no other Sub-Saharan country available in the LIS Database. The closest country to compare the trends against is Israel, where ethnic divide is also rather pronounced. Apart from lower level of aggregate inequality, also in the latter country inequality seems declining (see Figure A.4 in the Appendix).

[Figure 9 about here]

[Figure 10 about here]

Overall, we may conclude that inequality trend in BRICS countries exhibits different patterns, with Latin America and Eastern Europe on a declining trend while Eastern Asia and South Africa on the rise. Despite the enormous differences in economic structure among these countries, we make an attempt to investigate whether common causes may underlie these changes. We resort to the common within/between decomposition analysis based on Mean Log Deviation index, and we focus on personal labour earnings, where the identification of common sources of inequality (like gender and education) is easier. In Figure 11 we report the result of such decomposition, where two covariates are considered, gender and education, and consequently the employed population is divided into six groups (2 genders  $\times$  3 educational attainments). The between-group component (namely the inequality that would have been observed had each group member an income equivalent to the group mean) is intended to capture the return to education and/or the gender gap contributions to inequality: in all countries this dimension declines, though from different starting points. In previously centrally planned economies this dimension of inequality was almost non existing, while for the other countries it reached between one third and one half of observed earnings inequality. The within-group inequality (namely the inequality that can be attributed to unobserved components) is on the rise in India and South Africa, while a strong reduction can be observed, especially in China. The limit of this approach is the exclusion of people without labour earnings from the analysis, which makes these countries not strictly comparable. Nevertheless, the decomposition indicates that traditional inequality drivers, gender and education, see a reduction of their explanatory power with reference to income inequality.

[Figure 11 about here]

## 5.2 Regression results

We now turn to a multivariate analysis, but let us first briefly summarize the main micro and macro indicators that enter our regressions either as outcome variables or covariates. Detailed summary statistics are reported in Table 1. We can see that our countries differ in terms of the computed inequality indicators (e.g. Gini index of the household disposable income ranges from 0.20 to

---

<sup>16</sup> “Using Sequential Regression Multiple Imputation (SRMI) to impute values for reported zero or missing incomes, Yu (2009) found a strong increase (seven or eight points) in the Gini coefficient between 1996 and 2001 (Table 3). Supporting evidence comes from other studies employing alternative measures: Leibbrandt et al (2006) found an increase in the Gini from 0.68 to 0.73 using one method, and from 0.74 to 0.79 using another; Simkins (2004) found that the Gini coefficient for households grew from 0.66 to 0.69; and Ardington et al (2005) concluded that the Gini coefficient rose from 0.74 to 0.82. There is thus agreement about the trends, though the levels vary widely.” (Van Der Berg 2010, p.12).

0.63), but also in terms of underlying social conditions (e.g. life expectancy ranges from 53 to 83 years).

[Table 1 about here]

We also explore existing correlations between the considered income inequality measure and a set of institutional variables.<sup>17</sup> We present four scatterplots in Figure 12. Results of these unconditional correlations suggest that income inequality (captured by the Gini index of household disposable income) is negatively correlated with the log of GDP per capita (measure din current US \$), share of public expenditure (% of GDP) and life expectancy (years). On the other hand, age dependency ratio is positively linked with income inequality. This might suggest that in countries with an aging population and limited replacement pension systems, income inequality becomes an issue. We are fully aware that such graphs do not imply any causal relationships and does not consider potential covariance among these variables. To partly cope with these problems, we move to multivariate regressions.

[Figure 12 about here]

The results of our multivariate analysis are presented through Tables 2, 3 and 4. In the first table we consider three outcome variables (factor household income, disposable household income and personal labour earnings), three inequality indicators (Gini index, Atkinson index with alternative risk aversion coefficients –  $\epsilon = 0.5$  and  $\epsilon = 2$  – and alternative income shares). Since we do not control for country fixed effects (though residuals are clustered at country level), these are to be intended as simple conditional correlations, meant to explore the data. Among the most persistent results we notice that compositional variables (employment rates and educational attainment by gender) exhibit significant correlations, though with inconsistent patterns. On the contrary GDP per capita turns out always insignificantly correlated with any inequality measure, similarly to proxies for production composition (agriculture share, trade openness) and public expenditure. Two variables exhibit positive correlation with inequality: one is a proxy for financial development (the market capitalization of listed domestic companies) and the other is a measure of urbanization (though it is positively associated to total income inequality, but negatively associated to labour earnings). However the main limit of this approach is the missing values on some variables (as it can easily be detected in Table 1, where market capitalization and public expenditure are missing in one third of the sample).

[Table 2 about here]

For this reason in Table 3 we adopt a more parsimonious model, in order to raise the number of observations. We also abandon the Atkinson index [with  $\epsilon = 0.5$ ] give its high correlation with Gini index ( $\rho = 0.98$  for household factor income) and we introduce country fixed effects.<sup>18</sup> We

<sup>17</sup> Notice that from now onward we extend the sample of country/year in order to get a more precise estimation of the variance correlations. We are thus working with 21 countries and 150 observations, with an average of 7.4 surveys per country. The countries are (in brackets the number of surveys): Australia (8), Austria (9), Brazil (4), Chile (12), China (2), Colombia (4), Hungary (8), India (2), Israel (11), Italy (12), South Korea (4), Mexico (12), Paraguay (6), Peru (4), Poland (9), Russia (5), Slovenia (6), South Africa (4), Taiwan (11), US (12) and Uruguay (5).

<sup>18</sup> Year fixed effects cannot be included since LIS surveys are available in neighboring years, but not necessarily coincident ones.

currently find that educational attainment in the male population tends to polarise the income distribution, thus raising inequality, while employment rates reduce it. The GDP per capita and the population urban share are now negatively associated to labour market inequality, though still uncorrelated with total income inequality. However, the time span covered by these regressions is rather wide, spanning from 1974 to 2016. Therefore, we have chosen to restrict the sample period to most recent observations, in order to obtain a model more compatible with the BRICS sample (which is only observed after the year 2000).

[Table 3 about here]

In Table 4 we present our preferred model, which contains a limited number of statistically significant coefficients, especially when looking at inequality in factor incomes or in personal labour earnings. However more consistent results emerge when considering household disposable incomes (columns 6 and 7 of Table 4), irrespective of whether we use the Gini index or the Atkinson index with  $\epsilon = 2$  which focusses more on lower values: inequality declines when more women enter the formal labour market (female employment rate), when public expenditure increases and when GDP per capita rises.<sup>19</sup>

[Table 4 about here]

## 6 Concluding remarks

In this paper we have presented recent developments in income and consumption inequality focusing on a set of middle-income countries (Brazil, China, India, Russia and South Africa) that have been recently added to the LIS Database and for which there is a limited coverage and comparability in the empirical literature.

We start by discussing the main challenges when harmonizing income and consumption survey microdata from the middle-income countries and what implications it has for the analysis of economic inequality.

In our empirical exercise, we first estimate a variety of income (consumption) inequality indicators separately for each country and year for the whole population as well as for subpopulation groups. We describe the trends of these five countries against of the trends of neighbouring countries, finding declining trends in inequality in Brazil, Russia and South Africa, against rising trends in East Asia (India and China).

By then merging inequality indicators with World Bank Indicators data, we create an unbalanced panel database covering around 21 countries over the time period from 1976 to 2016. Our panel data analysis updates findings of Roine et al. (2009) who estimate macroeconomic determinants of economic inequality for 15 high-income countries.

---

<sup>19</sup> We have also considered a random-effect model, as well as relaxing the error clustering assumption, without finding more statistical significance in the results. Only life expectancy obtains a significant negative sing, but we acknowledge that causality may go in both directions.

Results from the country-level panel regressions reveal the following patterns. For our analysed sample of countries, the relationship between income inequality and GDP per capita exhibits a negative correlation, jointly with the (female) employment rate. It is also negatively correlated with public expenditure in the GDP, while other controls (like trade openness, share of agriculture, financial openness) come out not significant when country fixed effect are taken into account. None of the country-level correlation results implies a causal relationship. However, the robustness of the results would be reinforced with more countries available in the sample – especially the low-income ones which differ in many economic circumstances from the high-income countries. Therefore, in the future the LIS Datacenter aims to acquire more microdata from low- and middle-income countries, in addition to the already covered ones in the database.

## References

- Adams Jr, R. H., & Page, J. (2003). Poverty, inequality and growth in selected Middle East and North Africa countries, 1980–2000. *World Development*, 31(12), 2027-2048.
- Alvaredo, F., & Gasparini, L. (2015). Recent trends in inequality and poverty in developing countries. In *Handbook of income distribution* (Vol. 2, pp. 697-805). Elsevier.
- Assaad, R., Krafft, C., Nazier, H., Ramadan, R., Vahidmanesh, A., & Zouari, S. (2017). Estimating Poverty and Inequality in the Absence of Consumption Data: An Application to the Middle East and North Africa. In Economic Research Forum Working Papers (No. 1100).
- Assaad, R., Krafft, C., Roemer, J., & Salehi-Isfahani, D. (2016, May). Inequality of opportunity in income and consumption: the Middle East and North Africa Region in comparative perspective. In Economic Research Forum Working Paper Series (No. 1003).
- Atkinson, A. B., & Brandolini, A. (2001). Promise and pitfalls in the use of "secondary" data-sets: Income inequality in OECD countries as a case study. *Journal of Economic Literature*, 39(3), 771-799.
- Atkinson, A. B., Piketty, T., & Saez, E. (2011). Top incomes in the long run of history. *Journal of Economic Literature*, 49(1), 3-71.
- Cornia, G.A. (ed) (2014). *Falling inequality in Latin America. Policy changes and lessons*. Oxford University Press
- Cowell, F. (2011). *Measuring inequality*. Oxford University Press.
- Davies, J. B., Lluberas, R., & Shorrocks, A. F. (2017). Estimating the level and distribution of global wealth, 2000–2014. *Review of Income and Wealth*, 63(4), 731-759.
- Gasparini, L., Cruces, G., Tornarolli, L., & Mejía, D. (2011). Recent trends in income inequality in Latin America [with comments]. *Economia*, 11(2), 147-201.
- Ghosh, Madhusudan (2012) Regional Economic Growth and Inequality in India during the Pre- and Post-reform Periods, *Oxford Development Studies*, 40:2, 190-212,

- Gornick, J.C., Jantti, M., & Leguizamón, A. (2009). Adding Five Latin American Datasets to the Luxembourg Income Study (LIS) Database: Data Assessment and First Comparative Results. Technical Paper No. 3. Luxembourg Income Study Technical Paper Series.
- Jain-Chandra, Sonali, Niny Khor, Rui Mano, Johanna Schauer, Philippe Wingender and Juzhong Zhuang. (2018). Inequality in China – Trends, Drivers and Policy Remedies. IMF Working Paper WP/18/127
- Krishna, Pravin and Guru Sethupathy. 2012. Trade and Income Inequality in India in J.Bhagwati and Panagariya (eds). *India's Reforms: How they Produced Inclusive Growth*, Oxford University Press,
- Lakner, C., & Milanovic, B. (2016). Global Income Distribution, From the Fall of the Berlin Wall to the Global Recession. *World Bank Economic Review*, 30, 203–32.
- Novokmet, F, T Piketty and G and Zucman (2017). From Soviets to Oligarchs: Inequality and Property in Russia, 1905-2016. NBER Working Paper 23712
- Piketty, T., & Qian, N. (2009). Income inequality and progressive income taxation in China and India, 1986-2015. *American Economic Journal: Applied Economics*, 1(2), 53-63.
- Roine, J., Vlachos, J., & Waldenström, D. (2009). The long-run determinants of inequality: What can we learn from top income data?. *Journal of Public Economics*, 93(7-8), 974-988.
- Van der Berg, Servaas. 2010. Current poverty and income distribution in the context of South Africa History. Stellenbosch Economic Working Papers: 22/10.

## Tables and Figures

**Table 1: Descriptive statistics of variables used in the empirical analysis**

Variable	Obs	Countries	Mean	Std.Dev.	Min	Max
Gini index equivalised household factor income	148	21	0.43	0.08	0.27	0.71
Gini index equivalised disposable household income	150	21	0.37	0.10	0.21	0.63
Atkinson index household equivalised factor income e=0.5	148	21	0.17	0.06	0.06	0.43
Atkinson index household equivalised factor income e=1	148	21	0.35	0.09	0.12	0.74
Atkinson index household equivalised factor income e=2	148	21	0.82	0.13	0.31	1.00
Income share bottom 50% (based on household factor income)	144	21	0.19	0.05	0.03	0.30
Income share top 10% (based on household factor income)	147	21	0.33	0.07	0.22	0.60
Income share top 5% (based on household factor income)	147	21	0.21	0.06	0.12	0.43
Income share top 1% (based on household factor income)	147	21	0.07	0.03	0.03	0.17
Gini index personal paid employment income	143	20	0.40	0.08	0.21	0.57
Atkinson index personal incomes e=0.5	143	20	0.14	0.05	0.04	0.27
Atkinson index personal incomes e=1	143	20	0.28	0.10	0.08	0.50
Atkinson index personal incomes e=2	143	20	0.60	0.19	0.18	0.97
MLD personal paid employment income	143	20	0.34	0.15	0.08	0.69
MLD personal paid employment income - between 6 groups(sex & education)	143	20	0.08	0.04	0.02	0.22
MLD personal paid employment income - within 6 groups(sex & education)	143	20	0.26	0.11	0.07	0.53
Decile ratio p90/p10 of personal paid employment income	143	20	9.92	7.18	2.50	32.00
Employment rate among men (16-65)	150	21	0.72	0.09	0.40	1.00
Employment rate among women (16-65)	150	21	0.51	0.12	0.20	1.00
Share female low educ 15-65	150	21	0.25	0.11	0.07	0.47
Share female high educ 15-65	150	21	0.08	0.06	0.01	0.28
Share male low educ 15-65	150	21	0.23	0.10	0.08	0.46
Share male high educ 15-65	150	21	0.07	0.05	0.01	0.19
Age dependency ratio (% of working-age population)	138	20	0.52	0.09	0.36	0.90
Life expectancy at birth (years)	138	20	75.13	4.85	53.72	83.09
log GDP per capita	135	20	9.30	0.95	6.43	10.96
Urban population (% of total)	139	20	0.73	0.13	0.29	0.95
Government expenditure (% of GDP)	129	20	0.29	0.11	0.11	0.51
Government spending on education (% of GDP)	103	19	4.63	0.99	2.25	6.70
Agriculture, forestry, and fishing, value added (% of GDP)	122	20	0.05	0.04	0.01	0.20
Market capitalization of listed domestic companies (% of GDP)	107	19	61.96	49.29	2.08	246.47

Source: LIS Database; World Bank Indicators Database

**Table 2: Regression analysis – OLS including all contextual variables**

Variables	1	2	3	4	5	6	7	8	9	10	11	12
	household factor income (labour + capital)						household disposable income			personal wage income		
	Gini index	Atkinson e=0.5	Atkinson e=2	share bottom50	share top90	share top95	Gini index	Atkinson e=0.5	Atkinson e=2	Gini index	Atkinson e=2	p90p10
Employment rate among men (16-65)	0.13 [0.120]	0.07 [0.106]	-0.40* [0.214]	0.19** [0.070]	0.09 [0.086]	0.09 [0.073]	0.26 [0.180]	0.19 [0.117]	-0.24 [0.504]	0.63*** [0.138]	1.28*** [0.298]	47.16*** [10.954]
Employment rate among women (16-65)	0.06 [0.114]	0.06 [0.101]	0.34** [0.133]	-0.15** [0.061]	-0.01 [0.078]	-0.02 [0.062]	-0.07 [0.120]	-0.04 [0.080]	0.3 [0.386]	-0.09 [0.093]	0.06 [0.227]	2.49 [13.946]
Share female low educ 15-65	-1.22*** [0.302]	-0.68*** [0.230]	0.76 [0.677]	0.41 [0.264]	-0.88*** [0.277]	-0.79*** [0.199]	-0.96 [0.572]	-0.55 [0.351]	-2.51*** [0.835]	-0.67** [0.254]	0.35 [0.588]	-19.8 [45.905]
Share female high educ 15-65	-0.60** [0.236]	-0.43** [0.199]	-0.52 [0.707]	0.2 [0.225]	-0.39* [0.213]	-0.34* [0.174]	-0.29 [0.311]	-0.18 [0.199]	-1.52* [0.855]	-1.05*** [0.282]	-1.82 [1.198]	-85.09** [37.238]
Share male low educ 15-65	1.25*** [0.290]	0.80*** [0.234]	-0.32 [0.742]	-0.34 [0.309]	0.82*** [0.275]	0.75*** [0.210]	0.97* [0.545]	0.60* [0.336]	2.40*** [0.658]	0.67*** [0.191]	0.29 [0.773]	28.92 [43.557]
Share male high educ 15-65	0.01 [0.330]	0.06 [0.243]	0.64 [0.662]	0.29 [0.291]	-0.13 [0.332]	-0.12 [0.286]	-0.01 [0.390]	0 [0.252]	0.99 [0.700]	1.71** [0.623]	4.42* [2.226]	145.17* [70.498]
Age dependency ratio (% of working-age population)	0.27** [0.098]	0.14* [0.079]	-0.31* [0.151]	-0.19*** [0.063]	0.16* [0.076]	0.1 [0.059]	0.26* [0.137]	0.13 [0.093]	0.43 [0.363]	0.30* [0.155]	0.3 [0.465]	29.95 [20.890]
Life expectancy at birth (years)	-0.01*** [0.003]	-0.01** [0.003]	0 [0.004]	0.00** [0.002]	-0.01*** [0.003]	-0.01*** [0.002]	-0.01* [0.005]	-0.01* [0.003]	0 [0.010]	-0.01*** [0.003]	-0.01 [0.006]	-0.53* [0.278]
log GDP per capita	0.01 [0.018]	0.01 [0.016]	0.09** [0.033]	0 [0.012]	0 [0.014]	0 [0.013]	-0.02 [0.023]	-0.02 [0.016]	0.02 [0.065]	0.01 [0.026]	0.04 [0.067]	3.24 [3.560]
Urban population (% of total)	0.15** [0.064]	0.07 [0.054]	-0.17 [0.116]	-0.12*** [0.037]	0.17** [0.062]	0.15** [0.060]	0.17* [0.082]	0.09 [0.062]	-0.32* [0.184]	-0.09 [0.096]	-0.71*** [0.192]	-32.75*** [8.638]
Government expenditure (% of GDP)	-0.08 [0.067]	0.03 [0.049]	0.06 [0.207]	0.08 [0.050]	-0.07 [0.053]	-0.06 [0.044]	-0.11 [0.133]	-0.03 [0.086]	-0.14 [0.255]	-0.21** [0.093]	-0.40* [0.218]	-8.21 [13.887]
Government spending on education (% of GDP)	0 [0.010]	0 [0.008]	-0.04** [0.013]	-0.01 [0.008]	0.01 [0.008]	0 [0.006]	0 [0.015]	0 [0.009]	-0.09* [0.047]	0.03*** [0.006]	0.06** [0.023]	2.73*** [0.752]
Agriculture, forestry, and fishing, value added (% of GDP)	0.2 [0.314]	-0.07 [0.268]	-0.02 [0.644]	-0.37* [0.181]	0.22 [0.293]	0.23 [0.284]	-0.06 [0.522]	-0.2 [0.362]	-0.6 [1.155]	0.07 [0.617]	-1.23 [1.573]	19.6 [75.907]
Trade openness (% of GDP)	0.02 [0.028]	0.01 [0.026]	-0.05 [0.047]	0 [0.019]	0.01 [0.024]	0.01 [0.024]	0 [0.039]	0 [0.027]	-0.27*** [0.084]	0.05 [0.047]	0.16 [0.094]	6.82 [5.833]
Market capitalization of listed domestic companies (% of GDP)	0.06*** [0.014]	0.05*** [0.013]	0.01 [0.030]	-0.02*** [0.007]	0.05*** [0.012]	0.04*** [0.012]	0.08*** [0.019]	0.05*** [0.014]	0.16** [0.057]	0.09*** [0.025]	0.14** [0.056]	7.28** [2.604]
Observations	75	75	75	74	75	74	77	77	77	71	71	71
R-squared	0.85	0.793	0.515	0.795	0.88	0.872	0.857	0.848	0.556	0.831	0.669	0.632

Note: Robust standard errors in brackets. Standard errors clustered at country level. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1.

Source: LIS Database; World Bank Indicators

**Table 3: Regression analysis – country fixed effects selecting some contextual variables**

Variables	1	2	3	4	5	6	7	8	9	10
	household factor income (labour + capital)					household disposable income		personal wage income		
	Gini index	Atkinson e=2	share bottom50	share top90	share top95	Gini index	Atkinson e=2	Gini index	Atkinson e=2	p90p10
Employment rate among men (16-65)	-0.26** [0.113]	0.12 [0.138]	0.18** [0.076]	-0.19* [0.101]	-0.11 [0.080]	-0.13 [0.095]	-0.71** [0.325]	-0.34** [0.122]	-0.66 [0.410]	3.72 [15.151]
Employment rate among women (16-65)	0.12 [0.087]	-0.40*** [0.129]	-0.04 [0.069]	0.06 [0.082]	0.02 [0.072]	0 [0.079]	0.26 [0.220]	0.28** [0.120]	0.44 [0.400]	-1.84 [16.612]
Share female low educ 15-65	0.14 [0.144]	-0.48 [0.426]	-0.12 [0.101]	0.14 [0.091]	0.19** [0.085]	0.02 [0.114]	-0.86 [0.552]	0.13 [0.094]	-0.13 [0.267]	-15 [17.744]
Share female high educ 15-65	-0.08 [0.212]	-0.44 [0.639]	0.32 [0.238]	-0.11 [0.213]	-0.1 [0.150]	0.03 [0.209]	-0.04 [0.777]	-0.23 [0.160]	-0.42 [0.778]	-92.87*** [31.871]
Share male low educ 15-65	0.23** [0.102]	-0.13 [0.161]	-0.03 [0.108]	0.18 [0.108]	0.16* [0.087]	0.04 [0.133]	0.59** [0.265]	0.08 [0.094]	-0.42 [0.332]	-22.99 [19.263]
Share male high educ 15-65	0.97*** [0.177]	0.99 [1.045]	-0.63*** [0.185]	0.84*** [0.175]	0.70*** [0.160]	0.4 [0.234]	3.25** [1.278]	1.09*** [0.340]	3.33*** [0.980]	136.31*** [40.248]
Age dependency ratio (% of working-age population)	0.05 [0.112]	-0.31** [0.150]	-0.09 [0.122]	0.02 [0.112]	0.06 [0.084]	0.03 [0.116]	0.34 [0.417]	-0.18** [0.077]	-0.63*** [0.158]	-33.59** [12.440]
Life expectancy at birth (years)	0.01 [0.005]	0.02** [0.006]	0 [0.004]	0 [0.005]	-0.01 [0.004]	0 [0.004]	0.01 [0.009]	0.01** [0.004]	0.02** [0.008]	1.62** [0.720]
log GDP per capita	-0.02 [0.016]	-0.01 [0.025]	0.01 [0.012]	-0.02 [0.015]	-0.01 [0.014]	-0.02 [0.014]	-0.09** [0.033]	-0.05** [0.017]	-0.13*** [0.039]	-4.57** [1.987]
Urban population (% of total)	0.08 [0.328]	-0.4 [0.659]	-0.4 [0.311]	0.37 [0.305]	0.44** [0.209]	0.06 [0.305]	-0.12 [1.072]	-0.98** [0.374]	-3.15*** [0.809]	188.85*** [61.310]
Government expenditure (% of GDP)	-0.19* [0.098]	-0.2 [0.220]	-0.03 [0.076]	-0.15 [0.093]	-0.11 [0.075]	-0.12 [0.104]	-0.51 [0.313]	-0.12 [0.125]	-0.07 [0.335]	3.18 [15.923]
Observations	127	127	123	126	126	129	129	122	122	122
R-squared	0.289	0.271	0.326	0.234	0.216	0.179	0.331	0.373	0.332	0.311
Number of countries	21	21	21	21	21	21	21	20	20	20

Note: Robust standard errors in brackets. Standard errors clustered at country level. Country fixed effects included. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1.

Source: LIS Database; World Bank Indicators Database

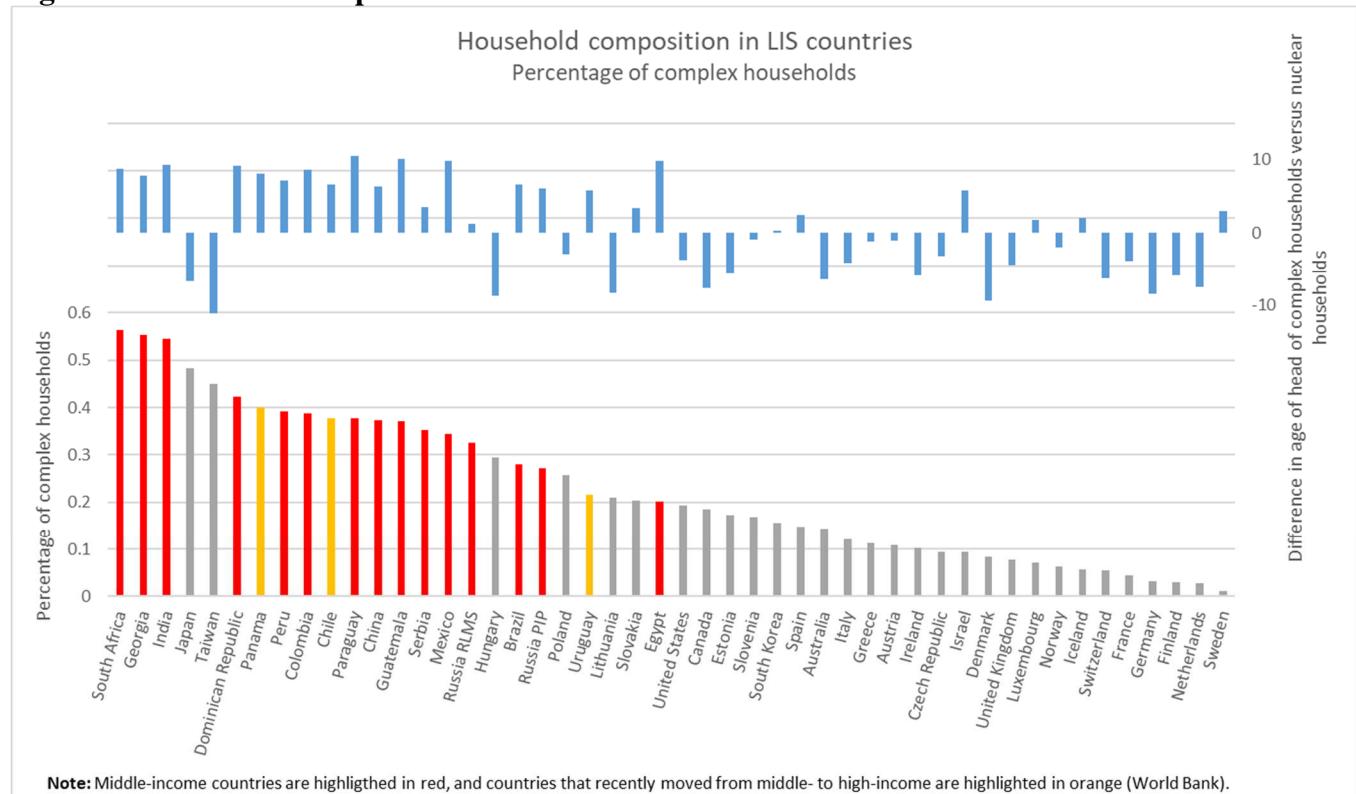
**Table 4: Regression analysis – country fixed effects selecting some contextual variables - most recent observations (year>2000)**

Variables	1	2	3	4	5	6	7	8	9	10
	household factor income (labour + capital)					household disposable income		personal wage income		
	Gini index	Atkinson e=2	share bottom50	share top90	share top95	Gini index	Atkinson e=2	Gini index	Atkinson e=2	p90p10
Employment rate among men (16-65)	0.11 [0.170]	0.54 [0.348]	0.07 [0.114]	0.1 [0.143]	0.13 [0.145]	0.16 [0.108]	0.37 [0.389]	-0.06 [0.146]	-0.42 [0.508]	2.05 [26.774]
Employment rate among women (16-65)	-0.25 [0.146]	-0.79* [0.388]	0.12 [0.115]	-0.21 [0.133]	-0.22 [0.127]	-0.24** [0.092]	-0.78** [0.365]	0.05 [0.166]	0.07 [0.629]	9.72 [32.745]
Share female low educ 15-65	-0.29 [0.565]	-2.05 [1.984]	-0.04 [0.421]	-0.11 [0.502]	-0.2 [0.474]	-0.62 [0.413]	-5.16*** [1.798]	-0.08 [0.488]	-0.68 [2.336]	-29.26 [83.711]
Share female high educ 15-65	0.07 [0.400]	0.47 [1.279]	-0.07 [0.286]	0.19 [0.292]	0.14 [0.227]	0.11 [0.261]	-0.8 [1.124]	-0.22 [0.731]	-1.59 [1.956]	-101 [130.952]
Share male low educ 15-65	0.11 [0.527]	0.11 [1.345]	0.14 [0.357]	0.18 [0.513]	0.31 [0.500]	0.43 [0.434]	1.66 [1.175]	-0.33 [0.450]	-1.65 [1.987]	-96.12 [66.203]
Share male high educ 15-65	-0.31 [0.433]	-1.45 [1.039]	0.33 [0.201]	0.08 [0.332]	0.22 [0.289]	0.02 [0.347]	1.9 [1.269]	-0.47 [0.695]	0.33 [2.645]	-99.34 [100.619]
Age dependency ratio (% of working-age population)	0.13 [0.189]	-0.86* [0.428]	-0.07 [0.140]	-0.04 [0.147]	-0.12 [0.116]	0.02 [0.151]	0 [1.007]	0.4 [0.267]	1.31* [0.689]	66.45 [50.825]
Life expectancy at birth (years)	0 [0.005]	0.01 [0.008]	0 [0.004]	-0.01 [0.005]	0 [0.004]	0 [0.003]	0 [0.008]	0 [0.005]	0.01 [0.014]	1.04 [0.936]
log GDP per capita	-0.01 [0.014]	-0.07*** [0.024]	0.01 [0.008]	-0.01 [0.013]	-0.01 [0.011]	-0.02* [0.010]	-0.17*** [0.049]	-0.01 [0.020]	-0.02 [0.050]	1.35 [2.713]
Urban population (% of total)	-0.08 [0.641]	-0.62 [1.211]	-0.03 [0.393]	0.11 [0.499]	-0.07 [0.429]	-0.39 [0.481]	-0.72 [1.005]	-0.08 [0.662]	-0.32 [1.734]	-137.18 [105.797]
Government expenditure (% of GDP)	-0.09 [0.095]	-0.52 [0.405]	-0.04 [0.092]	-0.05 [0.080]	-0.06 [0.065]	-0.17** [0.074]	-0.89** [0.335]	-0.1 [0.157]	-0.72* [0.391]	-2.74 [21.288]
Observations	83	83	81	83	82	83	83	79	79	79
R-squared	0.306	0.243	0.354	0.359	0.287	0.428	0.346	0.196	0.15	0.198
Number of countries	21	21	21	21	21	21	21	20	20	20

Note: Robust standard errors in brackets. Standard errors clustered at country level. Country fixed effects included. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1.

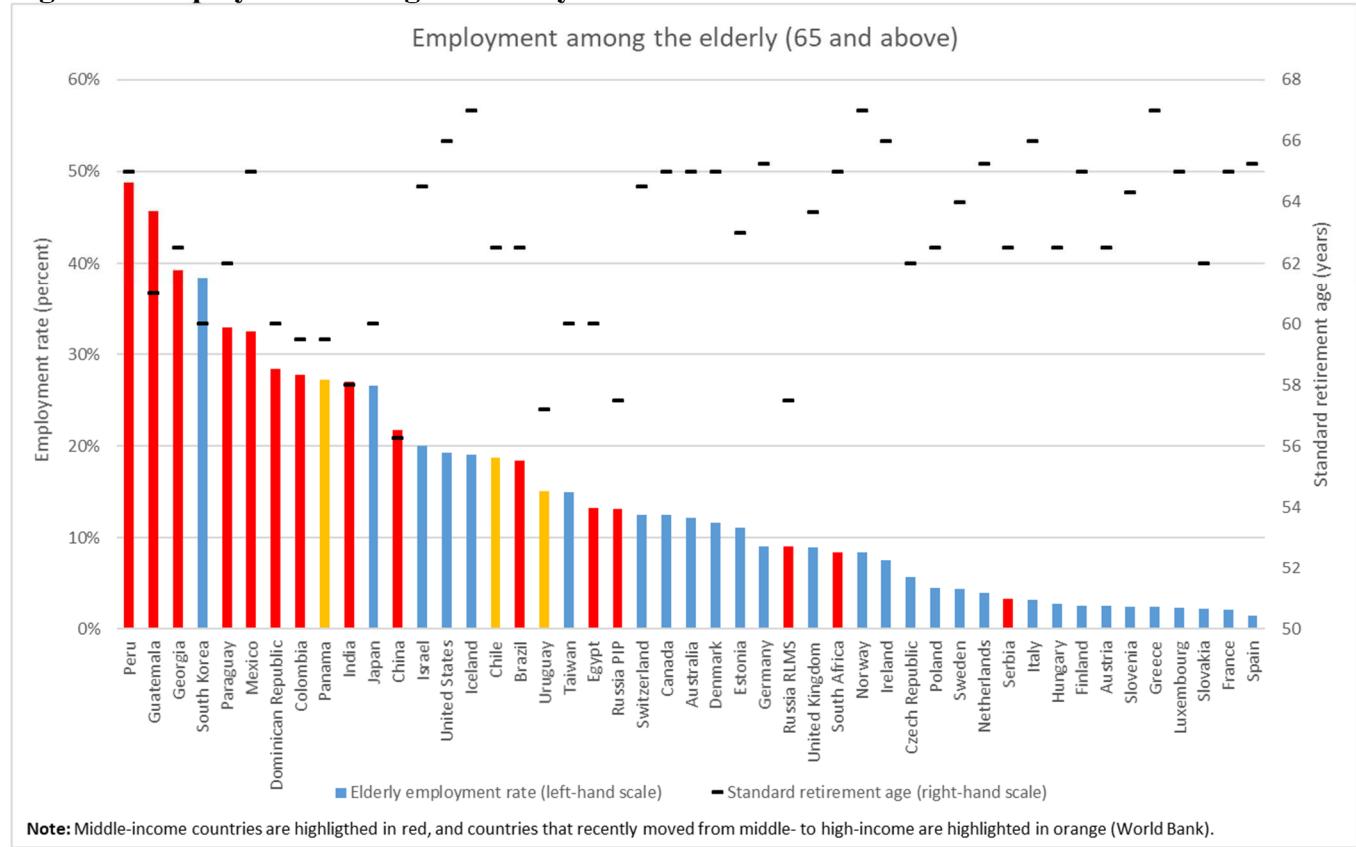
Source: LIS Database; World Bank Indicators Database

**Figure 1: Household composition in LIS countries**



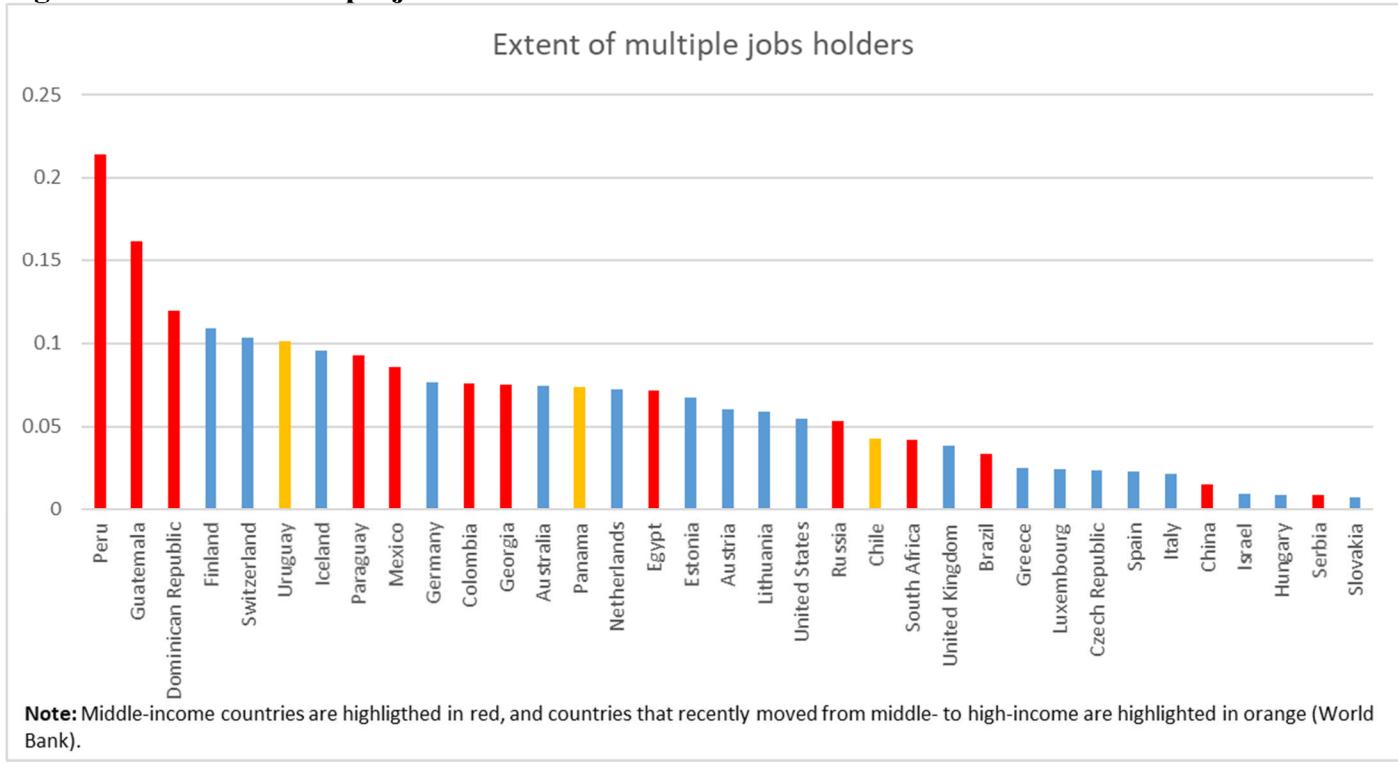
Source: LIS Database

**Figure 2: Employment among the elderly**



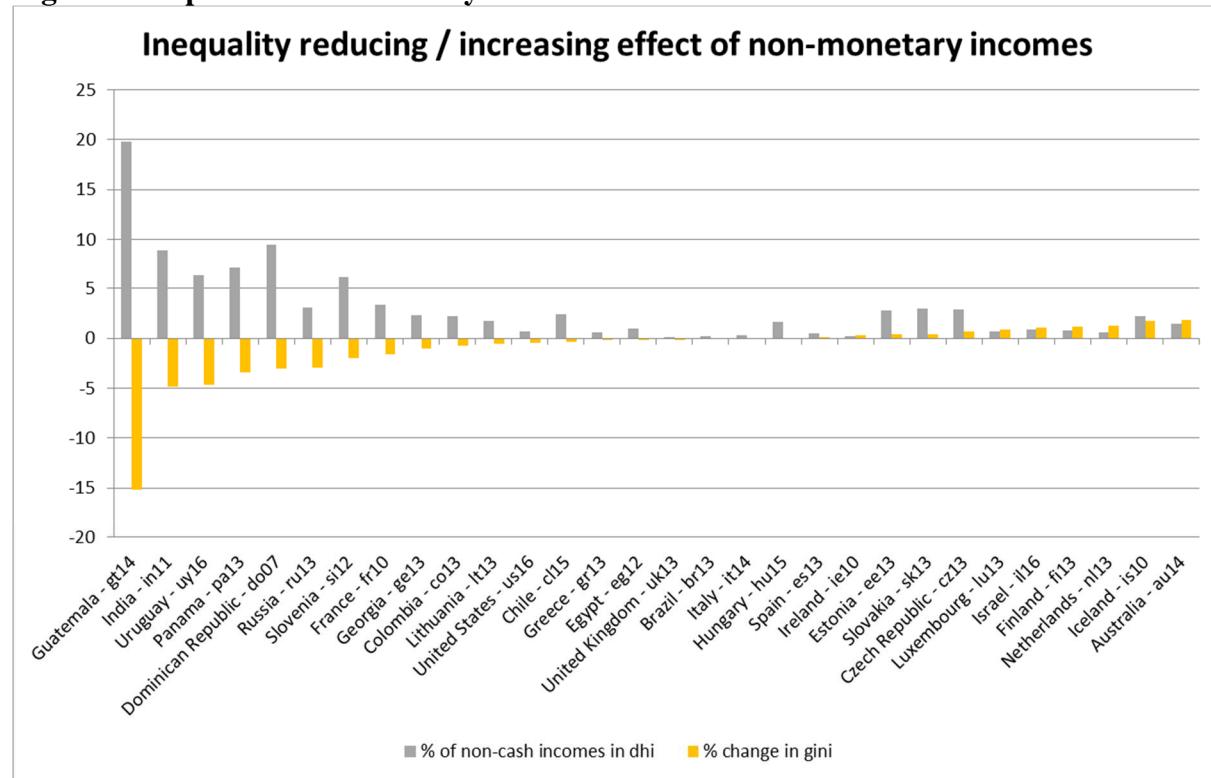
Source: LIS Database

**Figure 3: Extent of multiple jobs holders**



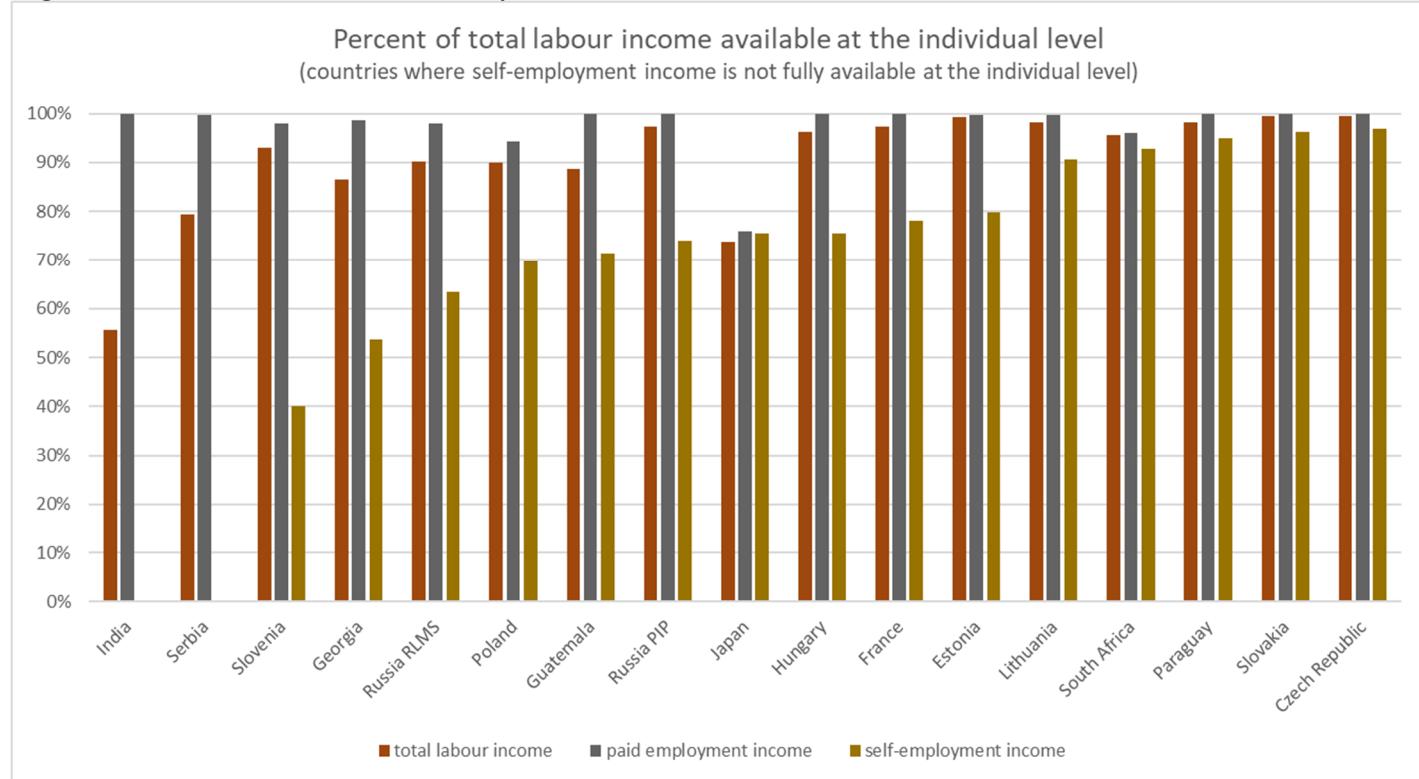
Source: LIS Database

**Figure 4: Impact of non-monetary incomes**



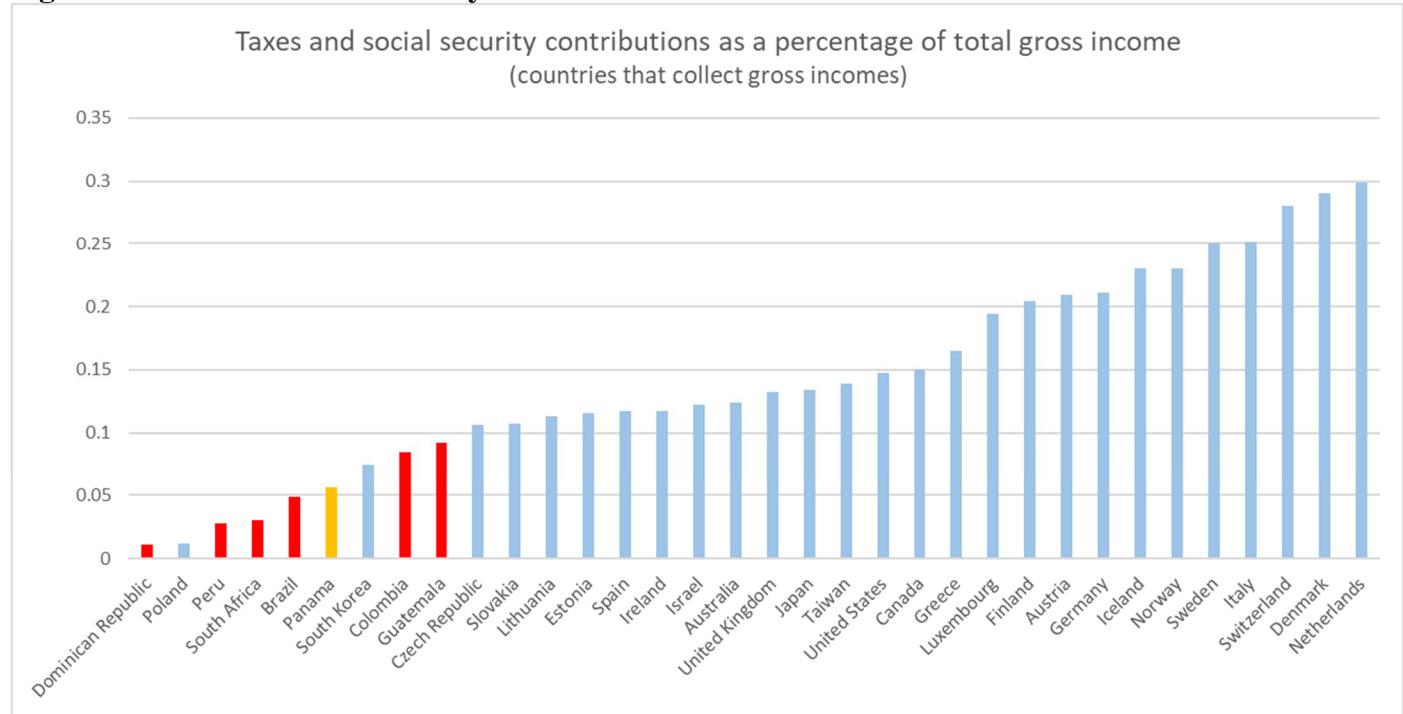
Source: LIS Database

**Figure 5: Labour income availability at the individual level**



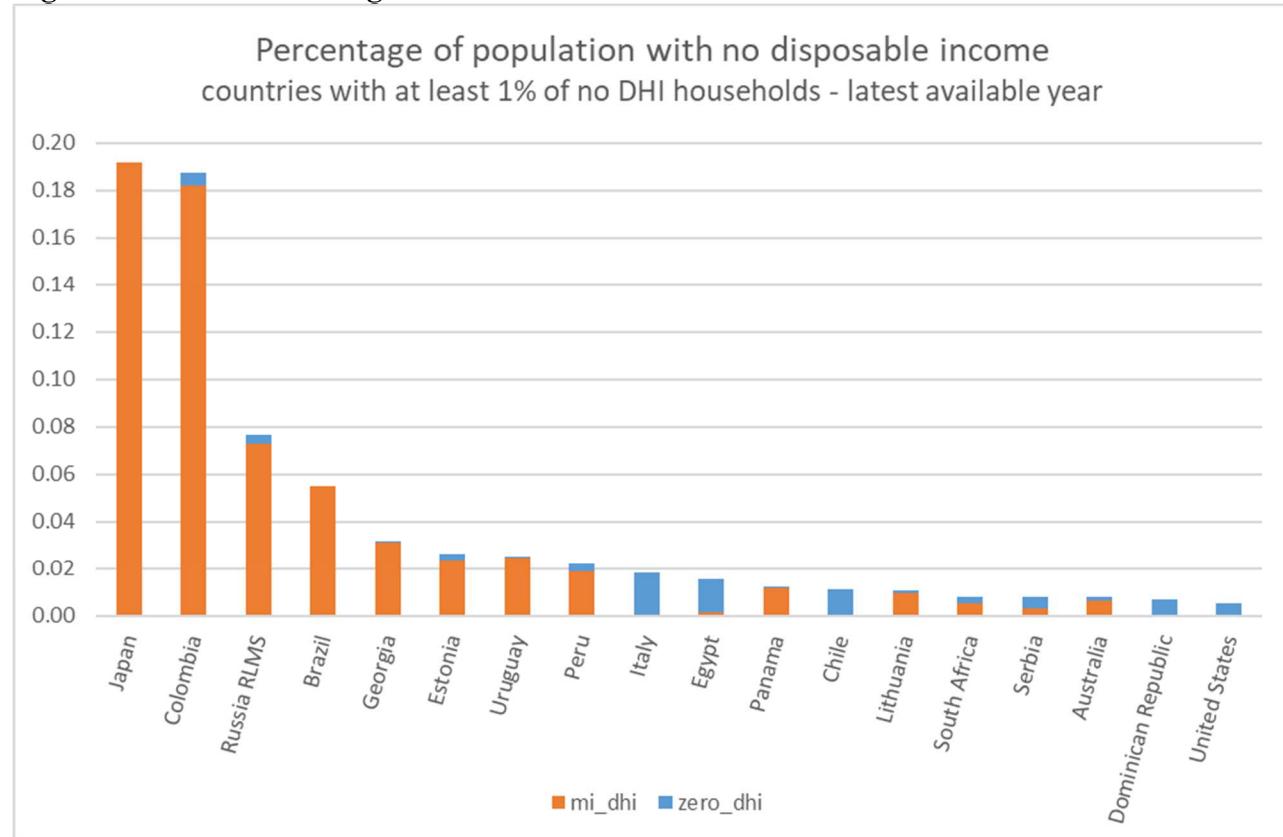
Source: LIS Database

**Figure 6: Taxes and social security contributions**



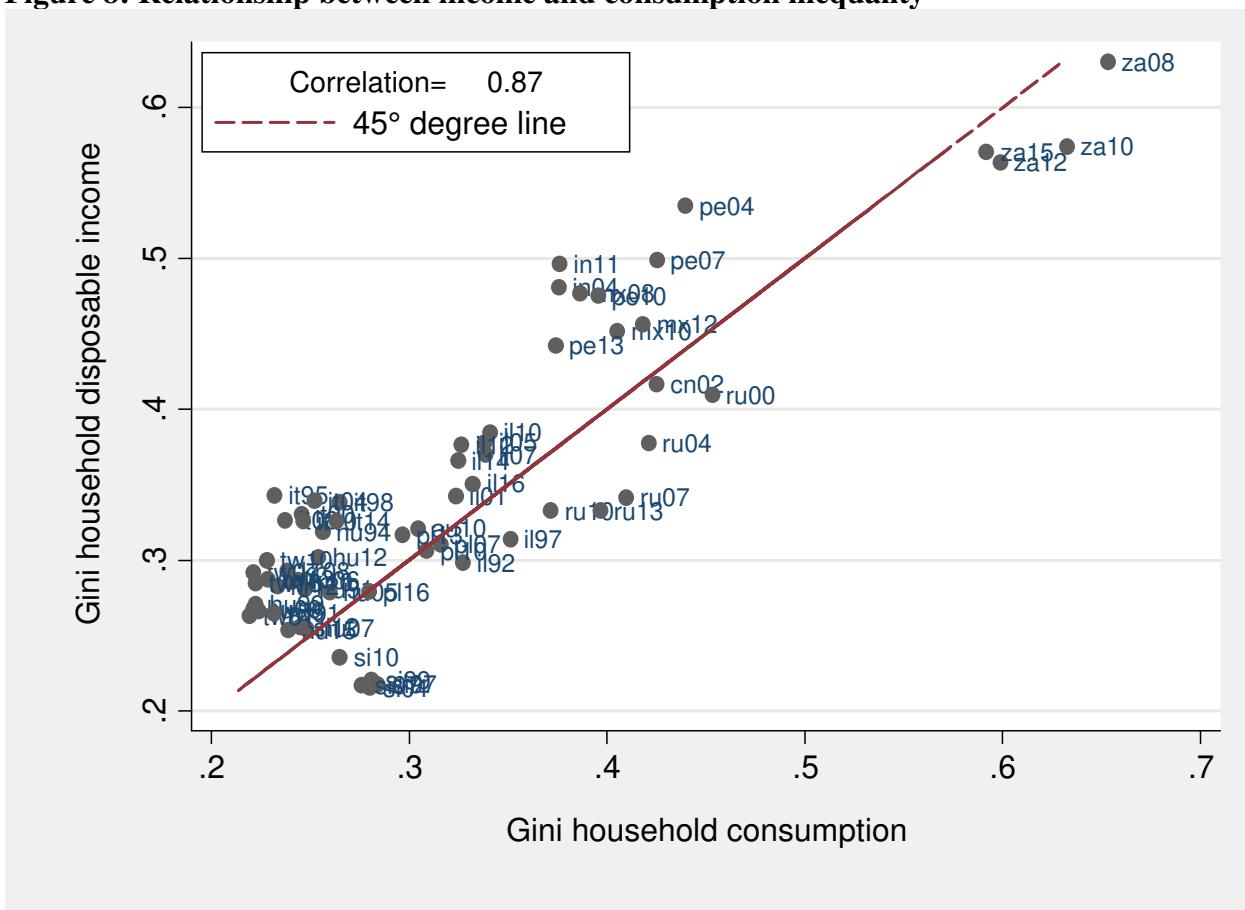
Source: LIS Database

**Figure 7: Extent of missing or zero income**



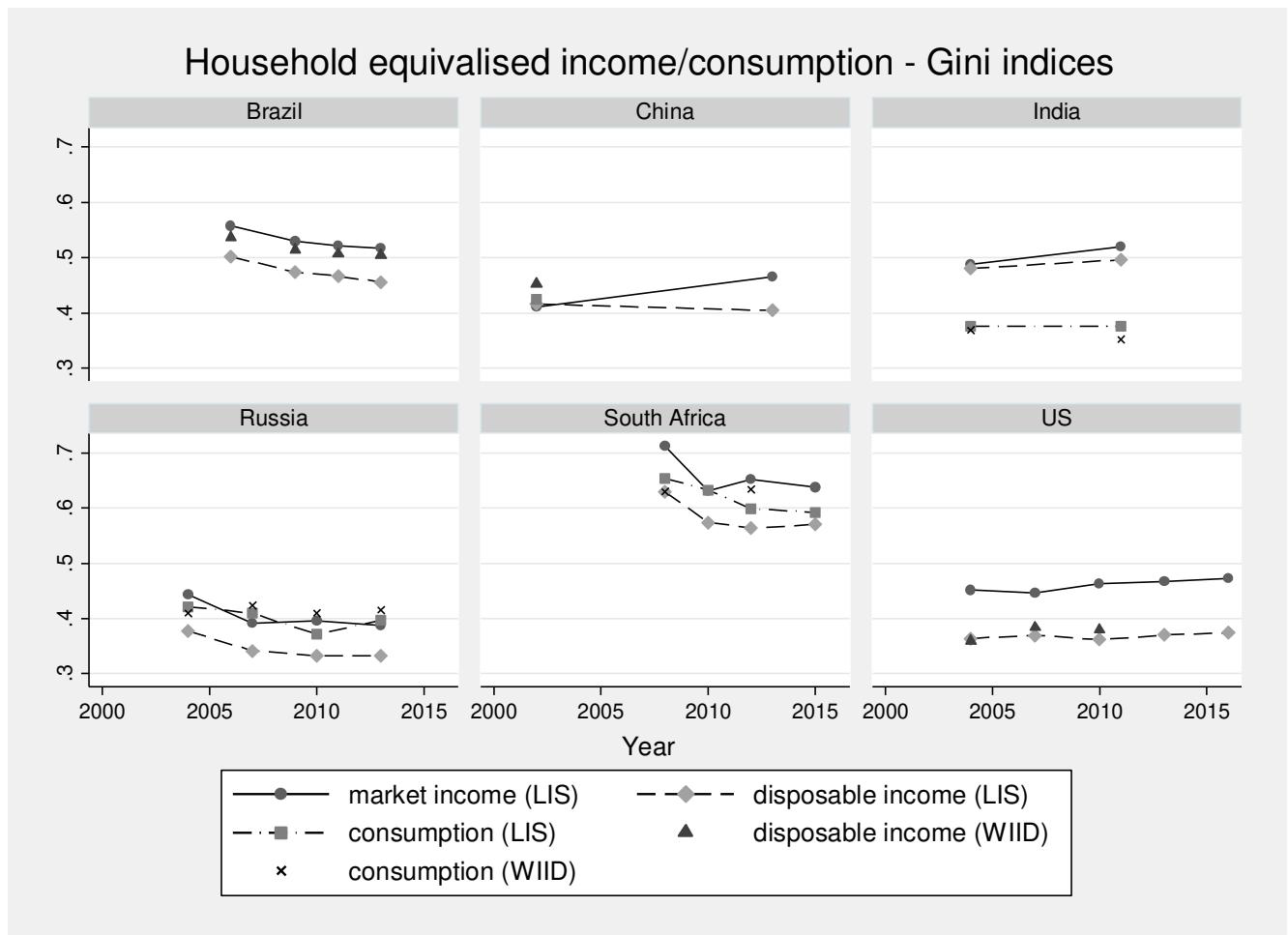
Source: LIS Database

**Figure 8: Relationship between income and consumption inequality**



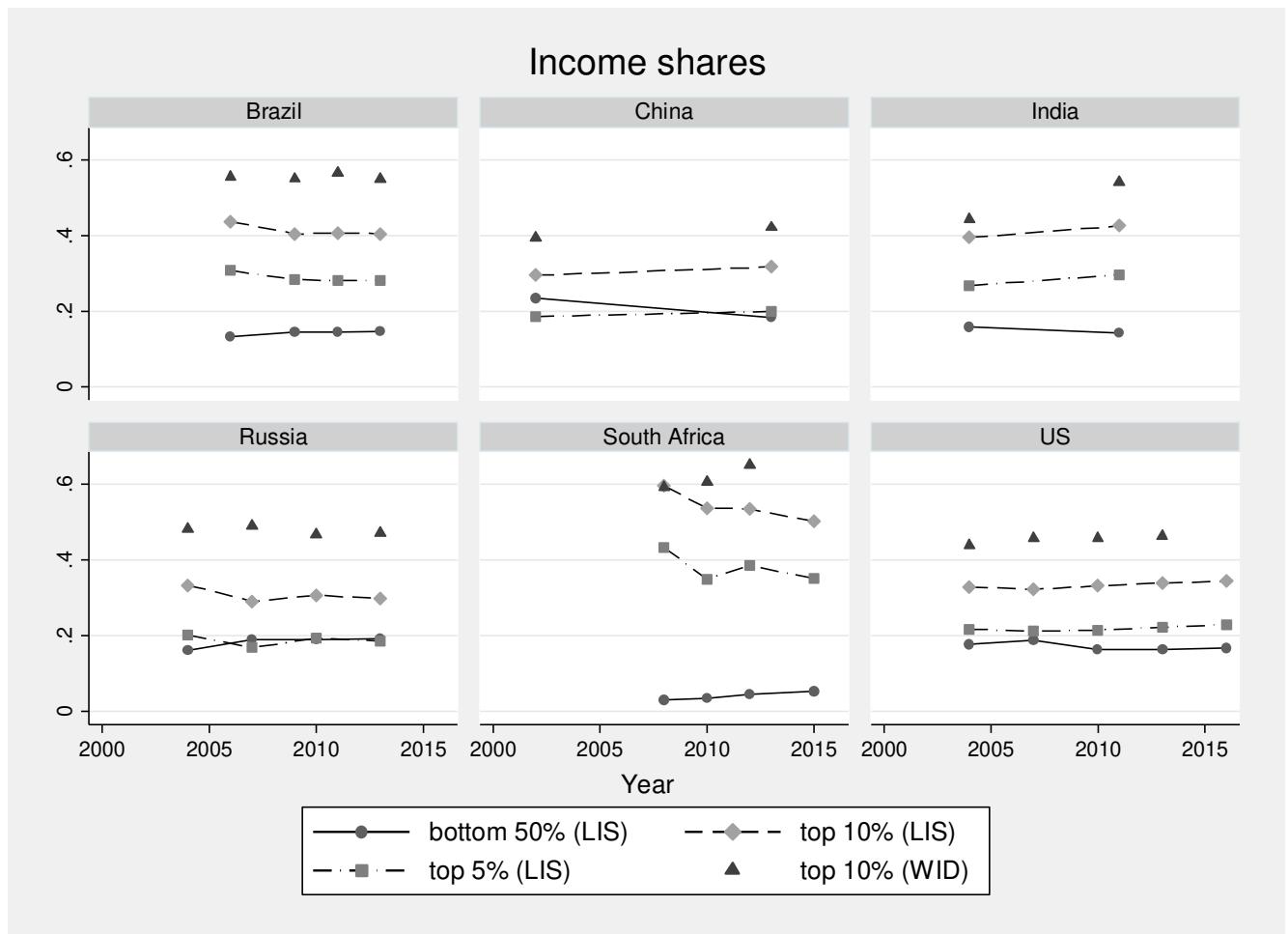
*Source: LIS Database*

**Figure 9: Trends in income inequality (Gini index) in selected middle-income countries (US as benchmark)**



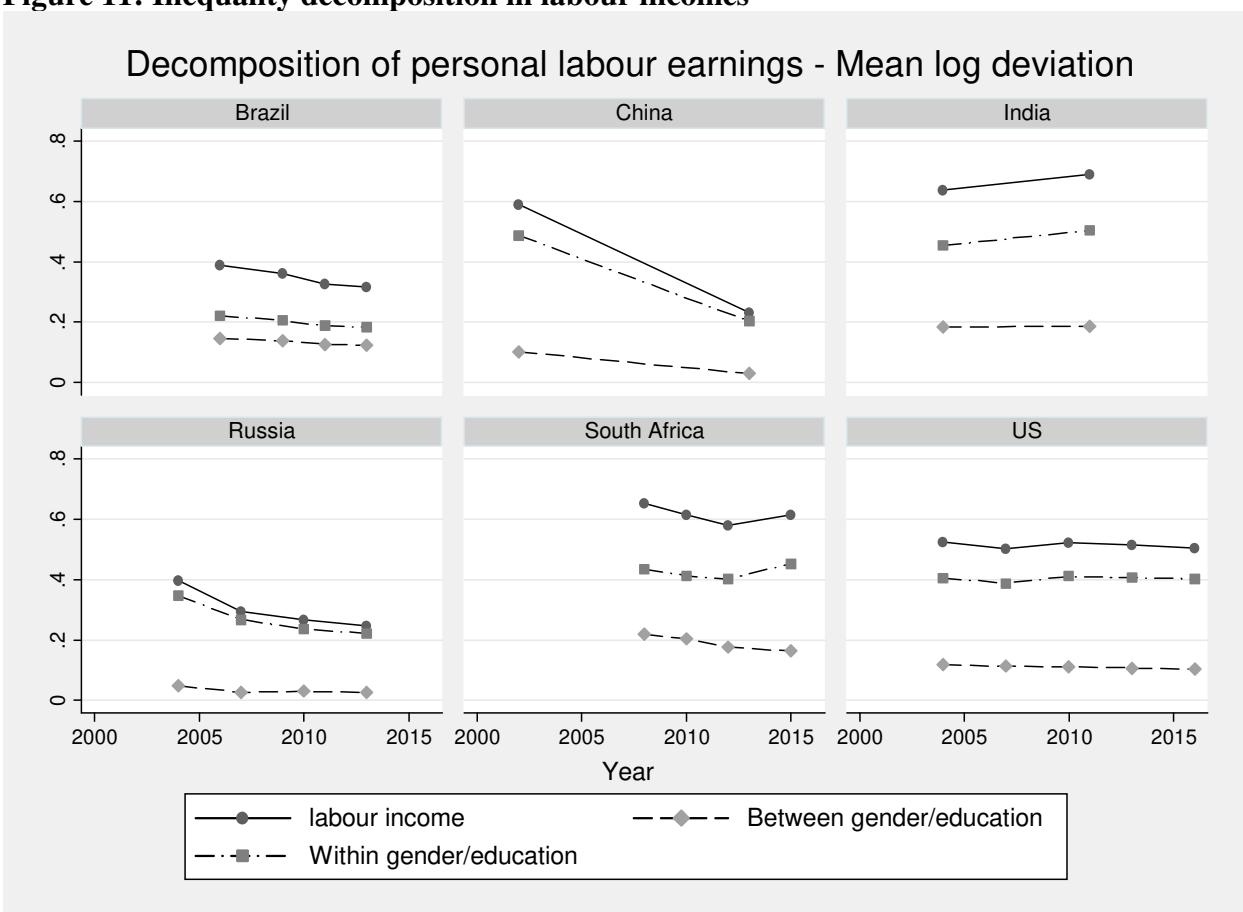
Source: LIS Database; WIID Database

**Figure 10: Evolution of income shares held by households with incomes below 50<sup>th</sup>, above the 90<sup>th</sup> and 95<sup>th</sup> percentiles**



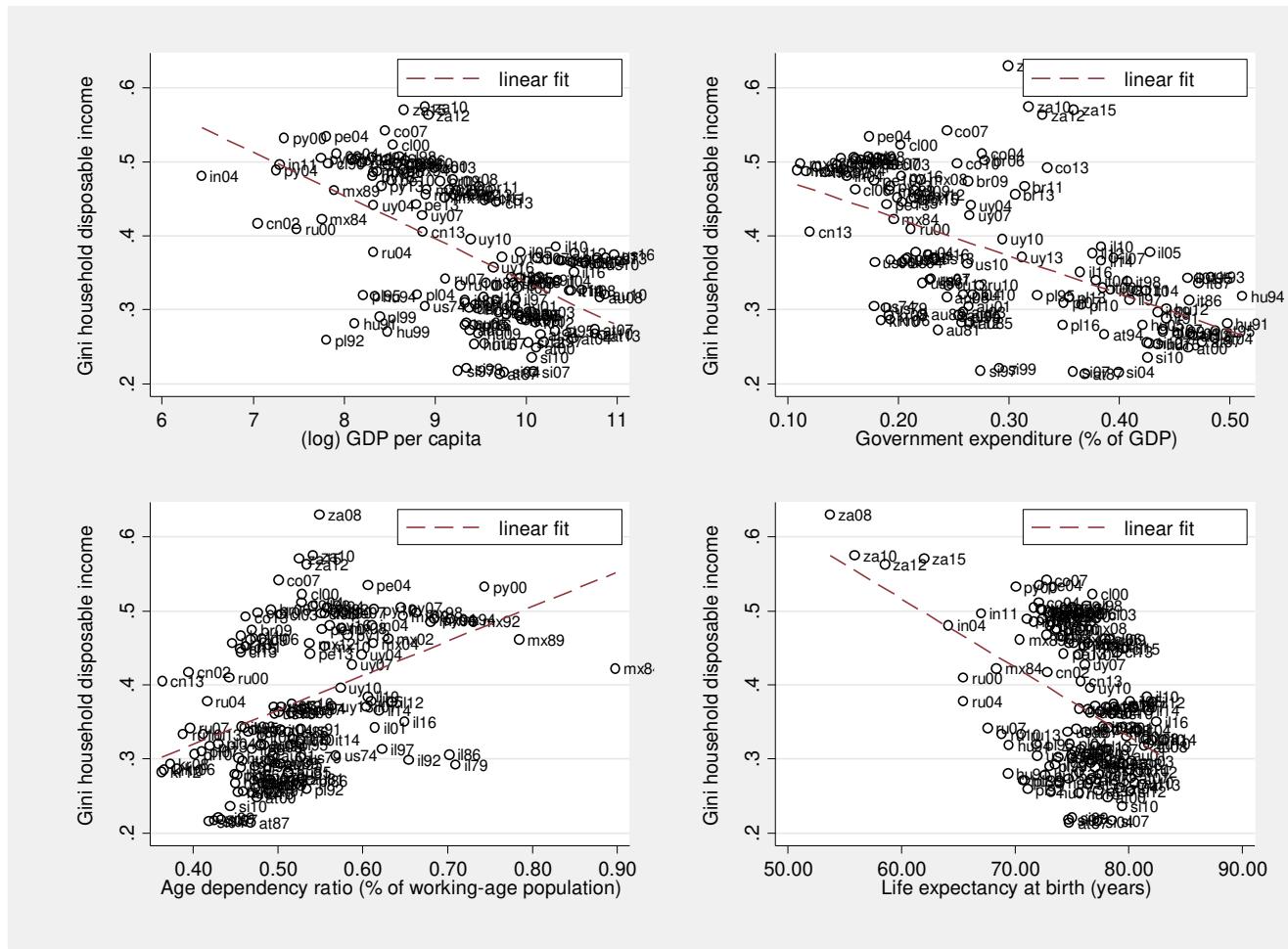
Source: LIS Database; WID Database

**Figure 11: Inequality decomposition in labour incomes**



Source: LIS Database

**Figure 12: Income inequality (Gini index) versus selected macroeconomic country characteristics**



*Note: Gini index of disposable household income (adjusted for outliers and equalized according to the OECD equivalence scale) presented on the vertical axis.*

*Source: LIS Database, and World Bank Indicators Database*

## Appendix: Additional tables and figures

**Table A.1: Overview of sample countries used in the empirical analysis**

Country	Historical data	Wave I (~ 1980)	Wave II (~ 1985)	Wave III (~ 1990)	Wave IV (~ 1995)	Wave V (~ 2000)	Wave VI (~ 2004)	Wave VII (~ 2007)	Wave VIII (~ 2010)	Wave IX (~ 2013)	Wave X (~ 2016)
Austria (AT)	n.a.	n.a.	AT 87	n.a.	AT 94 AT 95 AT 97	AT 00	AT 04	AT 07	AT 10	AT 13	n.a.
Australia (AU)	n.a.	AU 81	AU 85	AU 98	AU 95	AU 01	AU 03	AU 08	AU 10	n.a.	n.a.
Brazil (BR)	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	BR 06	BR 09 BR 11	BR 13	n.a.
Chile (CL)	n.a.	n.a.	n.a.	CL 90 CL 92	CL 94 CL 96 CL 00	CL 98	CL 03	CL 06	CL 09 CL 11	CL 13	CL 15
China (CN)	n.a.	n.a.	n.a.	n.a.	n.a.	CN 02	n.a.	n.a.	n.a.	CN 13	n.a.
Colombia (CO)	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	CO 04	CO 07	CO 10	CO 13	n.a.
India (IN)	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	IN 04	n.a.	IN 11	n.a.	n.a.
Israel (IL)	n.a.	IL 79	IL 86	IL 92	IL 97	IL 01	IL 05	IL 07	IL 10	IL 12 IL 14	IL 16
Italy (IT)	n.a.	n.a.	IT 86 IT 87	IT 89 IT 91	IT 93 IT 95	IT 98 IT 00	IT 04	IT 08	IT 10	IT 14	n.a.
Mexico (MX)	n.a.	n.a.	MX 84	MX 89 MX 92	MX 94 MX 00 MX 96	MX 98	MX 04	MX 08	MX 10	MX 12	n.a.
Paraguay (PY)	n.a.	n.a.	n.a.	n.a.	n.a.	PY 00	PY 04	PY 07	PY 10	PY 13	PY 16
Peru (PE)	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	PE 04	PE 07	PE 10	PE 13	n.a.
Poland (PL)	n.a.	n.a.	PL 86	PL 92	PL 95	PL 99	PL 04	PL 07	PL 10	PL 13	PL 16
Russia (RU)	n.a.	n.a.	n.a.	n.a.	n.a.	RU 00	RU 04	RU 07	RU 10	RU 13	n.a.
Slovenia (SI)	n.a.	n.a.	n.a.	n.a.	SI 97	SI 99	SI 04	SI 07	SI 10	SI 12	n.a.
South Africa (ZA)	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	ZA 08	ZA 10	ZA 12	ZA 15
South Korea (KR)	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	KR 06	KR 08	KR 10	KR 12	n.a.
Taiwan (TW)	n.a.	TW 81	TW 86	TW 91	TW 95 TW 97	TW 00	TW 05	TW 07	TW 10	TW 13	TW 16
United States (US)	US 74	US 79	US 86	US 91	US 94 US 97	US 00	US 04	US 07	US 10	US 13	US 16
Uruguay (UY)	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	UY 04	UY 07	UY 10	UY 13	UY 16

*Note: countries with microdata not available are labelled by “n.a.”.*

*Source: Luxembourg Income Study Database*

**Table A.2: Description of micro-variables used in the empirical analysis**

Variable	LIS codes	Description
Market income	FACTOR	Total current monetary and non-monetary income from labour and capital
Disposable household income	HI - HXIT	Total monetary and non-monetary current income net of income taxes and social security contributions
Household monetary consumption	HMC	Total consumption from expenditures (monetary consumption), i.e. consumption of goods and services that have been purchased by the household
Total individual income	PILE	Paid employment personal income. This includes: monetary payments and value of non-monetary goods and services received from regular and irregular dependent employment
Education	EDUC	Recode of the highest completed level of education into three categories: - low: less than secondary education completed (never attended, no completed education or education completed at the ISCED levels 0, 1 or 2) - medium: secondary education completed (completed ISCED levels 3 or 4) - high: tertiary education completed (completed ISCED levels 5 or 6)
Employment status	EMP	Indicator of any employment activity in the current period
Gender	SEX	Classification of persons according to their sex
Age	AGE	Age in years. Note that when original data provide age in intervals, values given are the lowest value of the interval. For example, the intervals 10-14 and 15-19 will be coded as 10 and 15, respectively
Household size	NHHMEM	Number of household members. Note that in most dataset this corresponds to the number of persons in the survey unit
Weight	HPOPWGT	Population household cross-sectional weight: this weight inflates the result to reflect the total household population covered by the dataset

Source: METIS (*METadata Information System*) of the Luxembourg Income Study Database

(<http://www.lisdatacenter.org/frontend#/home>)

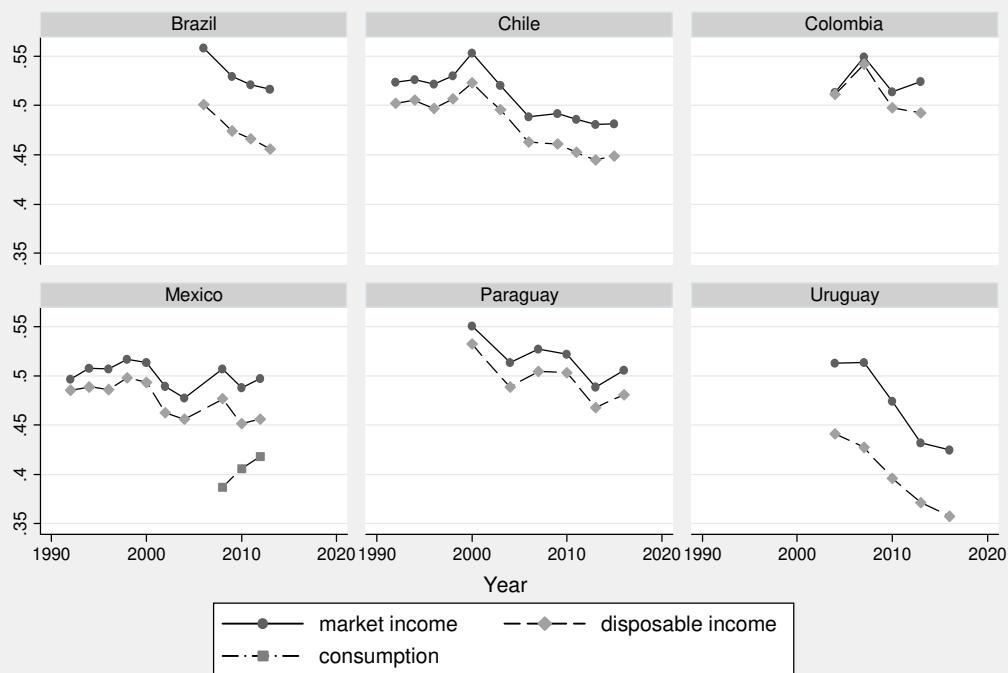
**Table A.3: Description of institutional variables used in the country-level analysis**

Variable	Description
GDP per capita	GDP per capita is gross domestic product divided by midyear population. GDP is the sum of gross value added by all resident producers in the economy plus any product taxes and minus any subsidies not included in the value of the products. It is calculated without making deductions for depreciation of fabricated assets or for depletion and degradation of natural resources. Data are in current U.S. dollars.
Agriculture, forestry, and fishing, value added (% of GDP)	Agriculture corresponds to ISIC divisions 1-5 and includes forestry, hunting, and fishing, as well as cultivation of crops and livestock production. Value added is the net output of a sector after adding up all outputs and subtracting intermediate inputs. It is calculated without making deductions for depreciation of fabricated assets or depletion and degradation of natural resources. The origin of value added is determined by the International Standard Industrial Classification (ISIC), revision 3. Note: For VAB countries, gross value added at factor cost is used as the denominator.
Age dependency ratio (% of working-age population)	Age dependency ratio is the ratio of dependents (people younger than 15 or older than 64) to the working-age population (those aged 15-64). Data are shown as the proportion of dependents per 100 working-age population.
Market capitalization of listed domestic companies (% of GDP)	Market capitalization (also known as market value) is the share price times the number of shares outstanding (including their several classes) for listed domestic companies. Investment funds, unit trusts, and companies whose only business goal is to hold shares of other listed companies are excluded. Data are end of year values.
Imports of goods and services (% of GDP)	Imports of goods and services represent the value of all goods and other market services received from the rest of the world. They include the value of merchandise, freight, insurance, transport, travel, royalties, license fees, and other services, such as communication, construction, financial, information, business, personal, and government services. They exclude compensation of employees and investment income (formerly called factor services) and transfer payments.
Exports of goods and services (% of GDP)	Exports of goods and services represent the value of all goods and other market services provided to the rest of the world. They include the value of merchandise, freight, insurance, transport, travel, royalties, license fees, and other services, such as communication, construction, financial, information, business, personal, and government services. They exclude compensation of employees and investment income (formerly called factor services) and transfer payments.
Government expenditure	Expense is cash payments for operating activities of the government in providing goods and services. It includes compensation of employees (such as wages and salaries), interest and subsidies, grants, social benefits, and other expenses such as rent and dividends.
Life expectancy	Life expectancy at birth indicates the number of years a newborn infant would live if prevailing patterns of mortality at the time of its birth were to stay the same throughout its life.
Urban population (% of the total population)	Urban population refers to people living in urban areas as defined by national statistical offices.

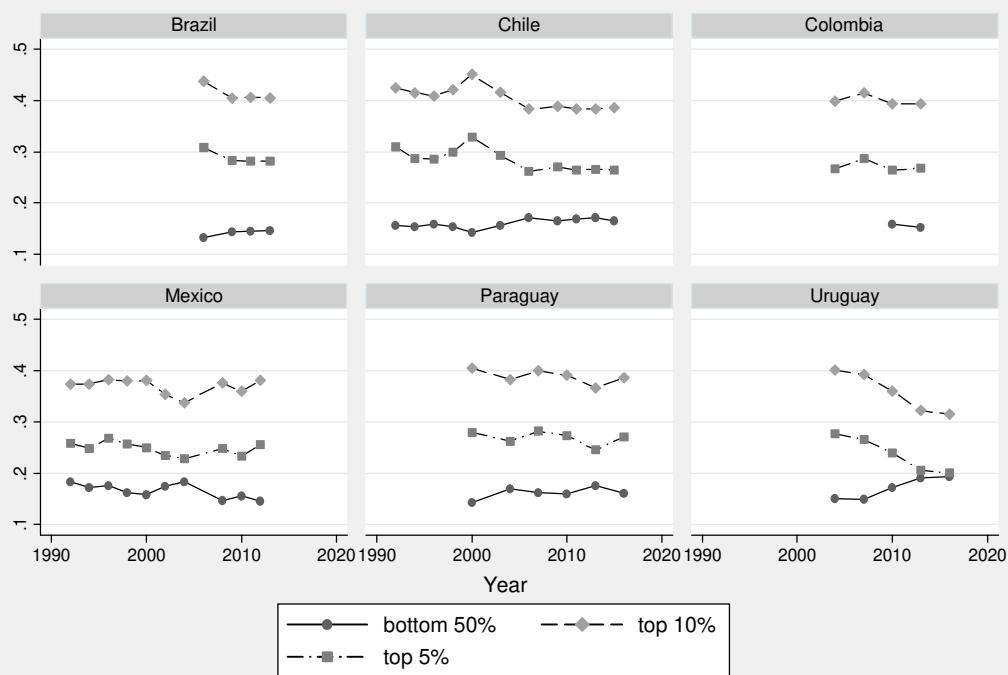
Source: World Bank Indicators Database

**Figure A.1: Inequality trends and evolution of income shares in Latin America**

### Household equivalised income/consumption - Gini indices

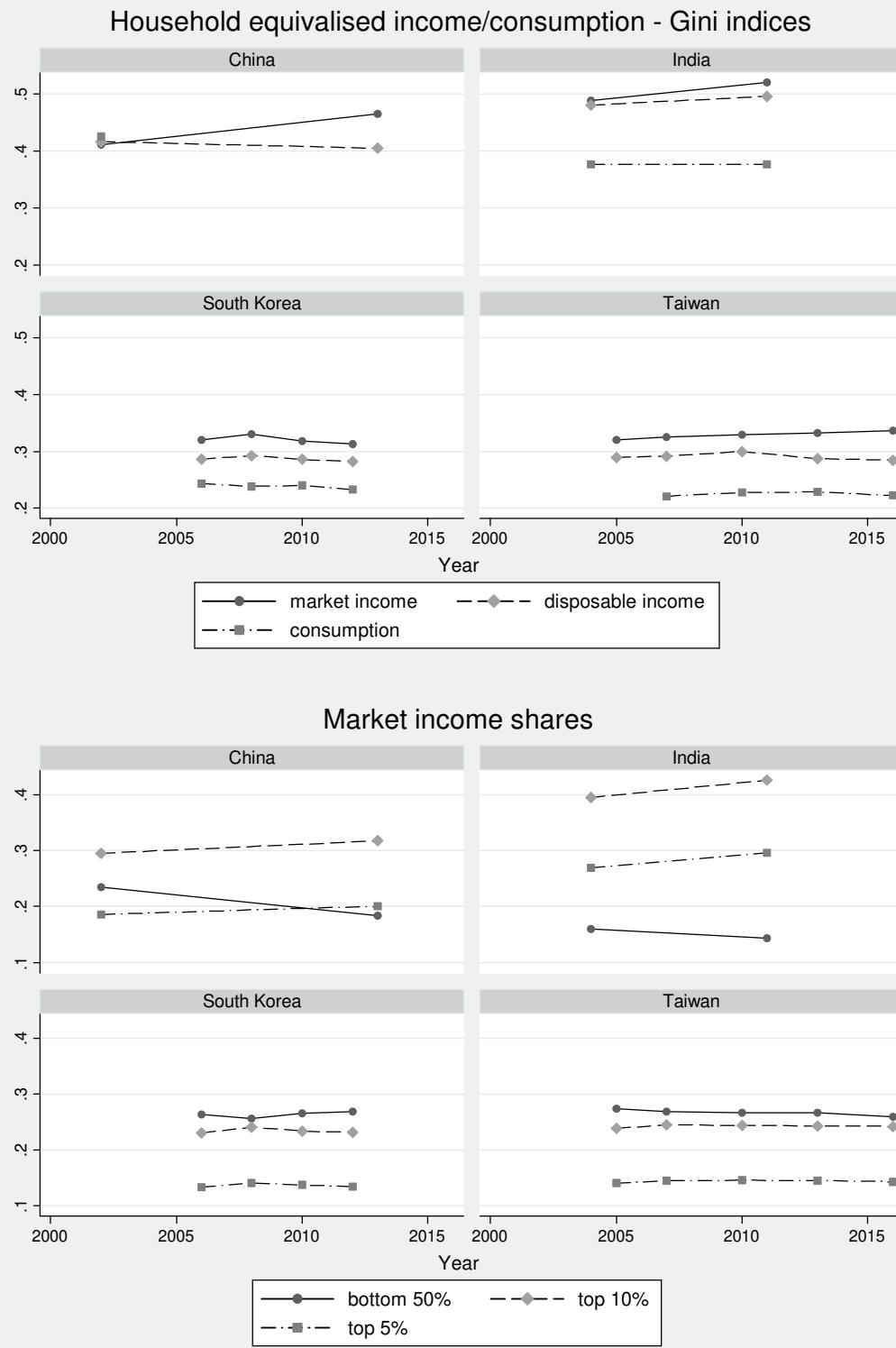


### Market income shares



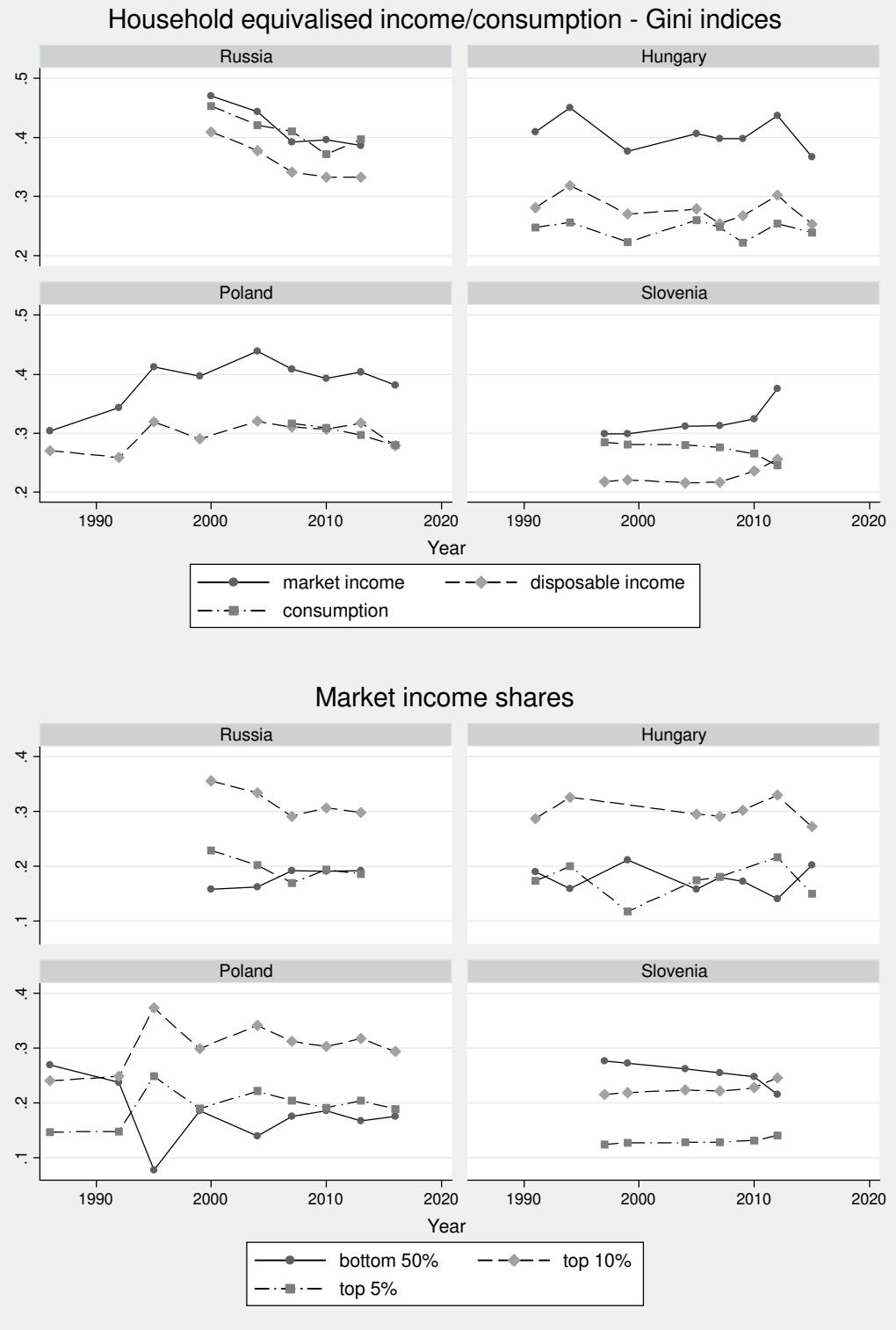
Source: LIS Database

**Figure A.2: Inequality trends and evolution of income shares in South-East Asia**



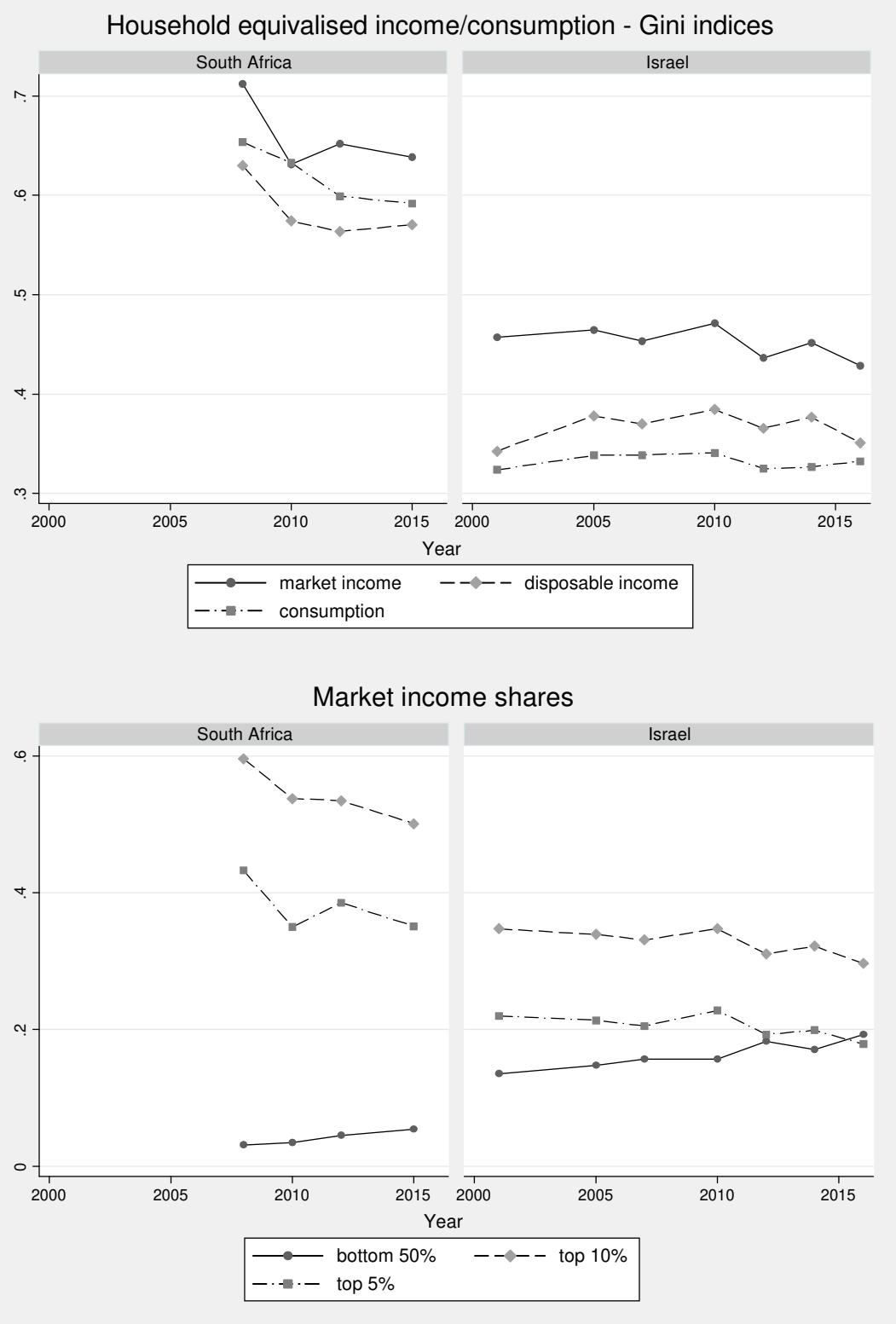
Source: LIS Database

**Figure A.3: Inequality trends and evolution of income shares in Central and Eastern Europe**



Source: LIS Database

**Figure A.4: Inequality trends and evolution of income shares in South Africa and Israel**



Source: LIS Database