

# LIS

## Technical Working Paper

No. 8

### **Unit nonresponse bias in inequality measurement: Worldwide analysis using Luxembourg Income Study database**

Vladimir Hlasny

Revised July 2017



CROSS-NATIONAL  
DATA CENTER  
*in Luxembourg*

---

Luxembourg Income Study (LIS), asbl

---

# **Unit nonresponse bias in inequality measurement: Worldwide analysis using Luxembourg Income Study database**

Vladimir Hlasny\*

July 12, 2017

## **Extended abstract**

This study evaluates the magnitude of biases in the measurement of income inequality due to the tendency of households at the top ends of national income distributions not to complete surveys. This unit nonresponse bias is evaluated and corrected for through a recently proposed probabilistic model utilizing information on regional nonresponse rates and income distributions. To this end, regional nonresponse rates were collected from individual national statistical agencies and are reported in this study. The study also uses 76 microdatasets from 38 high and middle-income countries across the world harmonized and made available by Luxembourg Income Study. The large number of surveys is used to infer consequences of unit nonresponse for the measurement of inequality in the industrialized and middle-income world and in major world regions. Analysis performed on survey data disaggregated geographically into regions for which nonresponse rates are available yields sizeable, systematic corrections for nonresponse. Nonresponse bias is estimated to affect national Gini coefficients typically by 1–8 percentage points, but sometimes by 15 percentage points or more, in line with previous estimates. Other inequality measures such as top income shares are affected more severely. Winsorizing performed by Luxembourg Income Study in producing key figures affects negatively not only Gini estimates but also the estimated corrections for nonresponse. This suggests that specific measurement issues such as the nonresponse bias evaluated here are unlikely to be identified and corrected fully in winsorized income distributions. The degree of geographic disaggregation appears to affect the nonresponse correction negatively, corroborating existing evidence. This phenomenon has not been explained satisfactorily so far and calls for further inquiry.

38 surveys, encompassing 891,271 household records in 422 regions, are used to derive back-of-the-envelope estimates of the Ginis for blocs of countries and worldwide. Gini coefficient for North America – of 44.72 after correction of unit nonresponse – appears quite robust to nonresponse, being merely 0.34 percentage points above the uncorrected statistic. Ginis for Europe and Latin America (excluding low income countries), however, are quite sensitive to the correction, rising from 32.11 to 37.63 (5.52 pc.pt.) in the European Single Market, from 35.25 to 42.97 (7.71 pc.pt.) in Europe including former Soviet-Union and Balkan nations, and from 49.12 to 61.36 (12.24 pc.pt.) in Latin America. Similarly, the Gini for the 35 OECD countries rises from 38.65 to 44.58, for a correction of 5.93 percentage points. Correction for unit nonresponse also matters for the estimation of the worldwide Gini. The corrected estimates, 57.35–57.77, are 3.11–4.14 percentage points above the uncorrected values, 53.64–54.24. Overall, these numbers from across countries, world regions and worldwide appear quite consistent with each other and suggest that unit nonresponse biases Gini coefficients downward by 1–8 percentage points.

**JEL:** D31, D63, N35.

**Keywords:** Top incomes, world Gini, inequality measurement, survey unit nonresponse, Luxembourg Income Study.

\* Department of Economics, Ewha Womans University (vhlasny@ewha.ac.kr). The author conducted a part of the research leading to these results on sabbatical leave as a visiting researcher at Faculty of Applied Economics, University of Antwerp, and at Luxembourg Income Study. The research received partial support under the European Commission's 7<sup>th</sup> Framework Programme (FP7/2013-2017) under grant agreement n°312691, InGRID – Inclusive Growth Research Infrastructure Diffusion. The author wishes to thank all staff of Luxembourg Income Study, and especially Carmen Petrovici, for help with the collection of documentation and data. Support and advice from Paolo Verme (World Bank) at the outset of the project are also acknowledged. Finally, this study wouldn't have been possible without the help of the many national statistical agencies with computing and making available regional nonresponse rates either directly in survey documentation or on the author's request.

## 1. Introduction

Household income surveys have traditionally been used to evaluate inequality in incomes. These surveys are known to contain various measurement problems, particularly at lower and upper ends of national income distributions (Atkinson and Micklewright 1983; Groves and Couper 1998). Respondents at both ends of the income distribution may have residencies that are difficult to reach, may not be present during visits by survey administrators, may fail to complete the survey in its entirety, or may be recorded in the sample with a wrong value for their income. Survey response is affected by households' characteristics and circumstances, and by survey design (Platek 1977). Misreporting may happen due to the households' incentives to hide their true incomes, their reliance on atypical sources and forms of compensation that are difficult to capture in standardized surveys, or the tendency of some statistical agencies to apply winsorizing (or top-coding or *rank proximity swapping*<sup>1</sup>) of incomes or individual income components to comply with privacy norms (Burkhauser *et al.* 2009, 2010, 2011, 2012).

Relatively little is known about the true distribution of incomes at the bottom and top ends of countries' income distributions even in industrialized nations, in part because of the rarity of these households in survey samples. The problem may be even more acute in developing countries with small survey samples, inappropriate design and administration of surveys, and earnings from informal labor and capital markets. At the same time, accounting well for bottom and top income households is important in all countries, because their influence on inequality

---

<sup>1</sup> Values above the cutoff for top-coding are swapped one for another within bounded intervals, and then rounded off, as is the case with US Current Population Survey (Moore 1996; Jenkins *et al.* 2011). US Current Population Survey suffers from other data issues including under-reporting of government assistance programs (Meyer, Mok and Sullivan 2009; Tiehen, Jolliffe and Smeeding 2013; Meyer and Mittag 2014), and unit and item nonresponse (Korinek *et al.* 2006; Dixon 2007). The US Census Bureau provides a limited correction for unit nonresponse by reweighting observations within adjustment cells (central and noncentral districts within metropolitan statistical areas, and urban and rural districts in non-MSAs) by the density of nonresponding households. This accounts for differences in response rates across adjustment cells but not for systematic differences across income groups within individual cells (Census & BLS, 2002, Ch.10-2).

measurement is substantial. Existing literature has established that unit nonresponse biases measurement of inequality (Mistiaen & Ravallion 2003; Korinek *et al.* 2006, 2007; Hokayem *et al.* 2012), and may be one of the factors explaining discrepancies in consumption data between national accounts and household surveys (Deaton 2005). This problem is particularly acute today given the calls for action in countries worldwide in response to the polarization of societies. Evidence in upper- and middle-income countries shows that the aggregate-income share of top-income households has risen significantly in recent years, and that the middle class is shrinking.

This study contributes to existing literature by evaluating how sensitive various national income distributions are to systematic unit nonresponse by top income households. Correction for nonresponse is performed by applying a recently proposed probabilistic model utilizing information on regional nonresponse rates and income distributions. Households are assigned weights according to their estimated inverse probability of response, which is a function households' income (Mistiaen & Ravallion 2003; Korinek *et al.* 2006, 2007; Hlasny & Verme 2013, 2015). Performance of this method is evaluated with respect to characteristics and structure of the data at hand. Adjustments are compared across countries to derive regularities and laws emerging from the corrections.

As a second contribution of this study, corrected measures of inequality at the national level, at the level of major world regions, as well as globally across the industrialized and middle-income world are presented. The study thus tries to add to the growing number of studies estimating global inequality and pointing out shortcomings with previous estimates (Sala-i-Martin 2002, 2006; Milanovic 2002a,b; Anand and Segal 2008, 2015; Nino-Zarazua *et al.* 2014; Bourguignon 2015; Lahoti *et al.* 2015). The study uses a worldwide sample of surveys from 38 upper- and middle-income countries that were harmonized and made available by Luxembourg Income Study (LIS). The analysis was partially undertaken on site at the LIS office in Luxembourg, which allowed the correct matching of geographic regions between surveys in the LIS database and nonresponse-rate information provided by national statistical agencies.

The third contribution of this study is thus to present research community with the regional nonresponse rates collected from individual national statistical agencies and prepared for merging on to the LIS database. When alternative measures of nonresponse are available, these alternatives are presented with their sources, and an attempt is made to align comparable nonresponse rates across countries.

The paper is organized as follows. The next section outlines key issues relevant to the study and the model used in the empirical section. Section three briefly describes the data under analysis, and discusses empirical challenges. Section four presents main results and section five concludes with their policy implications. An extended appendix provides regional unit nonresponse rates for all included national surveys, and information on the delineation of world regions.

## 2. Empirical Approach

Unit nonresponse refers to households that were selected into the sample but did not participate in the survey. Interviewers generally have lists of addresses that can be used to replace the missing household but this practice is not always sufficient to complete the survey with the full expected sample. Most available household surveys, particularly in developing countries, suffer from substantial unit nonresponse.<sup>2</sup>

### *Nonresponse-bias correction methods*

In the absence of any information about non-responding households, we can correct the bias either by imputing the incomes of non-responding households, or by re-scaling the incomes or the mass of responding households in a way to encompass the mass of non-responding households. Under the first option, estimation of the expected value of income for non-responding households would entail integrating incomes weighted by the corresponding probabilities of nonresponse across all possible incomes. With the imputed incomes of non-responding households, we would obtain the full income distribution on which to estimate measures of inequality. The problem with this method is that the results are sensitive to our assumption regarding the domain of incomes, and representativeness of the estimated income–probability relationship to counterfactual income levels.

An alternative to imputing incomes of nonresponding units is to apportion to responding units the known volume of aggregate income underreported due to nonresponse. A group of studies have attempted to align household survey results with those from national accounts by scaling up survey incomes to match aggregate national statistics (Bhalla 2002; Bourguignon and Morrisson 2002; Sala-i-Martin 2002; Lakner and Milanovic 2013). This method hinges on the restrictive assumption that the difference between survey and tax-record incomes is distribution-neutral, so that it would be appropriate to scale up all survey incomes by a fixed factor.

The third option entails imputing the true distribution of incomes by correcting the mass of each observation for its probability of being sampled. In this study we take this last approach. Inverted response probabilities estimated for each household serve as the appropriate household weights. In the income distribution imputed in this way, the derived measures of inequality converge to their true values as long as our sample is representative of the underlying population, and our behavioral assumptions are satisfied.

All of these three methods utilize external information about the population or about responding households and effectively extrapolate it to the mass of non-responding units. To infer

---

<sup>2</sup> For some surveys, the reason for nonresponse is recorded and sometimes this reason is used to correct the weights when the survey is completed if households have not been replaced. Reason for nonresponse can also be accounted for in the estimation of response probability. Korinek *et al.* (2006 and 2007) correct for the presence of type-A unit nonresponse in US CPS. Mistiaen & Ravallion (2003) add the count of item-nonresponse households to the type-A unit-nonresponse households, and use the methodology on them jointly.

households' probability to respond to a survey, some studies use information about the number of times it takes to reach each responding household, and model non-responding households as similar to the marginally responding ones – those reached at the last attempt (Bates and Creighton 2000). Some studies note which households serve as replacement households for non-responding units and weight households by their ease to reach, or use information on the reported reason for nonresponse. The problem with these methods is that the necessary external information may not be available consistently in all surveys at hand, and so these methods cannot be applied. Atkinson and Micklewright (1983) review a method that relies only on information about nonresponse rates across regions, whereas the mass of respondents in a region is 'grossed up' uniformly by the regional nonresponse rate, approach used by some national statistical agencies including the US Bureau of the Census (Census & BLS, 2002, Ch.10-2). The problem with this method is that it accounts only for inter-regional differences in nonresponse rates, and not for systematic differences across units within individual regions. This method also fails to model individual households' decisions to participate in a survey in a behaviorally sensible way.

*Correction using regional nonresponse rates*

Mistiaen and Ravallion (2003), and Korinek et al. (2006, 2007) propose a method to estimate response probability of individual units using basic information on regional nonresponse rates. This approach takes advantage of the variation in household nonresponse rates and the variation in the distribution of household characteristics (such as income or expenditure per capita) across geographic areas. Households' response probability is estimated by observing the propensity of households with similar characteristics across all regions to participate in the survey, by inferring regional populations from the estimated response probabilities of participating households, and by fitting the estimated populations to regions' actual known populations.

In what follows, we adopt the benchmark specifications from Korinek *et al.*'s (2006, 2007) and Hlasny and Verme's (2013, 2015) studies. The probability of a household  $i$  to respond to the survey,  $P_i$ , may be modeled as a logistic function of its arguments:

$$P_i(x_i, \theta) = \frac{e^{g(x_i, \theta)}}{1 + e^{g(x_i, \theta)}}. \quad (1)$$

Here  $g(x_i, \theta)$  is a stable function of  $x_i$ , the observable characteristics of a responding household  $i$ , and the corresponding parameters  $\theta$ .  $g(x_i, \theta)$  is assumed to be twice continuously differentiable in  $\theta$ , not necessarily monotonic. Equation 1 is estimated for all responding households, but the same behavioral relationship between households' demographics and their response probability is assumed to hold for nonresponding households as well. Parameters  $\theta$  can be estimated by fitting the estimated and actual number of households in each region using the generalized method of moments as

$$\hat{\theta} = \arg \min_{\theta} \sum_j [(\hat{m}_j - m_j) w_j^{-1} (\hat{m}_j - m_j)]. \quad (2)$$

Here  $m_j$  is the reported number of households in region  $j$ ,  $\hat{m}_j$  is the estimated number of households in the region, and  $w_j$  is a region-specific analytical weight proportional to  $m_j$ . The estimated number of households,  $\hat{m}_j$ , can be imputed as the sum of inverse response probabilities of all responding households in the region from equation 1.

$$\hat{m}_j = \sum_{i=1}^{N_j} s_i^{-1} \hat{P}_{ij}^{-1}. \quad (3)$$

Here  $\hat{P}_{ij}$  is the estimated response probability of responding household  $i$  in region  $j$ .  $N_j$  is the number of responding households in a region. If the sample is extracted from a larger population, the imputed true number of households should be divided by the sampling rate for the underlying population,  $s_i$ , to obtain population estimates in each region.

Assuming that household characteristics and the support of incomes in the sample are representative for those in the underlying population in each region, and assuming that the functional form of  $g(x_i, \theta)$  is stable for all households and regions, the estimator  $\hat{\theta}$  is consistent for the true  $\theta$ . Estimated values of  $\hat{\theta}$  that are significantly different from zero would serve as an indication of a systematic nonresponse bias. Inverted household response probabilities,  $\hat{P}_{ij}^{-1}$ , could then be used to reweight income observations to encompass the mass of non-responding households, and thus to correct for the non-response bias.

In the following empirical analysis, logistic functional form for  $P_i$  and natural-logarithmic specification for  $g(x_i, \theta)$  will be assumed in line with Korinek *et al.*'s (2006, 2007) and Hlasny and Verme's (2013, 2015) studies. We do not evaluate these specifications against possible alternatives, instead focusing on the comparative results of the benchmark model across datasets with very diverse characteristics, and implications for the measurement of inequality in the entire industrialized and middle-income world. Disposable household income per capita will be used as the only explanatory variable  $x_i$ , in recognition of the fact that this may be the only demographic variable available consistently across all household surveys.<sup>3</sup> In LIS database, disposable household income is computed as total monetary and non-monetary current income for the household, net of income taxes and social security contributions, and is made available and equalised across all LIS surveys.  $g(x_i, \theta)$  will thus take the form

---

<sup>3</sup> This parsimonious specification is thought to be robust and quite efficient in the measure of fit achieved. Since the explanatory variable, income per capita, is available in all income surveys (while household expenditures or other demographic information may not be), this specification is also preferable as most useful to practitioners. Using of more complicated forms would not improve the performance of the analysis, and would only distract from more pressing considerations.

If incomes suffer from mismeasurement or top-coding – such as the case of *rank proximity swapping* of incomes in the US CPS – while other demographic variables do not (or vice versa), using non-income explanatory variables (say, consumption) to correct income distribution for nonresponse bias would introduce measurement error.

$$g_i = \theta_1 + \theta_2 \log ey_i \quad (4)$$

where  $ey_i$  is equivalised disposable household income ( $dhi$ ) computed using LIS-selected adult-equivalent scale, square root of household size ( $dhi/\sqrt{nhhmem}$ ) (Atkinson, Rainwater and Smeeding 1995).

Population weights of households ( $hpopwgt$ ) are used as inverse sampling rates  $s_i^{-1}$ . Because disposable incomes may be negative, due to negative self-employment or capital income or due to tax considerations, only incomes of \$1 or higher are kept in the analysis.

#### *Choice of welfare aggregate and inequality measure*

Disposable household income per capita will be used as the welfare aggregate, in agreement with existing literature as well as the measure chosen by LIS. Gini coefficient will be used as a summary measure of income inequality. It is well recognized that some measures of inequality are more sensitive to issues in the tails of income distributions than others. The Gini index places less weight on topmost incomes and is thus less sensitive to nonresponse and other data issues in the upper tail than, say, indices in the generalized entropy (GE) class including the Theil index or GE(2) (half the squared coefficient of variation) (Cowell & Victoria-Feser 1996a,b; Cowell & Flachaire 2007). The Gini index is most sensitive to variations in income around the mode of the distribution. However, even Gini is systematically misestimated when household surveys do not capture top incomes precisely (Burkhauser *et al.* 2009), a problem that motivates this study. To the extent that we find systematic unit-nonresponse biases to the Gini, we will interpret them as lower estimates of the biases to other measures of inequality.

The following empirical analysis will be used to report Gini coefficients computed using two versions of national income distributions: the original distributions made available by national statistical agencies, and distributions winsorized using LIS methods.<sup>4</sup> The nonresponse correction model will be estimated on original income data, but the estimated nonresponse correction weights will be applied to both the original and the LIS-winsorized income distributions. Because our correction for unit non-response is sensitive to the presence of other measurement issues and arbitrary censoring among top incomes, we believe that the original income distributions will yield more accurate estimates of nonresponse bias and corrections for it. To evaluate the significance of the issue of data-censoring, both Gini estimates are provided and compared.

#### *Degree of demographic disaggregation*

---

<sup>4</sup> To account for potential measurement issues with bottom and top incomes, LIS bottom- and top-codes income distributions when reporting results such as the *key figures* at [www.lisdatacenter.org/data-access/key-figures](http://www.lisdatacenter.org/data-access/key-figures). This practice is on top of any winsorizing performed by some statistical agencies in the production of the original income distributions.

In order to compare income distributions and nonresponse rates of different demographic groups, as the above method essentially proposes, it is not necessary to divide survey sample by geographic regions. One may alternatively perform the analysis by other demographic criteria observable even for nonresponding units such as the degree of urbanization, survey rotational group, date of interview, etc. The success of analysis performed using these alternative data demarcations depends on how well behavioral assumptions are satisfied within and across the demographic groups, and empirically on the availability of nonresponse rates for the distinct groups, and availability of a sufficiently large number of groups with a sufficiently large number of units in each of them. These empirical constraints are binding. While nonresponse rates are often reported by survey rotational group, this yields only 3–4 groups and thus only 3–4 points to fit in equations 2 and 3. Differentiation of areas by the degree of urbanization also yields too few groups (e.g., thin/intermediate/dense population in EU-SILC), which may moreover differ behaviorally from one another.

Geographic disaggregation of nonresponse rates is reported in more surveys, and is available for as many as 51 subnational regions in surveys included in LIS database, often at more than one level of disaggregation (refer to table 1). Geographic disaggregation is also thought to satisfy the behavioral requirement that income distributions and nonresponse rates across all distinct groups be linked by a stable probabilistic relationship (equations 1 and 4). Finally, samples are often stratified at a top regional level (NUTS-1), ensuring representativeness for the underlying population and coverage of the entire support of incomes.

Having argued that geographic disaggregation has good properties for the task of correcting bias using group-level nonresponse rates, we next consider the appropriate degree of disaggregation. The choice of disaggregation involves a trade-off between the number of  $j$  data points in equation 2, and the number and distribution of within- $j$  observations vis-à-vis the characteristics of nonresponding units. On the one hand, observations must be behaviorally similar to nonresponding households within region  $j$  and to households with similar values of explanatory variables in surrounding regions, calling for smaller and more numerous geographic areas. Also, disaggregation into more regions makes estimation less sensitive to any outlying regions with atypical income distributions or nonresponse rates, as it translates into more residuals being minimized in equation 2. On the other hand, the sample of respondents must cover the entire support of values of incomes and explanatory variables of nonrespondents, calling for larger geographic areas. One consequence of this requirement is that the sample of households should be stratified at the regional level.

The correct choice over the degree of disaggregation is ultimately an empirical question, contingent on the characteristics of data at hand. Hlasny and Verme (2013, 2015) found that disaggregation into smaller regions yields systematically smaller bias corrections and – using simulation results from three different surveys – argued for intermediate or more detailed degrees of disaggregation. In this study we will attempt to balance between these considerations and, when nonresponse rates are available at multiple geographic levels, use geographic

disaggregation that yields a sufficient number of regions as well as adequate sample sizes within regions, whose samples are representative of underlying populations.

*Correcting estimates of inequality in major world regions and worldwide*

Next we will attempt to infer general patterns in nonresponse biases from our sample of national surveys and their implications for measurement of inequality at the level of blocs of countries, world regions and worldwide. We will use a subsample of surveys and variables that are most comparable among themselves, and most informative of cross-border and worldwide inequality. Several rules are followed in the selection of surveys and in specifications used in the empirical analysis. Only surveys with information on the unit nonresponse rate are used. For the analysis of worldwide inequality, surveys are selected to cover all major world regions (Europe, North America and Latin America explicitly, plus limited coverage of Africa, Middle East, East Asia and Australia) and as much as possible of worldwide distribution of incomes, including high, upper-middle and lower-middle income countries.

Low income countries are notably missing from the LIS database, and so this analysis cannot be applied to understanding inequality or nonresponse bias in these countries. Even lower-middle income countries are represented very lightly. More generally, the group of countries in the LIS database are not representative of countries or households around the world. To account for the undue presence of high and upper-middle income countries in the analysis, stratification by income bracket is applied: Weights are assigned to the three respective groups of countries – high, upper-middle and lower-middle income – under the highly-stylized assumption that countries missing in each income bracket have a similar distribution of incomes, nonresponse rates, and behavioral relationship between them, as those studied. In particular, population sampling weights of households are divided by a subsampling rate  $ss$  according to the portion of world working-age population covered by LIS-database countries in our study.

$$ss_K = \frac{\sum_{k \in K \cap LIS} m_k}{\sum_{k \in K} m_k} \quad (5)$$

Subscript  $k$  corresponds to countries (or national regions), and  $K = \{HIC, UMC, LMC\}$  is an index of sets of countries (or regions) classified as high, upper-middle and lower-middle income, respectively.<sup>5</sup>

Gini coefficients for the OECD bloc of countries, the entire industrialized and middle income world, as well as for three major world regions – Europe, North America and Latin America – are estimated. Stratifying the set of surveys into high, upper-middle and lower-middle income

---

<sup>5</sup> For instance, if Guatemala is the only lower-middle income country in the LIS database available for analysis, its working population of 9.3 million in 2014 yields a subsampling rate of 0.005, using the total working-age population of lower-middle income countries of 1,829.5 million (World Bank 2015c).

countries (following World Bank's, 2015c, classification) is used to capture the main lines according to which national surveys differ, in terms of income levels, nature of income inequality and survey characteristics. Countries are classified as belonging to three world regions that are adequately covered in the LIS database – Europe, North America and Latin America – considering economic interlinkages among countries, actual flows of trade and labor across countries, and the grouping of countries into effective trading blocs during the first decade of the 21<sup>st</sup> century. Clearly, the income-bracket and geographic demarcations used below are not without controversy (e.g., the status of Cuba, Turkey, Russia etc.), but no clearer, theoretically superior alternatives appear on offer.

Only one survey wave per country is used. To the extent that available surveys come from various years, only the most recent surveys or those most proximate in time are used. These surveys may then be viewed as informative of worldwide inequality in a given, recent point in time. Incomes are also converted to a single constant-price currency, year-end 2005 purchasing power parity international dollars, using currency conversion factors and US GDP-deflator inflation (World Bank 2015a,b).<sup>6</sup> This affects measurement of inequality and even the Gini coefficient when income distributions are censored at absolute levels.

Variables used in the analysis – namely disposable household incomes per capita, nonresponse rates and household weights – must be harmonized and comparable. LIS database serves this purpose well as incomes from all surveys, as much as possible, are constructed using the same conventions and same income components. LIS makes a note of any residual differences, and carefully marks surveys with gross, net and mixed incomes.<sup>7</sup> Household weights in the LIS database are inflated to national population for use in cross-country studies.

Importantly for this study, definition of nonresponse rates varies across national surveys and the responsible statistical agencies. An attempt is made to distinguish what precisely they comprise and, when alternative definitions are available, to choose nonresponse rates comparable across years and countries.<sup>8</sup> To further mitigate the problem of heterogeneous income and nonresponse-rate distributions across country surveys, we include country indicators in the behavioral relation between households' income and their probability to respond to a survey (equation 4):  $g_i = \theta_1 + \theta_2 \log ey_i + \sum_{(k \neq 1) \in K} \theta_{3k} I(k = \text{country}_i)$ , where  $I(\cdot)$  is an indicator function for households'

---

<sup>6</sup> For Belgium ('95, '00), Estonia ('04–'10), Romania ('97) and Slovakia ('04, '07), currency conversion factors are taken from European Central Bank exchange list archive.

<sup>7</sup> The issue of accounting for taxes and contributions, either by direct observation or imputation, in all income specifications (gross), in total incomes but not in income components (mixed), or not at all (net).

<sup>8</sup> In addition, one may wish to evaluate how comparable household weights are to those in other countries, whether they lead to regional or national representativeness of samples, and whether they themselves account for unit nonresponse. This latter property would be worrisome as it would effectively double count a portion of the nonresponse bias and potentially lead to overcorrection for the problem. Some statistical agencies, including the US CPS, adjust the initial household sampling weights to correct for unit nonresponse in demographic cells (Census Bureau 2002, Chapter 10). Eurostat does not correct for it. To preserve space, this study does not explore this issue further.

country. This allows for cross-country intercept differences in the relation, but it keeps assuming the same marginal impact of percent changes in income on response probability.

Finally worth noting, the model correcting for unit nonresponse was estimated using a custom-written program in Stata statistical software. The following section describes the data used in detail.

### 3. Data

This study relies on over 70 household surveys from 38 countries collected and harmonized by LIS. As of October 2015, LIS offers public access to over 250 income distributions for 45 countries, and additional surveys are being added several times a year. The datasets are harmonized and can be studied jointly both across years and across countries. In this study, only countries for which subnational regional nonresponse rates are available are used, allowing us to perform the correction for nonresponse at national as well as cross-national level. Only the most recent waves of national surveys are used, to focus on inequality at its level today, and to ensure comparability in our attempt to infer inequality worldwide.<sup>9</sup>

In the LIS database, we use information on disposable household income, normalized household sampling weights, household size, and administrative-region or population-density identifiers (variables *dhi*, *hpopwgt*, *nhhmem*, *region\_c* and *locsz\_c*). Beside variables included in the LIS data, information on unit non-response rates at the subnational geographic level was collected from survey documentation or from individual national statistical agencies responsible for administering the surveys. Regional non-response rates were merged onto the survey data using regional indicators *region\_c* and *locsz\_c*. Refer to tables A1 and A2 in appendix 1.

Surveys in the LIS database can be viewed as covering most of the spectrum of national surveys encountered around the world in terms of the level of economic development, inequality, sample size, unit nonresponse rate as well as potential for geographic disaggregation into regions. As table 1 shows (among surveys that allow geographic disaggregation), LIS database covers surveys with various definitions of incomes (net, mixed, gross); of various sample sizes ranging from 2,024 to 121,163; with various potential for geographic disaggregation – from none to three degrees of disaggregation into as many as 51 regions; with nonresponse rates ranging from 0 to 70 percent. Figures A1 and A2 in appendix 1 illustrate.

Definitions of nonresponse rates also differ in important ways across surveys worldwide. The majority of surveys report the ratio of responding households to those eligible or initially drawn, accounting for refusals, temporary absences, households unable to locate or access, and other

---

<sup>9</sup> Selected older surveys are used in order to evaluate robustness of results. Refer to table 1. As another robustness check, six countries without information on subnational regions are added to the analysis of worldwide inequality: *at04*, *kr06*, *lu10*, *nl10*, *pl10* and *ro97*. These additional countries are treated as regions comparable to subnational regions in other surveys.

interview failures among the drawn households. Some surveys account for households found not to belong to sampling universe, or even respondents' death. A small number of surveys include only refusals and partial incompletions. Some surveys – particularly panel surveys – (attempt to) replace non-responding households with reserve units, and report nonresponse rates among either or both of these original and replacement groups (refer to appendix 1, table A1).<sup>10</sup> These definitions may change over time, particularly in panel surveys. Finally worth noting, the reported definitions may differ systematically from the situation on the ground, depending on the efforts and motives of survey contracting companies and individual interviewers.

The LIS database covers income distributions on all world continents, in lower-middle to high income countries (mean incomes ranging from 2,457 in Romania as of 1997 to 49,244 in Luxembourg in 2010, 2005US\$ PPP); and with very low to very high income inequality, both on winsorized and non-winsorized income data. In light of this heterogeneity, the LIS database is an ideal testing ground for inequality estimation methods such as the one applied in this study.

### *Degree of demographic disaggregation*

For surveys in table 1, national statistical agencies reported unit non-response rates by smaller regions often following nomenclature of territorial units for statistics (NUTS) at the level of NUTS-1 (divisions/provinces/states) or NUTS-2 (counties/districts). In some cases more limited degree of disaggregation is possible, into areas with dense, intermediate or thin population (e.g., Iceland). The possible degree of disaggregation depends both on the reporting practices by national statistical agencies as well as on the precision of regional indicators in individual LIS datasets, because the two sets of information must be merged at the level of regions. Some statistical agencies (e.g., Brazil, Estonia, Peru, Slovakia, South Africa, US) release household registers including both responding and nonresponding units, sometimes even with information on reason for nonresponse, so the degree of geographic disaggregation depends only on the availability of regional indicators in the LIS version of the surveys. Many statistical agencies (including some of those administering EU-SILC surveys) report nonresponse rates by survey rotational group.

As section 2 noted, the appropriate geographic disaggregation should yield regional samples and income distributions that are representative for the underlying populations, and therefore be stratified at the level of regions. This is not precisely the case in many countries. In EU-SILC surveys, regional indicators follow the level of stratification only approximately.<sup>11</sup> Sweden does not apply

---

<sup>10</sup> Japan and Mexico report a 'quasi-nonresponse' rate among both the original and reserve units. In Taiwan's Survey of Family Income and Expenditure (2010), refusing and nonresponding households were replaced until the desired sample size was achieved. In Hungarian Household Monitor Survey (2005), similarly, dropped out units are replaced with 'extra addresses.' Information on the number of nonresponding units is missing. In panel surveys, including for Germany, Russia and South Africa, we should ideally distinguish 'longitudinal' nonresponses (attrition) and 'cross-sectional' nonresponses among households contacted for the first time, as these nonresponses may be driven by different factors. The methods of dealing with households that moved to new addresses, and with households that became split or connected, also differ across surveys.

<sup>11</sup> In EU-SILC surveys, information on sample design is incomplete. The original stratification variable (DB050) is

stratification explicitly. Even though it is unclear theoretically whether these issues affect the consistency of the correction method evaluated in this study, they affect the sizes of estimated standard errors adversely.

---

missing, but NUTS1 or NUTS2 regional indicators can be used as substitutes. These variables may underestimate the number of strata in some member-states. For instance, Finland applies stratification by other than geographic criteria. In Spain, regions Ceuta and Melilla must be grouped together as is the case in the real sample design.

Table 1. Sample summary for countries with subnational nonresponse rates

Country	Income ref. year	LIS dname	Curr= '05\$1	Net/ gross	Sample size	NUTS disaggr.	Regions (#)	Nonresp. rate	Avg. inc. (\$)	Gini (LIS) <sup>a</sup>	Gini <sup>b</sup>
Australia	2010	au10	1.367aud	gross	18,071	1	7	21.40	36,396	32.8	33.1
Austria	2008	at08	0.765eur	gross	6,187	1;2	3;9	27.67	--	--	--
	2010	at10	0.765eur	gross	6,187	1;2	3;9	21.18	--	--	--
Belgium	1995	be95	46.406bef	net	2,986	1	3	10.30	13,777	26.6	26.6
	2000	be00	40.355bef	net	2,697	1	3	10.30	19,540	27.9	31.6
Brazil	2006	br06	1.064brl	gross	118,590	1;2	5;27	2.66	9,963*	48.7	50.2
	2009	br09	1.189brl	gross	121,163	1;2	5;27	4.55	11,366*	46.3	47.4
	2011	br11	1.310brl	gross	111,986	1;2	5;27	6.84	12,170*	45.5	46.5
Canada	2004	ca04	1.272cad	gross	27,820	1;2	5;10	25.20	25,901	31.7	31.9
	2007	ca07	1.147cad	gross	26,745	1;2	5;10	28.20	33,241	31.4	31.8
	2010	ca10	1.109cad	gross	25,019	1;2	5;10	32.70	36,818	31.6	31.7
Colombia	2010	co10	1,009.805cop	gross	18,626	1;2	5;24	1.90	7,559*	48.3	51.9
	2013	co13	1,017.203cop	gross	20,631	1;2	5;24	2.09	~15,700*	49.1	--
Czech R.	2004	cz04	14.745czk	gross	4,351	2;3	8;14	35.20	11,851	26.6	26.7
	2007	cz07	13.177czk	gross	11,294	2;3	8;14	16.90	23,057	25.2	--
	2010	cz10	12.681czk	gross	8,865	2;3	8;14	17.15	27,007	25.6	--
Domin.R.	2007	do07	14.836dop	gross	8,363	3	32	10.30	12,398*	--	--
Egypt	2012	eg12	1.570egp	net	12,060	1	6	4.36	14,528	46.4	--
Estonia	2004	ee04	7.477EEK	gross	4,169	2;3	5;16	34.10	8,948	34.6	34.7
	2007	ee07	9.026EEK	gross	4,744	2;3	5;16	37.40	12,588	31.5	31.5
	2010	ee10	12.885EEK	gross	4,993	2;3	5;16	26.00	9,072	32.5	32.5
Finland	2007	fi07	0.889eur	gross	10,472	2;3	5;20	18.22	27,706	26.4	26.7
	2010	fi10	0.828eur	gross	9,351	2;3	5;20	18.10	32,064	26.1	26.3
France	2005	fr05	0.923eur	mixed	10,240	1	8	31.30	21,198	27.9	28.0
	2010	fr10	0.779eur	mixed	15,797	1	13	22.40	30,824	28.8	29.0
Germany	2007	de07	0.785eur	gross	10,921	1;2	2;16	12.93	28,227	28.8	29.4
	2010	de10	0.724eur	gross	12,146	1;2	2;16	10.65	31,687	28.6	28.8
Greece	2004	gr04	0.718eur	net	5,568	1;2	4;13	22.39	18,067	32.5	32.5
	2007	gr07	0.679eur	gross	6,504	1;2	4;13	11.74	22,313	31.5	31.6
	2010	gr10	0.638eur	gross	6,029	1;2	4;13	29.16	23,969	32.9	33.1
Guatem.	2006	gt06	2.808gtq	gross	13,686	2	22	3.67	7,429**	49.0	50.8
Hungary	2007	hu07	124.107huf	gross	2,024	1;2;3	3;7;18	15.56	16,921	27.4	--
	2012	hu12	111.905huf	gross	2,061	1;2;3	3;7;18	23.71	21,447	28.9	--
Iceland	2004	is04	97.242isk	gross	2,928	1 <sup>c</sup>	2	24.87	26,727	25.4	25.6
	2007	is07	106.882isk	gross	2,887	1	2	26.18	34,704	27.6	27.9
	2010	is10	119.817isk	gross	3,018	1	2	24.83	32,454	24.5	24.6
Ireland	2007	ie07	0.906eur	gross	5,247	2;3	2;8	23.75	33,522	29.7	30.2
	2010	ie10	0.766eur	gross	4,333	2;3	2;8	19.58	33,134	29.4	29.4
Israel	2010	il10	3.609ils	gross	6,168	2;3	7;16	12.12	23,970	37.8	38.6
Italy	2008	it08	0.731eur	mixed	7,977	1;2	5;20	43.90	24,526	32.4	32.7

	2010	it10	0.709eur	mixed	7,951	1;2	5;20	47.30	24,797	32.7	32.7
	2012	it12	0.668eur	mixed	8,151	1;2	5;20	47.40	--	--	--
Japan	2008	jp08	108.300jpy	gross	4,022	1	8	58.34	30,730	30.2	30.2
Mexico	2002	mx02	7.089mxn	net	17,167	3	32	13.53	5,705*	46.8	47.5
	2004	mx04	7.401mxn	net	22,595	3	32	11.68	6,291*	45.7	47.1
	2008	mx08	6.923mxn	net	29,468	3	32	17.73	10,657*	45.1	46.5
	2010	mx10	6.970mxn	net	27,655	3	33	10.05	10,396*	43.2	43.9
	2012	mx12	6.935mxn	net	9,002	3	33	11.83	17,282*	45.9	--
Peru	2004	pe04	1.393pen	net	19,502	1	8	17.74	5,078*	50.1	52.3
	2007	pe07	1.322pen	net	22,204	1	8	18.46	11,588*	50.0	--
	2010	pe10	1.342pen	net	21,496	1	8	19.11	14,057*	47.0	--
	2013	pe13	1.318pen	net	30,453	1	8	18.37	17,146*	45.5	--
Russia	2007	ru07	13.216rub	net	3,933	1	8	56.80	9,752*	37.1	37.1
	2010	ru10	14.372rub	net	6,323	1	8	56.70	15,111*	35.3	35.7
	2013	ru13	15.880rub	net	6,147	1	8	61.00	29,139	33.1	--
Serbia	2006	rs06	27.186rsd	gross	4,560	1	4	5.00	13,222*	34.3	--
	2010	rs10	32.383rsd	gross	4,585	1	3	4.50	16,373*	32.4	--
	2013	rs13	35.817rsd	gross	4,517	1	3	5.90	17,434*	33.2	--
Slovakia	2004	sk04	17.390skk	gross	5,147	2	4	14.44	8,769	26.8	26.8
	2007	sk07	17.967skk	gross	5,450	2	4	12.09	11,939	24.7	24.7
	2010	sk10	0.464eur	gross	5,200	2	4	10.36	18,136	26.3	26.5
S. Africa	2008	za08	3.777zar	gross	7,302	2	9	31.00	8,289*	59.4	65.5
	2010	za10	4.152zar	gross	6,809	2	9	8.82	7,450*	59.1	61.0
	2012	za12	4.328zar	gross	8,040	2	9	9.05	13,224*	57.2	--
Spain	2004	es04	0.783eur	net	12,996	1;2	7;18	18.41	17,929	31.6	31.6
	2007	es07	0.688eur	gross	13,014	1;2	7;18 <sup>d</sup>	9.49	24,480	30.4	30.4
	2010	es10	0.652eur	gross	13,109	1;2	7;18 <sup>d</sup>	12.51	25,151	32.8	32.8
	2013	es13	0.586eur	gross	11,965	1;2	7;18 <sup>d</sup>	26.17	47,277	34.3	--
Sweden	2000	se00	10.257sek	gross	14,491	2;3	8;21	26.40	16,446	25.1	25.6
	2005	se05	9.378sek	gross	16,268	2;3	8;21	32.75	21,765	23.6	23.8
Switz.	2002	ch02	1.916chf	gross	3,726	1	7	70.38	27,223	27.0	27.1
	2004	ch04	1.810chf	gross	3,270	1	7	67.72	27,824	26.3	26.3
UK	2004	uk04	0.653gbp	gross	28,041	1	12	32.40	26,107	34.1	35.1
	2007	uk07	0.610gbp	gross	24,977	1	12	32.49	31,138	33.6	33.9
	2010	uk10	0.628gbp	gross	25,350	1	12	41.00	32,468	33.3	33.6
	2012	uk12	0.608gbp	gross	20,201	1	12	40.00	--	--	--
US	1979	us79	2.260usd	gross	65,238	1;2	9;51	4.30	4,381	30.9 <sup>e</sup>	30.9
	1986	us86	1.575usd	gross	58,258	1;2	9;51	5.43	9,877	33.9	33.9
	1991	us91	1.333usd	gross	59,219	1;2	9;51	5.02	14,688	34.5	34.5
	1994	us94	1.247usd	gross	56,941	1;2	9;51	4.53	17,846	36.0	36.9
	1997	us97	1.179usd	gross	50,348	1;2	9;51	7.79	21,560	36.0	37.2
	2000	us00	1.123usd	gross	78,054	1;2	9;51	8.03	25,879	35.6	36.9

	2004	us04	1.032usd	gross	76,447	1;2	9;51	9.01	31,295	36.3	37.3
	2007	us07	0.945usd	gross	75,872	1;2	9;51	7.82	38,448	36.9	37.7
	2010	us10	0.909usd	gross	75,188	1;2	9;51	8.12	40,338	36.6	36.9
	2013	us13	0.862usd	gross	51,498	1;2	9;51	10.26	46,785	38.2	39.2
Uruguay	2004	uy04	11.706uyu	net	18,392	1;2	5;19	4.85	7,490*	43.9	45.0
	2007	uy07	11.516uyu	net	49,136	1;2	5;19	1.93	18,350	42.3	--
	2010	uy10	13.003uyu	net	46,550	1;2	5;19	9.62	23,726	40.5	--
	2013	uy13	14.931uyu	net	46,622	1;2	5;19	10.68	27,428	37.2	--

Notes: at10, co13, it12 and uk12 are included even though they are presently missing from the LIS database for the case that they are added in the future (sample sizes are only approximate). Ginis computed by the author from LIS database on 23 January 2015 except au10, eg12, mx12, pe07, pe10, pe13, ru13, uy07, uy10, uy13 and za12 computed on 13 October 2015. Summary statistics account for household sampling weights and household size. Incomes converted to year-end 2005 international dollars, purchasing power parity, using currency conversion factors and US GDP-deflator inflation rate (World Bank 2015a,b). For Belgium ('95, '00), Estonia ('04-'10) and Slovakia ('04, '07), currency conversion factors are taken from European Central Bank exchange list archive. Mean incomes may not be representative of those for the entire countries, as they omit non-responding households.

\* – classified by LIS as upper-middle; \*\* – lower-middle; rest – high income country.

<sup>a</sup> Gini (LIS) is computed using LIS method: Keep only nonzero disposable incomes and weights; censor small disposable incomes per capita at  $0.01 \times$  mean disposable income per capita; censor high disposable household incomes at  $10 \times$  median disposable household income, prior to dividing by adult equivalence scale; Adult equivalence scale is square root of household members; for analytical weight, count of household members is used. Results may differ from statistics reported by LIS because an older version of data may have been used.

<sup>b</sup> LIS method is partly adopted: Keep only disposable incomes of \$1 or greater, and positive weights; no top/bottom coding is performed; Adult equivalence scale is square root of household members; for analytical weight, count of household members is used. For clarity, Ginis and their jack-knife estimated standard errors are multiplied by 100.

<sup>c</sup> Only information on thinly vs. densely populated areas is available for Iceland. For Finland, only nonresponse rates at greater geographic level (NUTS2, 5 regions) are available.

<sup>d</sup> Out of 19: Ciudad Autónoma de Ceuta and Ciudad Autónoma de Melilla must be treated as one region as sample design (stratification and reporting of nonresponse rates) dictates.

<sup>e</sup> computed for gross household income because disposable income is missing.

### *Cross-country analysis*

Beside correcting national estimates of Gini coefficients using country-specific analyses, we attempt to take advantage of harmonization among LIS-database surveys to comment on worldwide inequality and biases to it introduced by unit nonresponses. Concurrent use of multiple LIS surveys facilitates a richer analysis on a larger sample both in terms of regions as well as potentially in terms of households within regions. This richness comes at a cost, however. Different national statistical agencies use different sample designs and different definitions of incomes, household weights and nonresponse rates. Some agencies apply methods such as hard top or bottom coding, imputation of values, trimming of sampling weights, or correcting for unit nonresponse in sampling weights. Different national surveys are also subject to various statistical issues, and the incomes they report on are subject to different geographic nature of inequality and distribution of nonresponse rates.

The presence of 25 European countries (including Denmark, Luxembourg, Netherlands, Norway, Poland and Russia) in the sample deserves a note. For ten European Union member states (Austria, Czech Rep., Estonia, Finland, Greece, Ireland, Luxembourg, Netherlands, Slovakia,

Spain) plus Iceland, data come from the EU Statistics on Income and Living Conditions (EU-SILC) survey coordinated by European Commission's statistical office, Eurostat. EU-SILC covers the largest common market but also one of the most heterogeneous sets of member states, including some of the world's most affluent nations as well as middle-income transitional economies. Nonresponse rates range from 9.5 to 43.3 percent across EU-SILC surveys evaluated here. Nonresponse rates therefore vary substantially across countries, even though they vary less within countries (refer to appendix 1, table A1). For most of these countries, nonresponse rates have been made publicly available, and computed in a consistent way using Eurostat guidelines. They are thus comparable to each other across the different EU-SILC surveys. (Appendix 1 provides more information on nonresponse rates in EU-SILC surveys.)

Unlike other surveys, all EU-SILC surveys can be used jointly only subject to the behavioral assumption that the response probability of households across all countries is the same stable, simple function of their income (in 2005 PPP dollars). Income data in EU-SILC surveys are of high quality, harmonized across countries, and not subject to modification by statistical agencies such as top-coding. One challenge specific to European countries is that incomes exhibit substantial cross-nation inequality, but relatively less inequality within nations. This affects the relative performance of analysis conducted on alternative levels of geographic disaggregation. Specifically, with little overlap between income distributions in the richest and the poorest member states, if the method proposed in the previous section were run at the level of states (rather than within-state regions), it would effectively adjust the mass of entire member states in the calculation of the Gini. The majority of households in rich states would be assigned a higher weight than the majority of households in poor states.<sup>12</sup>

Another notable set of surveys are the many waves of the US CPS. CPS has been collected systematically for over fifty years, in a consistent format with only infrequent minor changes since 1989.<sup>13</sup> Income distribution in the US CPS has also been consistent across years, with only a moderate steady drift in mean incomes and inequality. Incomes exhibit substantial inequality even within geographic regions, but relatively less of cross-region inequality. US Bureau of the Census applies winsorizing at the level of individual income components by the way of rank proximity swapping. Nonresponse rates are reported for each drawn household, and three types of nonresponse rates are distinguished. Type A includes refusals to be interviewed and failed interviews because nobody was at home. The following analysis takes only note of type A nonresponse. Omitted from the calculation are types B and C nonresponders, that is vacant or

---

<sup>12</sup> Other idiosyncratic problems with the EU-SILC data are that 1) Sampling weights in the EU-SILC are distributed very widely. This suggests that comparing unweighted, EU-SILC weighted, and our nonresponse-probability weighted income distributions may yield very different statistics, much more so than in other countries (for instance, incomes in the US Current Population Survey exhibit a high degree of inequality within rather than across states); 2) EU-SILC sampling weights are winsorized from below and from above to limit the extent to which individual observations can influence sample-wide statistics. This of course affects sample representativeness.

<sup>13</sup> This includes reclassification of income components and redesigning of relevant survey questions in 2011 and 2014 (Census 2015). This study utilizes all waves of US CPS in the LIS database for which state-level nonresponse rates can be computed, 1979–2013, and uses the revised dataset for 2011 (*us10i* rather than *us10ai*).

demolished housing units, and group quarters including boarding houses, transient residences (hotels), tents, trailer sites, barracks and other institutional residences.

In the following cross-country analysis, to the extent possible, we attempt to use concurrently surveys with comparable definitions of incomes, household weights and nonresponse rates. When alternative nonresponse rates are available, those comparable across years and to rates reported for other national surveys are used.<sup>14</sup> The following 28 national surveys will be used: Australia '10, Belgium '00, Brazil '11, Canada '10, Colombia '10, Czech Republic '04, Estonia '10, Finland '10, France '10, Germany '10, Greece '10, Guatemala '06, Iceland '10, Ireland '10, Israel '10, Italy '10, Japan '08, Mexico '10, Peru '04, Russia '10, Slovakia '10, South Africa '10, Spain '10, Sweden '05, Switzerland '04, UK '10, USA '10, Uruguay '04.<sup>15</sup>

All but six of these 28 selected surveys report gross incomes (including Italy reporting 'mixed incomes' at the level of income components, but accounting for taxes and contributions in total incomes), while the other six report incomes net of taxes and contributions. Twenty of these surveys come from years 2010-2011, seven from 2004-2008, and one from 2000, representing a comparable set of income distributions that may be viewed as informative of worldwide inequality during 2004–2011.

These 28 surveys come from all major world regions (Africa – 1, Pacific – 1, Central Eurasia – 1, East Asia – 1, Europe excl. Russia – 15, Middle East – 1, North America – 3, and Latin America excl. Mexico – 5). Twenty of them are for high income countries, seven for upper-middle income countries, and one for a lower-middle income country. In addition, some French territories are middle income regions. While this selection of surveys clearly oversamples richer – and particularly richer European – countries, data limitations simply do not allow broader or more balanced coverage. To mitigate this issue, stratification by country income bracket is applied when performing inference to other high and middle income countries.

#### *Correcting inequality estimates in Europe, North and Latin Americas, and OECD countries*

These selected surveys allow us to estimate inequality in several major world regions as well as worldwide.<sup>16</sup> For Europe, we observe sixteen countries. (Appendix 2 details how world regions and income groups of countries are delineated; Table A3 reports population of world regions.) In an attempt to extrapolate results from our selection of countries to the missing countries, we assign weights to groups of countries in different income brackets – high and upper-middle income countries – inversely proportional to the fraction of working-age population covered in

---

<sup>14</sup> Some agencies report both cross-sectional and longitudinal components of nonresponse (e.g., Russia, South Africa). Some agencies report nonresponse rates by survey rotational group.

<sup>15</sup> Six additional countries, with nonresponse rates known only at the national level, will be used in a supplemental analysis: *at04*, *kr06*, *lu10*, *nl10*, *pl10* and *ro97*.

<sup>16</sup> A similar exercise was performed for the set of OECD countries, recognizing that most research of inequality (Atkinson et al. 1995, 2011) deals with them. Since this is not a clear geographic bloc, these results are omitted but are available on request.

the selected surveys to Europe-wide population in countries in the same income-brackets (i.e., subsampling rate). Specifically, in high income countries Europe-wide, working-age population was estimated at 324.7 million in 2014.<sup>17</sup> In upper-middle income countries working-age population was estimated at 173.4 million. For the fifteen European high income countries in the LIS database, working-age population of 251.7 million, this yields a subsampling rate of 0.77509. For the one upper-middle income territory in the LIS database – Russia’s five European regions, population 79.0 million – the subsampling rate is 0.45544. There are no lower-middle or low income countries in Europe.

A related analysis is performed using only European Union Single Market countries (plus the city states of Andorra, Monaco and San Marino), that is, all of the above countries with the exception of former Soviet Union and Baltic states. This group of 35 countries – fifteen of them covered in the LIS database – are all high income countries with similar economic systems, interlinked economies, and similar survey design methods. This analysis is expected to yield slightly different results from the analysis that included Russia. The results can plausibly be extrapolated with greater confidence to other Single Market high income countries. The subsampling rate used for extrapolation is the same as for Europe-wide high-income countries in the previous paragraph, 0.77509.

For North America, we limit our attention to Canada, Mexico and the US, and we observe all three of them in the LIS database. For Latin America (excluding Mexico, and excluding low-income Haiti, to which inference cannot be made), we observe six states, including Brazil, Colombia, French Guiana (as a region in the French 2010 survey), Guatemala, Peru and Uruguay. To extrapolate to the missing countries, we again assign weights inversely proportional to the fraction of working-age population covered in our sample. Inference cannot be made to high income or low income countries, because the LIS sample does not include any such countries.<sup>18</sup> In upper-middle income countries throughout Latin America, working-age population was estimated at 234.7 million (29.0 mil. in lower-middle income countries). For the five Latin American upper-middle income countries in the LIS database – Brazil, Colombia, French Guiana, Peru and Uruguay (pop. 197.3 mil.) – this yields a subsampling rate of 0.84067. Finally, for the only Latin American lower-middle income country in the LIS database, Guatemala (pop. 9.3 mil.), this yields a subsampling rate of 0.32059.

The last bloc of countries for comparison is the member states of the OECD. The LIS database includes surveys for 30 out of the 35 members. For the other five, information from similar comparison groups is used. Namely, for Chile, Latvia, New Zealand, Portugal and Turkey,

---

<sup>17</sup> Working-age population (15–64) is considered for consistency with the adult equivalent scale (square root of household members) used in the analysis. Percent of population aged 15–64 is unavailable for Andorra, American Samoa, Bermuda, Cayman Islands, Dominica, Faeroe Islands, Greenland, Isle of Man, Liechtenstein, Marshall Islands, Monaco, Northern Mariana Islands, Palau, San Marino, Sint Maarten (Netherlands), St. Kitts and Nevis, St. Martin (France), Turks and Caicos Islands, and Tuvalu. For these territories, the 2014 median of 65.446% is used.

<sup>18</sup> While Uruguay has recently made the transition from upper-middle to high income status, our sample contains only pre-transition, year 2004 data.

Mexican, Estonian, Australian, Spanish and Greek surveys are used, respectively. These pairs of OECD member countries are similar in the level and inequality of incomes. Population sampling weights of households in mx10, ee10, au10, es10, gr10 are thus scaled up to cover the population of even the comparison countries. The scaling factors are 1.1975 for au10, 2.5754 for ee10, 1.2270 for es10, 7.4832 for gr10, and 1.1434 for mx10.

#### *Correcting inequality estimates worldwide*

For the analysis of worldwide inequality, to extrapolate results to the entire industrialized and middle-income world, weights are similarly assigned to the three respective income groups of countries – high, upper-middle and lower-middle income countries – inversely proportional to the fraction of working-age population covered in the selected surveys to population in countries in the same income-brackets worldwide. Specifically, in high income countries worldwide, working-age population was estimated at 828.4 million in 2014 (1,775.6 mil. in upper-middle, and 1,829.5 mil. in lower-middle income countries). For the twenty high income countries in the LIS database (work.-age pop. 586.8 mil.), this yields a subsampling rate of 0.70838. For the seven upper-middle income countries (pop. 416.0 mil.), the subsampling rate is 0.23429. For the single lower-middle income country, Guatemala, the subsampling rate is 0.00509.

#### *Utilizing additional country-level nonresponse rates*

The exercise described in the previous paragraph was limited by the number of LIS surveys with information on subnational geographic regions and external information on nonresponse rates for these same regions. In order to increase geographic coverage of our sample, ten surveys for which overall national nonresponse rate is reported are added.<sup>19</sup> For the other 28 countries, nonresponse rates at the sub-national regional level are used. Using this sample of 38 national surveys, the subsampling rates are 0.82395 for high income countries (0.24186 for upper-middle and 0.00509 for lower-middle income countries).

## 4. Results

Table 2 presents the benchmark results of this study, correcting distribution of incomes in the 66 household surveys (for which subnational nonresponse rates are available) for unit-nonresponse

---

<sup>19</sup> *at04* has mean income 25,006 and Gini 26.8; *dk10* has mean income 35,472 and Gini 25.2; *hu05* has mean income 9,490 and Gini 29.1; *kr06* has mean income 25,093 and Gini 30.4; *lu10* has mean income 49,244 and Gini 26.8; *nl10* has mean income 32,465 and Gini 25.4; *no10* has mean income 41,822 and Gini 24.8; *pl10* has mean income 15,311 and Gini 30.9; *ro97* has mean income 2,457 and Gini 28.4; *si10* has mean income 20,726 and Gini 25.2. With these additional 9 high and 1 upper-middle income countries, the respective worldwide income-bracket subsampling rates change to 0.824 and 0.242.

For *at04*, nonresponse rate is 37.30% (5,164 accepted/8,236 drawn). For *dk10* and *no10*, nonresponse rate is 0.00% (fully national-register data). For *hu05*, nonresponse rate is 28.00%. For *kr06*, average monthly nonresponse rate is 20.00%. For *lu10*, nonresponse rate for the cross-sectional sample is 43.34%. For *nl10*, nonresponse rate is 19.30% (EU-SILC information). For *pl10*, overall-sample no-contact or refusal rate is 28.10%. For *ro97*, nonresponse rate among sampled households is 9.60%. For *si10*, nonresponse rate is 35.30%.

using cross-region information. Following the lead of Korinek *et al.*'s (2006, 2007) and Hlasny and Verme's (2013, 2015), these models estimate survey-response probability as a simple logistic function of the logarithm of income per capita. Logarithmic specification allows the relationship between income and response probability to be highly nonlinear, with the response rate dropping rapidly in the highest range of incomes as one may expect.

The main finding in table 2 is that households' survey response probability is consistently related negatively to income per capita. The coefficients on income  $E(\theta_2)$  are estimated to be negative in all but four surveys,<sup>20</sup> and highly significant in the vast majority of surveys.

Reweighting households by the inverse of their estimated response probability (mean reported in column 5 in table 2) allows us to correct measures of inequality for the differential probability of rich and poor households to respond to the survey. The corresponding corrected Ginis are systematically higher than the uncorrected statistics in table 1. The correction to Ginis computed on LIS-winsorized income distributions is consistently positive – and the estimated bias is thus negative – with only four instances where the estimated bias is insignificant positive (*ca04–ca10* and *za08*).<sup>21,22</sup> The negative bias is significant in 41 of the 66 surveys, and ranges from 0.0 to 16.6 percentage points (average 3.6; median 2.0).

Using non-winsorized income distributions, the estimated biases due to unit nonresponse are consistently negative, with only three insignificant exceptions (*ca04–ca10*). The negative bias is significant in over one-half of all surveys, and ranges from 0.0 to 29.1 percentage points, with a large bulk of estimates in the 1–8 percentage-point range (average 5.9; median 3.7). The estimated bias is thus distributed widely across surveys. This is an artifact of the distribution of incomes and regional nonresponse rates, and the degree of geographic disaggregation of surveys. However, for countries where prior estimates of the bias exist – EU member states (Hlasny and Verme 2015), the UK (Jenkins 2016) and the US (Mistiaen & Ravallion 2003; Korinek *et al.* 2006, 2007) – these estimated biases are near or below those found previously, verifying their robustness.

While Ginis estimated on non-winsorized income distributions appear plausible for most national surveys, in Peru and Sweden they appear off the mark. Winsorizing of incomes leads to more plausible results in these surveys. This may be the result of particular statistical issues or top-income values in these surveys, and it serves as a warning that unit nonresponse is just one

---

<sup>20</sup> *au10*, *mx04*, *ch02* and *ch04*, of which only in *ch02* the positive coefficient is significantly different from zero. In these four countries, it is the lowest-income households who apparently have a lower tendency to respond than the highest-income households.

<sup>21</sup> Asterisks on the estimated biases indicate significance at the 5% level using jack-knife standard errors of uncorrected & corrected estimates, ignoring covariance between the estimates. The bias is just insignificant at the 5% level in the case of *ca04*, *ca07* and *ch02*.

<sup>22</sup> In *ca10*, 193 households have a missing regional identifier. In an alternative specification, these are classified as the 11<sup>th</sup> Canadian region and assigned the national average nonresponse rate of 32.7%. Results in this alternative specification are equivalent to those reported in table 2. A similar issue of unknown regions occurs to a lesser degree even in other surveys.

problem present in household income surveys, and that its correction is sensitive to the presence of nonrepresentative income values in the top end of income distributions and to other data issues.

Simple Pearson correlation between Gini corrections in LIS-winsorized versus non-winsorized income distributions is +0.81, implying imperfect correspondence between them, even accounting for the fact that the corrections in LIS-winsorized data are distributed more narrowly. Correlation between the bias correction (or the bias with a negative sign) and survey characteristics reveals that the correction is not related to log mean incomes (corr. +0.024) or mean count of regions (corr. -0.062), but is associated negatively with uncorrected Ginis (corr. -0.122 in non-winsorized data, -0.227 in winsorized data) and with mean sample-size per region (corr. -0.188), and positively with mean nonresponse rates (corr. +0.142).<sup>23</sup> There are no apparent systematic differences in the bias correction across world regions (figures A3 and A4 in appendix 3).

Table 3 builds on the results in table 2 by reporting on models estimated across national borders. Surveys from countries in North America, Latin America, Europe, EU Single Market and worldwide, respectively, are used concurrently here. Single sets of coefficients, applicable to all member countries are estimated, representing a major restriction on the models. The first two columns in table 3 report estimates of these universal coefficients. Agreeing with table 2, table 3 shows that across the world, households' probability of responding is related systematically negatively to their income. All estimates of  $\theta_2$  are negative and highly significant. To allow for differences in income levels and definitions of nonresponse rates across countries, country binary indicators are included in the models. Their coefficients are presented in appendix 3, table A4.

Under a stylized assumption that excluded countries have a similar distribution of incomes and of nonresponse rates, and a similar relationship between households' income and response probability, we project the results to the rest of the respective world regions (except for North

---

<sup>23</sup> To evaluate the claims regarding the role of geographic disaggregation more formally, we re-run the models for *cz04*, *es10*, *il10*, *se05* and *us10*, because these surveys have an adequate number (i.e., 7–9) of regions even at a more-aggregated level. These counts matter because they affect the effective regional distributions of incomes, and dictate the size and number of errors to be minimized in estimation.

In *cz04*, the 14 regions are aggregated as follows: 11; 21; 31+32, 41+42; 51+52+53; 61+62; 71+72; 81. In *es10*, the 18 regions are aggregated as follows: 11+12+13; 21+22+23+24; 30; 41+42+43; 51+52+53; 61+62+(63+64); 70. In *il10*, the 16 regions are aggregated as follows: 11; 21+23+24; 31+32; 41+42+43+44; 51+52+53; 61+62; 70. In *se05*, the 21 regions are aggregated as follows: 110; 121+122+123+124+125; 211+212+213+214; 221+224; 231+232; 311+312+313; 321+322; 331+332. In *us10*, the 51 regions are aggregated as follows: 11+12+13+14+15+16; 21+22+23; 31+32+33+34+35; 41+42+43+44+45+46+47; 51+52+53+54+55+56+57+58+59; 61+62+63+64; 71+72+ 73+74; 81+82+83+84+85+86+87+88; 91+92+93+94+95.

This alternative estimation reveals that the more detailed the degree of geographic disaggregation, the smaller the estimated bias due to unit nonresponse. All estimates of  $\theta_2$  are still negative, but standard errors on  $E(\theta_1)$  and  $E(\theta_2)$  are larger. The corrected estimates of Ginis are: in *cz04* it is 37.99 (s.e. 3.95); in *es10* it is 32.98 (s.e. 0.28); in *il10* it is 39.32 (s.e. 0.94); in *se05* it is 44.98 (s.e. 20.13); in *us10* it is 41.29 (s.e. 0.51). These estimates are consistently higher than those reported in table 2 with the exception of *se05*, where they are insignificantly just smaller than those in table 2. This implies a higher degree of geographic aggregation yields a consistently higher correction for nonresponse, confirming the findings of Hlasny and Verme (2013, 2015).

America where all countries are present in the LIS database, and EU, where the vast majority are present). In North America, the Gini appears quite robust to nonresponse, rising by merely 0.15–0.34 percentage points, to 44.28–44.72, above the uncorrected statistics for winsorized and non-winsorized income distributions. Gini coefficient for Latin America (excluding low income countries), however, is quite sensitive to the correction, rising from 49.12 to 61.36, or 12.24 percentage points (from 47.78 to 52.06, or 4.28 pc.pt., in LIS-winsorized data). In Europe, the correction is also substantial and statistically significant. Across all Europe including former Soviet-Union and Balkan nations, the Gini rises from 35.25 to 42.97, or 7.71 percentage points (from 35.05 to 38.99, or 3.94 pc.pt., in LIS-winsorized data). Limiting the analysis to European Single Market plus European city-states (Andorra, Monaco and San Marino), the Gini rises from 32.11 to 37.63, or 5.52 percentage points (from 31.90 to 34.89, or 2.99 pc.pt., in LIS-winsorized data).

**Table 2. Gini correction for unit non-response using within-country regional response rates**

	$E(\theta_1)$	$E(\theta_2)$	$\sigma^2$	Regns	Avg.wt	Corrected Gini (L) <sup>a</sup>	Bias	Corrected Gini <sup>b</sup>	Bias
au10	-9.605 (0.677)	1.153 (0.094)	12.166	7	1.296	38.714 (1.237)	-5.9*	38.998 (1.270)	-5.9*
be95	13.775 (1.195)	-1.216 (0.123)	0.007	3	1.115	27.836 (0.965)	-1.2	27.836 (0.965)	-1.2
be00	4.179 (36.892)	-0.194 (3.789)	1.473	3	1.101	28.044 (0.902)	-0.1	32.106 (4.612)	-0.5
br06	5.223 (8.645)	-0.186 (0.961)	0.769	27	1.028	48.713 (0.115)	-0.0	50.269 (0.170)	-0.1
br09	8.998 (15.237)	-0.654 (1.596)	4.749	27	1.052	46.796 (0.117)	-0.5*	48.196 (0.207)	-0.8*
br11	5.595 (10.992)	-0.327 (1.175)	5.227	27	1.074	45.722 (0.126)	-0.2	46.756 (0.172)	-0.3
ca04	9.822 (1.652)	-0.870 (0.161)	2.763	10	1.355	30.833 (0.305)	0.9	30.833 (0.305)	1.1
ca07	9.728 (1.848)	-0.854 (0.176)	4.266	10	1.413	30.458 (0.274)	1.0	30.459 (0.274)	1.3
ca10	12.941 (1.645)	-1.166 (0.153)	5.132	10	1.510	31.258 (0.470)	0.4	31.258 (0.470)	0.5
co10	11.250 (2.252)	-0.807 (0.238)	0.078	23	1.018	48.610 (0.497)	-0.3	54.375 (2.180)	-2.5
cz04	13.949 (1.345)	-1.425 (0.140)	4.849	14	1.539	33.726 (2.464)	-7.1*	34.372 (2.814)	-7.7*
ee04	19.062 (1.000)	-2.011 (0.103)	5.696	16	1.556	41.479 (2.261)	-6.9*	42.199 (2.918)	-7.5*
ee07	15.380 (1.303)	-1.570 (0.134)	5.298	16	1.556	35.043 (2.694)	-3.5	35.213 (2.858)	-3.7
ee10	20.846 (0.598)	-2.148 (0.061)	0.962	16	1.343	37.191 (1.277)	-4.7*	37.191 (1.277)	-4.7*
fi07	16.149 (0.306)	-1.412 (0.029)	0.063	5	1.223	30.907 (1.469)	-4.5*	33.213 (2.771)	-6.6*
fi10	14.930 (0.172)	-1.279 (0.016)	0.020	5	1.222	29.285 (0.973)	-3.2*	31.088 (1.716)	-4.8*
fr05	8.561 (3.720)	-0.780 (0.369)	10.158	8	1.454	29.205 (0.442)	-1.3*	29.452 (0.525)	-1.4*
fr10	12.167 (2.982)	-1.065 (0.283)	2.879	13	1.282	31.629 (0.750)	-2.9*	34.072 (2.897)	-5.1
de07	12.294 (1.168)	-1.007 (0.111)	0.313	16	1.148	30.746 (0.612)	-1.9*	33.194 (1.764)	-3.8*
de10	8.448 (1.448)	-0.612 (0.139)	0.148	16	1.119	28.956 (0.381)	-0.4	29.314 (0.482)	-0.5
gr04	22.658 (7.743)	-2.160 (0.740)	25.249	13	1.321	40.772 (1.902)	-8.3*	40.966 (2.007)	-8.5*
gr07	32.414 (2.255)	-2.922 (0.200)	1.209	13	1.125	44.733 (3.254)	-13.3*	46.409 (4.082)	-14.8*
gr10	25.087 (2.260)	-2.364 (0.205)	32.966	13	1.472	49.519 (3.067)	-16.6*	51.069 (3.800)	-18.0*
gt06	20.314 (0.950)	-1.837 (0.086)	0.738	22	1.041	53.381 (1.168)	-4.4*	66.618 (8.696)	-15.8
is04	20.163 (0.024)	-1.842 (0.002)	0.000	2	1.331	39.636 (5.285)	-14.3*	45.069 (10.437)	-19.4
is07	13.830 (0.002)	-1.222 (0.000)	0.000	2	1.368	33.761 (1.845)	-6.2*	35.298 (2.493)	-7.4*
is10	19.638 (556.65)	-1.762 (51.49)	0.015	2	1.330	34.037 (3.731)	-9.6*	34.862 (4.090)	-10.3*
ie07	6.617 (0.554)	-0.530 (0.053)	1.007	8	1.302	30.382 (0.854)	-0.7	31.240 (1.479)	-1.0
ie10	6.213 (0.481)	-0.466 (0.046)	0.298	8	1.239	29.504 (0.540)	-0.1	29.504 (0.540)	-0.1
il10	5.390 (1.629)	-0.342 (0.162)	0.910	16	1.137	37.817 (0.462)	-0.0	38.901 (0.783)	-0.3
it08	19.161 (1.506)	-1.846 (0.139)	48.863	20	1.786	45.336 (2.865)	-13.0*	52.117 (6.460)	-19.4*
it10	13.129 (1.418)	-1.277 (0.136)	22.558	20	1.867	36.755 (1.166)	-4.0*	36.758 (1.166)	-4.0*
jp08	14.226 (5.330)	-1.400 (0.501)	10.774	8	2.379	34.576 (1.132)	-4.4*	34.576 (1.132)	-4.4*
mx02	2.157 (0.675)	-0.034 (0.082)	0.964	32	1.153	46.773 (0.563)	-0.0	47.508 (0.716)	-0.0
mx04	0.956 (0.986)	0.124 (0.121)	1.024	32	1.136	45.665 (0.352)	-0.0	47.053 (0.464)	-0.0
mx08	6.157 (1.997)	-0.504 (0.215)	3.312	32	1.211	45.792 (0.330)	-0.7*	48.279 (0.817)	-1.8*
mx10	3.202 (1.953)	-0.111 (0.219)	1.355	32	1.109	43.233 (0.306)	-0.0	43.975 (0.371)	-0.0
pe04	18.108 (1.227)	-1.879 (0.120)	5.410	8	1.230	57.900 (0.609)	-7.8*	71.850 (3.840)	-19.5*
ru07	16.886 (3.811)	-1.831 (0.390)	96.235	8	2.162	38.201 (1.248)	-1.1	38.201 (1.248)	-1.1
ru10	13.947 (3.218)	-1.460 (0.318)	81.138	8	2.204	42.322 (2.745)	-7.1*	52.782 (10.936)	-17.1
sk04	7.434 (1.165)	-0.629 (0.129)	0.164	4	1.167	27.277 (0.561)	-0.5	27.303 (0.569)	-0.5
sk07	10.606 (3.359)	-0.926 (0.356)	0.706	4	1.138	25.146 (0.398)	-0.4	25.146 (0.398)	-0.4
sk10	16.784 (4.019)	-1.494 (0.394)	2.017	4	1.128	30.157 (2.482)	-3.8	33.572 (5.739)	-7.1
za08	6.856 (0.343)	-0.707 (0.037)	5.247	9	1.462	58.912 (0.466)	0.5	69.127 (1.659)	-3.6*
za10	15.647 (1.169)	-1.479 (0.117)	0.970	9	1.094	61.181 (0.831)	-2.1*	66.670 (1.871)	-5.7*
es04	31.462 (1.056)	-2.958 (0.094)	6.944	18	1.236	44.535 (4.536)	-13.0*	49.799 (9.652)	-18.2
es07	17.938 (2.220)	-1.548 (0.213)	0.566	18	1.105	31.173 (0.351)	-0.8*	31.173 (0.351)	-0.8*
es10	2.156 (3.072)	-0.024 (0.311)	2.032	18	1.147	32.787 (0.279)	-0.0	32.788 (0.279)	-0.0
se00	15.989 (1.003)	-1.526 (0.099)	1.686	21	1.339	33.739 (3.913)	-8.7*	54.652 (24.298)	-29.1
se05	8.754 (0.365)	-0.800 (0.035)	1.645	21	1.477	24.721 (0.385)	-1.1*	45.414 (20.554)	-21.6
ch02	-8.638 (0.373)	0.781 (0.038)	3.784	7	3.366	28.430 (0.708)	-1.4	28.458 (0.715)	-1.4
ch04	-1.785 (1.794)	0.105 (0.178)	10.255	7	3.053	26.337 (0.490)	-0.1	26.337 (0.490)	-0.1
uk04	7.179 (1.752)	-0.643 (0.171)	2.920	12	1.476	36.407 (0.393)	-2.3*	39.119 (0.943)	-4.0*
uk07	7.534 (3.085)	-0.667 (0.296)	8.884	12	1.480	35.441 (0.417)	-1.9*	36.275 (0.609)	-2.4*
uk10	10.319 (1.673)	-0.963 (0.158)	5.593	12	1.687	38.211 (0.578)	-4.9*	39.746 (0.945)	-6.1*
us79	22.744 (4.300)	-2.288 (0.477)	0.177	51	1.045	32.397 (0.253)	-1.5*	32.475 (0.283)	-1.6*
us86	16.961 (7.932)	-1.519 (0.825)	0.541	51	1.058	34.625 (0.166)	-0.7*	34.642 (0.168)	-0.7*
us91	23.696 (4.487)	-2.113 (0.435)	0.180	51	1.053	36.448 (0.226)	-2.0*	36.466 (0.229)	-2.0*
us94	20.361 (2.469)	-1.726 (0.227)	0.440	51	1.047	39.651 (0.452)	-3.6*	45.075 (1.266)	-8.2*

us97	16.216 (5.028)	-1.362 (0.476)	0.835	51	1.083	39.523 (0.412)	-3.6*	44.306 (0.974)	-7.1*
us00	16.752 (4.144)	-1.396 (0.384)	1.060	51	1.087	40.056 (0.403)	-4.4*	45.451 (0.923)	-8.5*
us04	21.617 (2.352)	-1.821 (0.206)	0.880	51	1.097	44.652 (0.643)	-8.4*	51.442 (1.132)	-14.2*
us07	14.521 (6.339)	-1.136 (0.582)	0.964	51	1.081	38.697 (0.240)	-1.8*	40.624 (0.416)	-2.9*
us10	12.263 (7.523)	-0.931 (0.696)	0.790	51	1.088	37.406 (0.177)	-0.8*	37.863 (0.218)	-1.0*
us13	11.483 (6.855)	-0.869 (0.627)	1.737	51	1.111	39.732 (0.274)	-1.5*	41.528 (0.415)	-2.3*
uy04	5.815 (2.104)	-0.323 (0.236)	0.143	19	1.052	43.997 (0.252)	-0.1	45.168 (0.363)	-0.2

$\sigma^2$  is the factor of proportionality, related to mean deviation of predicted to actual regional population, and computed as mean squared deviation of predicted to actual regional population, over mean squared regional population.

\* bias significant at the 5% level using standard errors of uncorrected & corrected estimates (ignoring covariance). For clarity, Ginis and their jack-knife estimated standard errors are multiplied by 100.

<sup>a</sup> Gini (L) is computed using LIS method: Keep only nonzero disposable incomes and weights; censor small disposable incomes per capita at 0.01×mean disposable income per capita; censor high disposable household incomes at 10×median disposable household income, prior to dividing by adult equivalence scale; Adult equivalence scale is square root of household members; for analytical weight, count of household members is used.

<sup>b</sup> LIS method is partly adopted: Keep only disposable incomes greater than 1 and non-negative weights; no top/bottom coding is performed; Adult equivalence scale is square root of household members; for analytical weight, count of household members is used.

**Table 3. Correction of world region Ginis using within-country regional response rates**

World Region	$E(\theta_1)$	$E(\theta_2)^a$	$\sigma^2$	Hhds.	Regns	Avg.non-resp.wt.	Gini (L) <sup>b</sup>	Corrected Gini (L)	Bias (L)	Gini <sup>c</sup>	Corrected Gini	Bias
North Am. (3)	12.125 (.028)	-0.918 (.003)	2.612	126,695	93	1.176	44.134 (.128)	44.284 (.148)	-0.150	44.381 (.139)	44.721 (.175)	-0.340
Latin Am. (6)	20.249 (.047)	-1.787 (.004)	5.700	173,569	100	1.075	47.783 (.118)	52.064 (.248)	-4.281*	49.117 (.192)	61.358 (1.972)	-12.241*
Europe (16)	12.339 (.018)	-1.301 (.002)	25.932	136,076	177	1.424	35.048 (.167)	38.989 (.514)	-3.941*	35.254 (.185)	42.965 (1.973)	-7.711*
EU Single Mkt. (15)	10.651 (.012)	-0.994 (.001)	10.043	132,283	172	1.434	31.897 (.145)	34.892 (.290)	-2.995*	32.107 (.167)	37.631 (1.698)	-5.524*
OECD (30)	18.507 (.026)	-1.500 (.002)	36.376	675,002	305	1.170	38.440 (.093)	41.580 (.280)	-3.140*	38.646 (.100)	44.577 (1.037)	-5.931*
World (28) <sup>d</sup>	15.596 (.013)	-1.237 (.001)	12.432	470,833	412	1.229	53.553 (.140)	54.014 (.300)	-0.461	54.238 (.190)	57.350 (1.240)	-3.112*
World (38) <sup>e</sup>	16.493 (.035)	-1.365 (.003)	73.220	891,271	422	1.232	52.947 (.140)	54.490 (.873)	-1.543*	53.638 (.193)	57.774 (1.440)	-4.136*

\* bias significant at the 5% level using standard errors of uncorrected & corrected estimates (ignoring covariance). For clarity, Ginis and their jack-knife estimated standard errors are multiplied by 100.

<sup>a</sup> Country binary indicators are also included, but are deferred to Appendix 3 to preserve space here. US is used as benchmark in North America, OECD and World (28) analyses, Brazil in Latin America, Russia in Europe, and UK in EU Single Market. In World (38), Island, Ireland, UK and US are used as benchmark.

<sup>b</sup> Gini (L) is computed using LIS method: Keep only nonzero disposable incomes and weights; censor small disposable incomes per capita at 0.01×mean disposable income per capita; censor high disposable household incomes at 10×median disposable household income, prior to dividing by adult equivalence scale; Adult equivalence scale is square root of household members; for analytical weight, count of household members is used.

<sup>c</sup> LIS method is partly adopted: Keep only disposable incomes greater than 1 and non-negative weights; no top/bottom coding is performed; Adult equivalence scale is square root of household members; for analytical weight, count of household members is used. For clarity, Ginis and their jack-knife estimated standard errors are multiplied by 100.

<sup>d</sup> 28 countries: *au10, be00, br11, ca10, ch04, co10, cz04, de10, ee10, es10, fi10, fr10, gr10, gt06, ie10, il10, is10, it10, jp08, mx10, pe04, ru10, se05, sk10, uk10, us10, uy04* and *za10*. Excluding low-income countries, for lack of data. Extrapolation to high, upper-middle and lower-middle countries done by augmenting household weights by author-computed subsampling rates of 70.838, 23.429 and 0.509%, respectively.

<sup>e</sup> 38 countries: *at04*, *dk10*, *hu05*, *kr06*, *lu10*, *nl10*, *no10*, *pl10*, *ro97* and *si10* added. For the first 28, subnational regional nonresponse rates are used. For the latter 10, national nonresponse rates are used – surveys are not disaggregated geographically. Excluding low-income countries, for lack of data. Extrapolation to high, upper-middle and lower-middle countries done by augmenting household weights by author-computed subsampling rates of 82.395, 24.186 and 0.509%, respectively.

Finally, correction for unit nonresponse also matters for estimating the worldwide Gini. In both model specifications – with 28 regionally disaggregated surveys only, or with 38 surveys including also ten non-disaggregated surveys – the corrected statistics (57.35 and 57.77) exceed the uncorrected values (53.64 and 54.24) significantly, by 3.11–4.14 percentage points.<sup>24</sup>

## 5. Conclusions

This study has re-evaluated income inequality in countries worldwide in the presence of systematic nonresponse by top-income households. The study has also proposed inequality estimates for world regions and for the entire industrialized and middle-income world. We have used a collection of 76 microdatasets from 38 high and middle-income countries across the world that were harmonized and made available by Luxembourg Income Study. In addition, we have custom-collected information on unit-nonresponse rates at the level of subnational regions, which is an essential input in the correction method used. These nonresponse rates are reported in the appendix for the benefit of the research community.

The study has found that the nonresponse bias affects national Gini coefficients typically by 1–8 percentage points, but sometimes by 15 percentage points or more. This is in line with prior estimates (Mistiaen & Ravallion 2003; Korinek et al. 2006, 2007). There are no clear differences in the estimated bias across world regions, or countries' income level. On the other hand, the degree of geographic disaggregation affects the bias systematically negatively, corroborating evidence in previous studies (Hlasny & Verme 2013, 2015). This phenomenon has not been explained satisfactorily as of yet and calls for further inquiry. Winsorizing performed by Luxembourg Income Study (on top of winsorizing performed by some national statistical agencies) affects negatively not only the original Gini estimates but also the estimated corrections for unit nonresponse (figure A5 in appendix 3). If we trust the uncensored values of top incomes as not systematically over-reported – plausible, as systematic underreporting is more probable – this would suggest that the unit-nonresponse bias is not likely to be identified and corrected fully in winsorized income distributions. Corrections for other specific measurement issues related to top incomes – such as winsorizing already performed by statistical agencies,

---

<sup>24</sup> An additional analysis was considered on non-disaggregated data, where each country is treated as a single region. This allows us to use a consistent degree of geographic aggregation across countries. Nevertheless, this specification is problematic because of heterogeneity in the definitions of incomes, weights and nonresponse rates across countries. The results of this model are not believed to be accurate, yet they provide a benchmark to which one may compare results in table 3. This model yields  $E(\theta_1)=28.613$ ,  $E(\theta_2)=-2.384$ ,  $\sigma^2=43.315$ , and avg. nonresponse weight 1.248. Compared to the uncorrected Gini of 54.58 (53.90 in LIS-winsorized data), the corrected Gini is 76.05 (62.67), or 21.47 pc.pt. (8.77) higher. This yet again shows that less geographically disaggregated data give rise to smaller corrections for unit nonresponse.

missing income components, random data-entry errors, etc. – may suffer the same fate. That said, winsorizing in the LIS database produces more plausible results in some national surveys, including Peru and Sweden. This may be the result of particular statistical issues or top-income values in these surveys, which may be assigned very high nonresponse-correction weights. The use of winsorizing should be decided on a case by case basis in view of data issues in the income distributions at hand.

The fact that the Gini coefficient was consistently corrected upward by 1–8 (or even 15+) percentage points is worrying since this measure of inequality is known to be less sensitive to changes in the tails, and most sensitive to perturbations in the middle of the income distribution. Other inequality measures may be affected far more. Table 4 shows the corrections to the top income shares in each country. The estimated income shares of the richest one percent of households often increase fivefold or even eight-fold, when the sampling weights of the richest households are corrected for unit nonresponse. Income shares of the following four, and nine percent of households (for the top 5, and 10% of households) are adjusted by a lower degree, and income shares of the households in the ninth decile often fall along with the income shares of the rest of the population. This confirms that the issue of unit nonresponse and the associated bias is by far most prevalent among households with the highest incomes. Table A5 in appendix 3 shows the same inequality measures – and equivalent results – for world regions at large. In addition, figure A6 reports Lorenz curves for world regions.

The other most important findings of this study relate to the estimation of Gini coefficients for world regions and worldwide. Using 38 surveys, encompassing 891,271 household records in 422 regions, we derived back-of-the-envelope estimates of the Ginis for world regions and worldwide. Gini coefficient for North America – of 44.72 after correction of unit nonresponse – appears quite robust to nonresponse, being merely 0.15–0.34 percentage points above the uncorrected statistic. Ginis for Europe and Latin America (excluding low income countries), however, are quite sensitive to the correction. In the European Single Market, the Gini rises from 32.11 to 37.63 (5.52 pc.pt.), and in Europe including former Soviet-Union and Balkan nations it rises from 35.25 to 42.97 (7.71 pc.pt.). In Latin America, the Gini rises from 49.12 to 61.36 (12.24 pc.pt.). Correction for unit nonresponse also matters for estimating the worldwide Gini. The corrected statistic, of 57.35–57.77, is 3.11–4.14 percentage points above the uncorrected value, 53.64–54.24.

Overall, these numbers from across countries, world regions and worldwide appear consistent with each other and with previous estimates, and suggest that unit nonresponse biases Gini coefficients downward by 1–8 percentage points. A number of robustness checks, related to the inclusion of various countries in the analysis or alternative disaggregation of national datasets into regions, confirm the direction and significance of the bias, and are generally consistent with the size of the bias reported in main models.

One limitation of this analysis was underrepresentation of surveys from middle-income and particularly low income countries. As a result, extrapolation to upper-middle and lower-middle income countries was speculative, and extrapolation to low income countries was impossible. The assumption used for extrapolation to the rest of world countries was clearly simplistic and most likely biased our estimates of worldwide inequality downward (Milanovic 2002b). This is because typically only richer countries among developing (lower-middle and upper-middle income) countries have the institutional capacity, willingness and adequate data quality to collaborate with LIS, while a large number of even poorer countries among industrialized (high income) countries collaborate. Coupled with the total exclusion of low income countries from this analysis, this implies that our Ginis for Latin America and the world are underestimates or lower bounds. External data on inequality in the omitted countries— something we viewed as out of the scope of the exercise performed here – would be needed to produce more believable results.

For the true worldwide bias due to nonresponse, our results may be viewed as upper-end estimates. To the extent that poorer countries typically have higher inequality, one may expect that they suffer higher nonresponse rates and greater biases due to them – since it is the far-off rich who are systematically missing from surveys. However, this does not appear to be the case, at least in the LIS sample. On the contrary, poorer countries and countries with higher income inequality tend to have lower nonresponse rates (figures A1 and A2 in appendix 1) and higher income inequality is associated with a lower bias estimate. To the extent that it is the households with high absolute levels of income – irrespective of the mean income in their country – that are less likely to respond to surveys, it is likely that the inclusion of the poorest countries would not affect the bias or would make it less influential compared to the inclusion of impoverished households and the jump in the level of inequality. Hence the finding that countries’ and world-regions’ Ginis should be corrected by 1–8 percentage points is the primary and robust conclusion of this study.

Table 4. Additional inequality measures: richest-household income shares

	Uncorrected income distributions					Unit-nonresponse corrected distributions				
	Top 0.5% inc. share	Top 1% share	Top 5% share	Top 10% share	Top 20% share	Top 0.5% share	Top 1% share	Top 5% share	Top 10% share	Top 20% share
au10	3.66	5.79	16.95	27.29	43.67	3.41	5.40	16.00	25.93	41.84
be95	2.92	4.79	14.87	24.56	40.48	4.98	7.20	18.26	28.14	43.96
be00	9.38	11.21	21.29	30.69	45.99	10.21	12.07	22.19	31.56	46.77
br06	6.00	9.55	26.46	39.01	55.43	6.09	9.67	26.68	39.26	55.67
br09	5.52	8.87	24.79	37.06	53.47	6.67	10.33	26.98	39.34	55.50
br11	5.37	8.70	24.37	36.39	52.74	5.78	9.27	25.38	37.50	53.80
ca04	2.50	4.39	15.08	25.26	41.65	3.99	6.63	19.33	30.39	47.16
ca07	2.49	4.35	15.00	25.20	41.62	4.11	6.68	19.43	30.59	47.45
ca10	2.45	4.35	15.07	25.30	41.70	5.33	8.69	22.93	34.48	51.25
co10	10.61	14.32	30.08	41.90	57.78	15.58	19.27	34.56	45.79	60.73
cz04	3.03	4.91	14.93	24.72	40.16	15.79	19.26	32.63	42.85	57.04
ee04	3.11	5.58	18.55	29.72	46.91	19.02	26.02	48.26	59.33	72.60
ee07	2.74	4.72	16.23	27.36	44.58	11.34	15.06	31.87	44.93	61.48

ee10	2.56	4.70	16.59	27.78	45.00	11.38	18.80	37.50	49.64	64.29
fi07	3.49	5.33	15.44	25.11	41.09	13.74	16.21	27.20	36.52	51.06
fi10	3.41	5.17	15.12	24.83	40.85	11.45	13.73	24.40	33.97	49.03
fr05	2.78	4.59	14.78	24.41	39.86	4.91	7.45	19.59	29.97	45.59
fr10	3.47	5.41	15.89	25.51	40.89	11.35	13.99	26.25	35.97	50.44
de07	4.10	6.04	16.85	26.78	42.10	9.88	12.19	23.70	33.48	48.01
de10	3.37	5.28	15.86	25.85	41.73	4.17	6.26	17.33	27.50	43.40
gr04	2.99	5.04	16.54	26.85	43.07	17.33	22.54	40.73	50.82	63.95
gr07	3.41	5.41	16.32	26.34	42.25	29.22	33.24	44.49	52.17	63.24
gr10	3.86	6.11	17.18	27.17	43.01	43.09	49.84	62.19	68.92	77.52
gt06	6.41	10.19	26.19	38.52	55.34	39.37	43.54	55.86	63.78	74.00
is04	3.73	5.71	15.69	24.99	40.27	33.93	37.37	47.09	54.11	64.68
is07	4.23	6.32	16.72	26.32	41.63	14.23	17.65	29.87	39.58	53.57
is10	3.19	4.93	14.51	23.92	39.44	19.86	22.61	33.26	42.05	55.29
ie07	3.82	5.56	15.98	25.56	41.99	5.44	7.46	18.85	28.79	45.36
ie10	2.40	3.91	14.28	24.21	40.51	2.85	4.56	15.60	25.94	42.51
il10	4.76	6.85	18.43	29.12	45.82	5.62	7.80	19.73	30.54	47.26
it08	3.58	5.69	16.97	27.13	43.04	38.30	44.33	58.70	66.50	76.26
it10	2.96	5.18	16.52	26.57	42.29	10.80	16.32	33.85	45.00	59.77
jp08	2.74	4.68	15.36	25.00	40.21	11.69	17.19	35.25	46.61	60.94
mx02	4.86	8.20	24.25	36.67	53.11	4.90	8.26	24.39	36.83	53.28
mx04	6.73	10.01	25.20	37.00	53.77	6.53	9.75	24.76	36.49	53.25
mx08	6.73	9.95	24.52	36.17	52.71	10.81	14.62	30.25	41.90	57.77
mx10	5.13	7.99	22.45	34.07	50.59	5.27	8.18	22.80	34.46	50.99
pe04	7.45	11.07	27.71	40.35	57.42	59.70	65.12	78.13	83.32	88.85
ru07	2.87	5.06	17.90	29.30	47.42	15.86	22.52	45.20	58.30	74.24
ru10	4.10	6.39	18.56	29.27	46.12	37.89	42.00	56.73	66.02	77.18
sk04	3.24	5.17	15.40	25.21	41.11	4.22	6.40	17.24	27.31	43.31
sk07	2.36	4.02	14.08	24.07	40.07	3.25	5.14	16.06	26.45	42.60
sk10	3.10	4.83	14.93	24.73	40.61	18.43	20.43	30.17	38.96	52.62
za08	8.41	14.13	37.83	52.67	69.56	19.37	29.10	58.60	71.81	84.01
za10	5.85	9.32	29.47	45.15	63.59	20.96	25.72	51.12	65.05	78.49
es04	2.52	4.33	14.92	25.18	41.55	35.25	38.69	50.34	58.25	68.78
es07	2.26	4.00	14.38	24.50	40.65	3.67	6.02	18.08	28.84	45.10
es10	2.20	3.97	15.09	25.53	41.98	2.20	3.98	15.13	25.58	42.04
se00	3.44	5.15	15.05	24.70	40.71	38.97	40.74	48.98	55.81	66.29
se05	2.89	4.52	14.09	23.64	39.54	32.43	33.93	41.66	48.86	60.29
ch02	2.71	4.37	13.70	22.95	38.13	1.45	2.50	9.40	16.87	30.17
ch04	2.48	4.02	13.31	22.48	37.59	2.23	3.67	12.61	21.53	36.45
uk04	5.26	7.69	19.42	29.73	45.58	11.44	14.76	28.24	38.81	54.01
uk07	4.38	6.76	18.27	28.53	44.52	8.79	12.12	25.54	36.28	52.00
uk10	4.44	6.79	18.40	28.69	44.47	14.69	19.01	33.83	44.54	59.08
us79	2.45	4.22	14.51	24.51	40.68	5.12	7.40	18.65	28.79	44.58
us86	2.73	4.59	15.40	25.71	42.20	4.13	6.39	18.14	28.73	45.14
us91	2.87	4.84	15.90	26.28	42.80	6.28	9.04	21.36	31.81	47.71
us94	4.97	7.12	18.55	28.88	45.04	19.46	21.76	32.46	41.45	55.13
us97	5.68	7.85	19.04	29.24	45.22	17.81	20.29	31.44	40.74	54.74
us00	6.06	8.31	19.29	29.42	45.31	20.88	23.49	34.04	42.98	56.44
us04	5.54	7.81	19.11	29.51	45.87	31.62	34.35	44.20	52.12	63.87
us07	5.18	7.47	19.21	29.69	46.13	10.47	13.17	25.55	35.85	51.47
us10	3.97	6.24	18.11	28.80	45.50	5.87	8.52	21.23	32.12	48.65
us13	6.00	8.64	20.71	31.27	47.62	10.09	13.28	26.07	36.58	52.34
uy04	5.35	8.21	22.56	34.17	50.54	5.62	8.55	23.14	34.83	51.19

Notes: Population shares refer to households in certain percentiles of disposable household income. Income shares are computed using disposable household income. Because the unit-nonresponse correction model is run using disposable household income *per capita* (using adult equivalent scale of square-root of household size), the ranking of households according to disposable household income may differ from the ranking using disposable household income per capita. Hence, aggregate income shares of the richest households according to disposable household income could fall under the correction for unit-nonresponse even when this correction identifies that households with the highest disposable household income *per capita* have the lowest survey response rate. Nevertheless, the

four surveys above where the top income shares fall (*au10, mx04, ch02, ch04*) are those where the nonresponse-correction model predicted poorer households to have lower survey response rates. Because of limited sample sizes, income shares of the top 0.5–1% of households may be sensitive to individual income observations and weights.

## References

Anand, S., and Segal, P. (2008) What do we know about global income inequality?, *Journal of Economic Literature* 46(1):57-94.

Anand, S., and Segal, P. (2015) The Global Distribution of Income, Ch. 11 in ed. A.B. Atkinson and F. Bourguignon, *Handbook of Income Distribution, Volume 2A*.

Atkinson, A.B., and Micklewright, J. (1983) On the reliability of income data in the family expenditure survey 1970–1977, *Journal of the Royal Statistical Society Series A* 146, 1, 33-61.

Atkinson, A.B., Rainwater L., and Smeeding T.M. (1995) Income distribution in OECD countries: Evidence from the Luxembourg income study, *Social Policy Studies*, Organisation for Economic Cooperation and Development, Paris.

Atkinson, A.B., Piketty T., and Saez E. (2011). Top Incomes in the Long Run of History, *Journal of Economic Literature* 49 (1): 3-71.

Bates, N. and Creighton, K. (2000) The last five percent: What can we learn from difficult interviews?, *Proceedings of the Annual Meetings of the American Statistical Association*, August 13-17, 2000. <http://www.fcsm.gov/committees/ihsng/asa2000proceedings.pdf>

Bhalla, S. (2002) Imagine there's no country: Poverty, inequality and growth in the era of globalization, *Institute for International Economics*, Washington, DC.

Bourguignon, F. (2015), *The Globalization of Inequality*, Princeton University Press.

Bourguignon, F. and Morrisson, C. (2002) Inequality among world citizens: 1820–1992, *American Economic Review* 92, 4, 727-744.

Burkhauser, R.V., Feng, S. and Jenkins, S.P. (2009) Using the P90/P10 ratio to measure US inequality trends with Current Population Survey data: a view from inside the Census Bureau vaults, *Review of Income and Wealth*, 55, 166-185.

Burkhauser, R.V., Feng, S. and Larrimore, J. (2010) Improving imputations of top incomes in the public-use current population survey by using both cell-means and variances, *Economic Letters*, 108, 69-72.

Burkhauser, R.V., Feng, S., Jenkins, S.P. and Larrimore, J. (2011) Estimating trends in US income inequality using the Current Population Survey: the importance of controlling for censoring, *Journal of Economic Inequality*, 9(3), 393-415.

Burkhauser, R.V., Feng, S., Jenkins, S.P. and Larrimore, J. (2012) Recent trends in top income shares in the United States: Reconciling estimates from March CPS and IRS tax return data, *Review of Economics and Statistics*, 94(2), 371-388.

Burricand, Carine (2013). Transition from survey data to registers in the French SILC survey. In *The Use of Registers in the Context of EU-SILC: Challenges and Opportunities* (eds Markus Jantti, Veli-Matti Tormalehto and Eric Marlier). European Union.

Census Bureau and Bureau of Labor Statistics, (Census & BLS, 2002) *Current Population Survey: Design and methodology, Technical paper 63RV* ([www.census.gov/prod/2002pubs/tp63rv.pdf](http://www.census.gov/prod/2002pubs/tp63rv.pdf)), U.S. Government Printing Office, Washington, DC.

Census Bureau, (Census, 2015) *2014 Annual Social and Economic Supplement* ([www2.census.gov/programs-surveys/cps/techdocs/cpsmar14.pdf](http://www2.census.gov/programs-surveys/cps/techdocs/cpsmar14.pdf)), U.S. Government Printing Office, Washington, DC, March 2015.

Cowell, F.A. and Victoria-Feser, M.-P. (1996) Poverty measurement with contaminated data: A robust approach, *European Economic Review*, 40, 1761-1771.

Cowell, F.A. and Victoria-Feser, M.-P. (1996) Robustness properties of inequality measures, *Econometrica*, 64, 77-101.

Cowell, F.A. and Flachaire, E. (2007) Income distribution and inequality measurement: The problem of extreme values, *Journal of Econometrics*, 141(2), 1044-1072.

Deaton, A. (2005) Measuring Poverty in a growing world (or measuring growth in a poor world), *The Review of Economics and Statistics*, 87(1), 1-19.

Dixon, J. (2007) Nonresponse bias patterns in the Current Population Survey, Bureau of Labor Statistics.

Groves, R.M., and Couper, M. (1998) *Nonresponse in Household Interview Surveys*, John Wiley and Sons, Inc.: New York, NY.

Hlasny, V., and Verme, P. (2013) *Top Incomes and the Measurement of Inequality in Egypt*, World Bank Policy Research working paper series #6557.

Hlasny, V., and Verme, P. (2015) *Top Incomes and the Measurement of Inequality: A Comparative Analysis of Correction Methods using EU, US and Egyptian Survey Data*, unpublished manuscript.

Hokayem, C., Ziliak, J.P., and Bollinger, C.R. (2012) A look at CPS non-response and trends in poverty, US Census SEHSD Working Paper 2012-21.

- Jantti, Markus, Tormalehto, Veli-Matti and Marlier, Eric (2013). The Use of Registers in the Context of EU-SILC: Challenges and Opportunities. European Union.
- Jenkins, S.P., Burkhauser, R.V., Feng, S., and Larrimore, J. (2011). Measuring inequality using censored data: a multiple-imputation approach to estimation and inference, *Journal of the Royal Statistical Society*, 174(1), 63-81.
- Jenkins, S.P. (2016). Pareto distributions, top incomes, and recent trends in UK income inequality", ISER Working Paper Series2016-07.
- Jolliffe, D., Datt, G., and Sharma, M. (2004) Robust poverty and inequality measurement in Egypt: Correcting for spatial-price variation and sample design effects, *Review of Development Economics*, 8(4), 557-572.
- Korinek, A., Mistiaen, J.A. and Ravallion, M. (2006) Survey nonresponse and the distribution of income, *Journal of Economic Inequality*, 4, 33-55.
- Korinek, A., Mistiaen, J.A. and Ravallion, M. (2007) An econometric method of correcting for unit nonresponse bias in surveys, *Journal of Econometrics*, 136, 213-235.
- Lahoti, R., Jayadev, A., and Reddy, S.G. (2015), The global consumption and income project (GCIP): an introduction and preliminary findings, DESA Working Paper 140, January 2015.
- Lakner, C. and Milanovic, B. (2013) Global income distribution from the fall of the Berlin Wall to the great recession, World Bank Policy Research working paper series #6719.
- Luxembourg Income Study (LIS) Database, <http://www.lisdatacenter.org> (multiple countries; microdata runs completed during 5-23 January, 2015). Luxembourg: LIS.
- Meyer, B.D., Mok, W.K.C. and Sullivan, J.X. (2009) The under-reporting of transfers in household surveys: Its nature and consequences, NBER Working Papers 15181, National Bureau of Economic Research, Inc.
- Meyer, B.D. and Mittag, N. (2014) Missclassification in binary choice models. AEA 2014 Proceedings.
- Milanovic, B. (2002a), The Ricardian vice: why Sala-i-Martin's calculations of world income inequality are wrong, Washington, DC: World Bank, November 2002.
- Milanovic, B. (2002b) True world income distribution, 1988 and 1993: First calculation based on household surveys alone, *Economic Journal* 112(476):51-92.
- Mistiaen, J.A. and Ravallion, M. (2003) Survey compliance and the distribution of income, Policy Research Working Paper #2956, The World Bank.

Moore, Richard A. Jr. (1996), Controlled Data-Swapping Techniques for Masking Public Use Microdata Sets, US Bureau of the Census research report rr96-4.

Nino-Zarazua, M., Roope, L., and Tarp, F. (2014) Global interpersonal inequality: trends and measurement, UNU-Wider project Working Paper 2014-004, January 2014.

Platek, R. (1977) Some factors affecting non-response. *Survey Methodology*. 3:191-214. *Bulletin of the International Statistical Institute*.- [Wechselnde Verlagsorte], ISSN 0373-0441, ZDB-ID 2150426. 47.1977, 3, 347-366.

Sala-i-Martin, X. (2002) The world distribution of income (estimated from individual country distributions), NBER Working Paper No. W8933.

Sala-i-Martin, X. (2006) The world distribution of income: falling poverty and ... convergence, period, *Quarterly Journal of Economics* 121(2):351-397.

Tiehen, L., Jolliffe, D. and Smeeding, T. (2013) The effect of SNAP on poverty, Brookings Institute Conference paper.

World Bank (2015a) Inflation: GDP Deflator (annual %), World Bank national accounts data, and OECD National Accounts data, World Development Indicators catalog, <http://data.worldbank.org/indicator/NY.GDP.DEFL.KD.ZG>.

World Bank (2015b) PPP conversion factor, GDP (LCU per international \$), International Comparison Program database, World Development Indicators catalog, <http://data.worldbank.org/indicator/PA.NUS.PPP>.

World Bank (2015c) World DataBank: World Development Indicators, <http://databank.worldbank.org/data/home.aspx>. Last updated 22 December 2015; extracted 6 January 2016.

## Appendix 1: Regional nonresponse rates

Table A1. Regional nonresponse rates for included surveys (all countries but US, %)

<i>dname</i>	<i>region_c</i>	Region name	Nonresp. rate	Notes
at08	11	Burgenland	35.82	EU-SILC 2009, NRh.
	12	N. Östrch.	30.62	
	13	Wien	24.29	
	21	Kärnten	32.50	
	22	Steiermark	25.28	
	31	O. Östrch.	27.85	
	32	Salzburg	26.77	
	33	Tirol	24.50	
	34	Vorarlberg	24.30	
		TOTAL	27.67	
at10	11	Burgenland	26.74	EU-SILC 2011, NRh.
	12	N. Östrch.	22.95	
	13	Wien	16.50	
	21	Kärnten	24.04	
	22	Steiermark	21.89	
	31	O. Östrch.	22.29	
	32	Salzburg	22.26	
	33	Tirol	20.27	
	34	Vorarlberg	18.83	
		TOTAL	21.18	
au10	1	N. S. Wales	26.40	Survey of Inc. & Hous.+Pension. & Benefic. Module+ Hh. Expend. Survey. hh. member illness or death; non-contact, lang. prob., refusal, nonresp. to key questions. Undercoverage of screening in pensioner and beneficiary modules unaccounted for.
	2	Victoria	29.70	
	3	Queensland	17.40	
	4	South Austr.	21.80	
	5	Western Australia	17.10	
	6	Tasmania	9.90	
	7	ACT	16.10	
		TOTAL	21.40	
be00	1	Brussels	15.20	Panel Study of Belgian Hhds/ BE-ECHP 5&10 ('96 & '01), nonresp./drawn (Dewilde <i>et al.</i> 1999; Ottoy <i>et al.</i> 2004)
	2	Vlaanders	7.06	
	3	Wallonia	14.63	
	TOTAL	10.30		
be95	1	Brussels	11.10	Panel Study of Belgian Hhds/ BE-ECHP 5&10 ('96 & '01), nonresp./drawn (Dewilde <i>et al.</i> 1999; Ottoy <i>et al.</i> 2004)
	2	Vlaanders	10.10	
	3	Wallonia	10.30	
		TOTAL	10.30	
br06	11	Rondônia	2.05	Pesquisa Nacional por Amostra de Domicílios, PNAD, 2006, 2009, 2011 (IBGE 2006, 2009, 2011); Nonresponding/drawn.
	12	Acre	0.94	
	13	Amazonas	3.31	
	14	Roraima	1.97	
	15	Pará	3.14	
	16	Amapá	0.23	
	17	Tocantins	1.65	
	21	Maranhão	5.44	
	22	Piauí	0.26	
	23	Ceará	3.45	
	24	Rio Grande do	2.75	
	25	Paraíba	0.14	
	26	Pernambuco	1.57	
27	Alagoas	0.85		
28	Sergipe	1.32		
29	Bahia	2.43		
31	Minas Gerais	2.24		
32	Espírito Santo	1.94		
33	Rio de Janeiro	3.34		
35	São Paulo	3.58		
41	Paraná	1.46		
42	Santa Catarina	1.09		
43	Rio Grande do	1.95		
50	Mato Grosso do	0.88		
51	Mato Grosso	3.38		
52	Goiás	5.91		
53	Distrito Feder	4.75		
	TOTAL	2.66		
br09	11	Rondônia	3.07	Pesquisa Nacional por Amostra de Domicílios, PNAD, 2006, 2009, 2011 (IBGE 2006, 2009, 2011); Nonresponding/drawn.
	12	Acre	3.43	
	13	Amazonas	2.18	
	14	Roraima	5.11	
	15	Pará	6.58	
	16	Amapá	1.00	
	17	Tocantins	1.67	
	21	Maranhão	4.33	
	22	Piauí	0.36	
	23	Ceará	4.48	
	24	Rio Grande do	1.51	
	25	Paraíba	0.92	
	26	Pernambuco	4.63	
27	Alagoas	1.19		
28	Sergipe	2.72		
29	Bahia	4.12		
31	Minas Gerais	2.23		
32	Espírito Santo	6.14		
33	Rio de Janeiro	4.53		
35	São Paulo	9.26		
41	Paraná	1.50		
42	Santa Catarina	2.77		
43	Rio Grande do	4.90		
50	Mato Grosso do	1.09		
51	Mato Grosso	9.11		
52	Goiás	5.47		
53	Distrito Feder	6.03		
	TOTAL	4.55		
br11	11	Rondônia	6.11	Pesquisa Nacional por Amostra de Domicílios, PNAD, 2006, 2009, 2011 (IBGE 2006, 2009, 2011); Nonresponding/drawn.
	12	Acre	8.69	
	13	Amazonas	10.66	
	14	Roraima	2.04	
	15	Pará	8.31	
	16	Amapá	4.08	
	17	Tocantins	6.41	
	21	Maranhão	3.39	
	22	Piauí	2.36	

23	Ceará	8.80	
24	Rio Grande do	1.15	
25	Paraíba	1.14	
26	Pernambuco	11.50	
27	Alagoas	4.65	
28	Sergipe	4.27	
29	Bahia	5.58	
31	Minas Gerais	4.26	
32	Espírito Santo	7.72	
33	Rio de Janeiro	6.85	
35	São Paulo	10.04	
41	Paraná	6.54	
42	Santa Catarina	7.88	
43	Rio Grande do	4.18	
50	Mato Grosso do	2.21	
51	Mato Grosso	6.75	
52	Goiás	7.04	
53	Distrito Feder	9.56	
	TOTAL	6.84	
ca04	10	Newfnd & Labrador	17.00
	11	Prince Edward Island	22.00
	12	Nova Scotia	21.20
	13	New Brunswick	24.20
	24	Quebec	25.10
	35	Ontario	29.00
	46	Manitoba	22.70
	47	Saskatchewan	20.90
	48	Alberta	24.20
	59	British Columbia	28.40
		TOTAL	25.20
ca07	10	Newfnd. & Labrador	18.00
	11	Prince Edward Island	22.60
	12	Nova Scotia	22.30
	13	New Brunswick	24.60
	24	Quebec	26.40
	35	Ontario	32.00
	46	Manitoba	28.00
	47	Saskatchewan	25.20
	48	Alberta	29.10
	59	British Columbia	32.00
		TOTAL	28.20
ca10	10	Newfnd. & Labrador	21.50
	11	Prince Edward Island	27.10
	12	Nova Scotia	25.00
	13	New Brunswick	25.50
	24	Quebec	30.10
	35	Ontario	35.80
	46	Manitoba	32.20
	47	Saskatchewan	32.50
	48	Alberta	36.40
	59	British Columbia	38.40
		TOTAL	32.70
ch02	1	Genfersee	72.12
	2	Espace Mittelland	69.27
	3	NWschweiz	67.43
	4	Zurich	70.23
	5	Ostschweiz	70.49
	6	Zentralschweiz	69.16
	7	Tessin	74.58
		TOTAL	70.38
ch04	1	Genfersee	70.64
	2	Espace Mittelland	64.81
	3	NWschweiz	66.36
	4	Zurich	67.76
	5	Ostschweiz	67.06

Survey of Labour & Inc. Dynamics, SLID, 2004; nonresponding/eligible hhds (Duddek & Storey 2006)

SLID 2007, 2010; nonresponding/eligible hhds (source: Statistics Canada, and Bastien 2009)

Income & Cons. Survey, EVE/ERC, 2002 & 2004; Stratified samples, nonresp. hhds/activated addresses

6	Zentralschweiz	64.96	
7	Tessin	73.59	
	TOTAL	67.72	
co10 <sup>a</sup>	108	Atlantic, Atlántico, Barranquilla	3.30
	113	Bolívar, Cartagena	0.70
	120	Cesar, Valledupar	1.86
	123	Córdoba, Montería	0.39
	144	La Guajira, Riohacha	4.21
	147	Magdalena, Santa Marta	4.31
	170	Sucre, Sincelejo	1.71
	215	Oriental, Boyacá, Tunja	0.00
	225	Cundinamarca	3.12
	250	Meta, Villavicencio	3.20
	254	N. de Santande, Cucuta	1.13
	268	Santander, Bucaramanga	2.63
	305	Central, Antioquia, Medellin	1.50
	317	Caldas, Manizales	1.72
	318	Caquetá, Florencia	0.94
	341	Huila, Neiva	2.11
	363	Quindio, Armenia	0.79
	366	Risaralda, Pereira	1.68
	373	Tolima, Ibague	1.40
	419	Pacific, Cauca, Popayan	1.31
	427	Chocó, Quibdó	0.74
	452	Nariño, Pasto	1.47
	476	Valle del Cauca, Cali	2.72
	511	Bogotá, Distr. Cap.	3.12
		TOTAL	1.90
co13 <sup>a</sup>	108	Atlantic, Atlántico, Barranquilla	2.85
	113	Bolívar, Cartagena	1.51
	120	Cesar, Valledupar	0.65
	123	Córdoba, Montería	4.38
	144	La Guajira, Riohacha	3.48
	147	Magdalena, Santa Marta	2.71
	170	Sucre, Sincelejo	2.90
	215	Oriental, Boyacá, Tunja	0.72
	225	Cundinamarca	2.06
	250	Meta, Villavicencio	1.91
	254	N. de Santande, Cucuta	1.54
	268	Santander, Bucaramanga	2.68
	305	Central, Antioquia, Medellin	2.47
	317	Caldas, Manizales	1.16
	318	Caquetá, Florencia	0.22
	341	Huila, Neiva	2.67
	363	Quindio, Armenia	1.55
	366	Risaralda, Pereira	2.12
	373	Tolima, Ibague	1.80
	419	Pacific, Cauca, Popayan	1.11
	427	Chocó, Quibdó	3.34
	452	Nariño, Pasto	1.46
	476	Valle del Cauca,	2.64

Gran Encuesta Integrada de Hogares, GEIH, January 2011 (DANE 2011); Nonresp./.(resp.+nonresp.)

GEIH, January 2013 (DANE 2013); Nonresp./.(resp.+nonresp.)

		Cali	
	511	Bogotá, Distr. Cap.	2.06
		TOTAL	2.09
cz04	11	Praha	48.90
	21	Středočeský	36.30
	31	Jihočeský	37.10
	32	Plzeňský	26.70
	41	Karlovarský	38.90
	42	Ústecký	35.40
	51	Liberecký	36.00
	52	Královéhradecký	37.10
	53	Pardubický	31.90
	61	Vysočina	26.50
	62	Jihomoravský	40.00
	71	Olomoucký	25.60
	72	Zlínský	32.70
	81	Moravskoslezský	26.10
		TOTAL	35.20
cz07	11	Praha	30.50
	21	Středočeský	18.90
	31	Jihočeský	13.00
	32	Plzeňský	17.70
	41	Karlovarský	16.40
	42	Ústecký	15.90
	51	Liberecký	16.70
	52	Královéhradecký	18.70
	53	Pardubický	15.00
	61	Vysočina	10.00
	62	Jihomoravský	16.40
	71	Olomoucký	16.00
	72	Zlínský	11.90
	81	Moravskoslezský	13.10
		TOTAL	16.90
cz10	11	Praha	22.40
	21	Středočeský	15.40
	31	Jihočeský	9.40
	32	Plzeňský	13.10
	41	Karlovarský	19.50
	42	Ústecký	16.00
	51	Liberecký	16.10
	52	Královéhradecký	17.60
	53	Pardubický	16.50
	61	Vysočina	8.60
	62	Jihomoravský	14.70
	71	Olomoucký	14.00
	72	Zlínský	12.30
	81	Moravskoslezský	11.40
		TOTAL	17.15
de07	1	Baden-Wuerttemb.	13.15
	2	Bayern	13.06
	3	Berlin	10.85
	4	Brandenburg	10.06
	5	Bremen	10.64
	6	Hamburg	10.05
	7	Hessen	14.83
	8	Mecklenburg-Vorp.	11.11
	9	Niedersachsen	13.11
	10	Nordrhein-Westfal.	14.80
	11	Rheinland-Pfalz	17.61
	12	Saarland	1.54
	13	Sachsen	9.98
	14	Sachsen-Anhalt	9.80
	15	Schleswig-Holstein	16.29
	16	Thueringen	9.60
		TOTAL	12.93
de10	1	Baden-Wuerttemb.	10.68
	2	Bayern	9.95
	3	Berlin	12.46

EU-SILC 2005, 2008, 2011 (CZSO 2007, 2010, 2012), not caught at home, language barrier, refusal due to health/age, other refusal

Sozio-ökonomische Panel, SOEP, 2008, 2011; longitudinal nonresp rate

	4	Brandenburg	10.12
	5	Bremen	12.90
	6	Hamburg	11.61
	7	Hessen	11.76
	8	Mecklenburg-Vorp.	8.36
	9	Niedersachsen	11.07
	10	Nordrhein-Westfal.	11.84
	11	Rheinland-Pfalz	9.91
	12	Saarland	8.61
	13	Sachsen	8.79
	14	Sachsen-Anhalt	8.12
	15	Schleswig-Holstein	10.15
	16	Thueringen	11.04
		TOTAL	10.65
do07	1	Distrito Nacional	29.03
	2	Azua	5.40
	3	Baoruco	8.96
	4	Barahona	5.71
	5	Dajabón	2.94
	6	Duarte	9.17
	7	Eliás Piña	3.61
	8	El Seibo	1.46
	9	Españat	6.94
	10	Independencia	4.23
	11	La Altagracia	8.77
	12	La Romana	11.06
	13	La Vega	11.55
	14	María Trin. Sánchez	5.65
	15	Monte Cristi	5.21
	16	Pedernales	6.25
	17	Peravia	7.26
	18	Puerto Plata	7.69
	19	Salcedo	6.06
	20	Samaná	7.50
	21	San Cristóbal	6.81
	22	San Juan	5.93
	23	San Pedro Macorís	9.01
	24	Sánchez Ramírez	5.22
	25	Santiago	15.49
	26	Santiago Rodríguez	6.52
	27	Valverde	7.83
	28	Monseñor Nouel	7.14
	29	Monte Plata	4.13
	30	Hato Mayor	3.23
	31	San José de Ocoa	5.00
	32	Santo Domingo	11.99
		TOTAL	10.30
ee04	101	Tallinn	52.60
	137	Harju	44.00
	439	Hiiu	26.70
	744	Ida-Viru	34.30
	849	Jõgeva	22.50
	651	Järva	25.60
	457	Lääne	39.30
	659	Lääne-Viru	25.00
	865	Põlva	12.50
	467	Pärnu	16.30
	670	Rapla	23.10
	474	Saare	30.60
	878	Tartu	34.50
	882	Valga	18.20
	884	Viljandi	37.50
	886	Võru	28.60
		TOTAL	34.10
ee07	1	Tallinn	42.60
	37	Harju	50.30
	39	Hiiu	36.10
	44	Ida-Viru	38.40
	49	Jõgeva	30.80
	51	Järva	20.00

Encuesta Nacional de Ingresos y Gastos de los Hogares 2007 (ONE 2008); Nonresponding/ drawn hhhd.

Estonian Social Survey, ESS, 2005, 2008 (Statistics Estonia 2010), due to differences in address accessibility, mobility of respondents as well as technical and other problems

	57	Lääne	30.40
	59	Lääne-Viru	30.80
	65	Põlva	18.20
	67	Pärnu	27.90
	70	Rapla	20.00
	74	Saare	29.30
	78	Tartu	40.90
	82	Valga	19.50
	84	Viljandi	41.20
	86	Võru	36.00
		<b>TOTAL</b>	<b>37.40</b>
ee10	1	Tallinn	41.20
	37	Harju	33.01
	39	Hiiu	30.16
	44	Ida-Viru	19.90
	49	Jõgeva	15.45
	51	Järva	15.45
	57	Lääne	21.98
	59	Lääne-Viru	26.29
	65	Põlva	7.85
	67	Pärnu	21.70
	70	Rapla	21.80
	74	Saare	16.40
	78	Tartu	32.14
	82	Valga	15.40
	84	Viljandi	23.64
	86	Võru	9.76
		<b>TOTAL</b>	<b>26.00</b>
eg12 <sup>a</sup>	1	Greater Cairo	9.79
	2	Alex., Suez Canal	5.66
	3	Urban Lower	4.88
	4	Urban Upper	4.47
	5	Rural Lower	2.06
	6	Rural Upper	2.44
		<b>TOTAL</b>	<b>4.36</b>
es04	11	Galicia	14.78
	12	Principado Asturias	12.15
	13	Cantabria	29.65
	21	País Vasco	17.87
	22	Com. Foral Navarra	13.16
	23	La Rioja	17.96
	24	Aragón	17.91
	30	Com. de Madrid	44.80
	41	Castilla y León	13.95
	42	Castilla-La Mancha	12.37
	43	Extremadura	9.33
	51	Cataluña	21.01
	52	Com. Valenciana	13.57
	53	Illes Balears	11.03
	61	Andalucía	16.10
	62	Región de Murcia	11.87
	63	C. Autón. Ceuta	9.86
	70	Canarias	18.11
		<b>TOTAL</b>	<b>18.41</b>
es07	11	Galicia	7.76
	12	Principado Asturias	6.60
	13	Cantabria	6.04
	21	País Vasco	15.30
	22	Com. Foral Navarra	4.87
	23	La Rioja	6.59

EU-SILC 2011, NRh

Egypt Labor Market Panel Survey 2012;  
Attrition rate: no hh re-interview due to  
refusal or moving to another known address  
where interview was then not conducted

Encuesta de Condiciones de Vida (ECV), and EU-SILC 2005, 2008,  
2011 (ES-INE 2005, 2008, 2011); Nonresponding/drawn.

	24	Aragón	9.54
	30	Com. de Madrid	14.37
	41	Castilla y León	7.02
	42	Castilla-La Mancha	7.42
	43	Extremadura	8.87
	51	Cataluña	10.82
	52	Com. Valenciana	6.44
	53	Illes Balears	10.77
	61	Andalucía	9.42
	62	Región de Murcia	11.85
	63	C. Autón. Ceuta	11.31
	64	C. Autón. Melilla	11.31
	70	Canarias	8.26
		<b>TOTAL</b>	<b>9.49</b>
es10	11	Galicia	15.07
	12	Principado Asturias	14.76
	13	Cantabria	8.97
	21	País Vasco	12.12
	22	Com. Foral Navarra	10.88
	23	La Rioja	4.89
	24	Aragón	7.13
	30	Com. de Madrid	15.21
	41	Castilla y León	12.94
	42	Castilla-La Mancha	4.07
	43	Extremadura	12.21
	51	Cataluña	12.75
	52	Com. Valenciana	8.56
	53	Illes Balears	25.72
	61	Andalucía	17.78
	62	Región de Murcia	11.91
	63	C. Autón. Ceuta	10.00
	64	C. Autón. Melilla	10.00
	70	Canarias	11.11
		<b>TOTAL</b>	<b>12.51</b>
es13	11	Galicia	22.69
	12	Principado Asturias	26.33
	13	Cantabria	37.63
	21	País Vasco	26.81
	22	Com. Foral Navarra	14.80
	23	La Rioja	23.46
	24	Aragón	24.76
	30	Com. de Madrid	26.12
	41	Castilla y León	22.62
	42	Castilla-La Mancha	29.59
	43	Extremadura	20.81
	51	Cataluña	31.14
	52	Com. Valenciana	29.23
	53	Illes Balears	31.68
	61	Andalucía	23.58
	62	Región de Murcia	20.74
	63	C. Autón. Ceuta	16.23
	64	C. Autón. Melilla	16.23
	70	Canarias	33.73
		<b>TOTAL</b>	<b>26.17</b>
fi07	30	FI19	16.38
	31	FI1D	16.26
	32	FI1C	17.28
	33	FI1B	22.64
	34	FI20	21.18
		<b>TOTAL</b>	<b>18.22</b>
fi10	30	FI19	16.48
	31	FI1D	17.17
	32	FI1C	16.98
	33	FI1B	21.81
	34	FI20	14.29
		<b>TOTAL</b>	<b>18.10</b>
fi05	1	DR45,78: Centre,	34.79

Living Conditions Survey (LCS) & EU-SILC 2014  
(ES-INE 2014); Nonresponding/drawn.

Income Distrib. Survey (IDS) /EU-SILC 2008 & 2011;  
Nonresponding/drawn.

		Ile de France	
2	DR14,21,51,76,80:	35.96	
	Basse & Haute		
	Norman.,		
	Champag.-Ard.,		
	Bourgogne, Picardie		
3	DR59 Nord, Calais	33.30	
4	DR25,54,67:	36.43	
	Franche-Comte,		
	Lorraine, Alsace		
5	DR35,44,86:	34.32	
	Bretagne, Pays de la		
	Loire, Poitou-Char.		
6	DR31,33,87: Midi-	27.98	
	Pyrenees,		
	Aquitaine, Limousin		
7	DR63,69:	21.68	
	Auvergne, Rhone-		
	Alpes		
8	DR13,34: Provence-	22.42	
	Alpes-Cote D'Azur,		
	Lang.-Roussil.		
	TOTAL	31.30	
fr10	1	DR45,78: Centre,	30.38
		Ile de France	
	2	DR14,21,51,76,80:	25.98
		Basse & Haute	
		Norman.,	
		Champag.-Ard.,	
		Bourgogne, Picardie	
	3	DR59 Nord, Calais	19.00
	4	DR25,54,67:	25.58
		Franche-Comte,	
		Lorraine, Alsace	
	5	DR35,44,86:	18.22
		Bretagne, Pays de la	
		Loire, Poitou-Char.	
	6	DR31,33,87: Midi-	21.21
		Pyrenees,	
		Aquitaine, Limousin	
	7	DR63,69:	24.87
		Auvergne, Rhone-	
		Alpes	
	8	DR13,34: Provence-	27.30
		Alpes-Cote D'Azur,	
		Lang.-Roussil.	
	9	DR9A Guadeloupe	17.70
	10	DR9B Martinique	19.40
	11	DR9C Guyane	22.90
	12	DR9D Reunion	15.90
	13	DR9F Mayotte	13.40
		TOTAL	22.40
gr04	11	GR11	7.56
	12	GR12	13.39
	13	GR13	4.81
	14	GR14	12.70
	21	GR21	14.16
	22	GR22	8.33
	23	GR23	12.62
	24	GR24	13.41
	25	GR25	15.38
	30	GR30	39.93
	41	GR41	20.00
	42	GR42	8.91
	43	GR43	13.95
		TOTAL	22.39
gr07	11	GR11	8.28
	12	GR12	14.01
	13	GR13	4.21
	14	GR14	6.25

EU-SILC 2005, 2008, 2011 (Statistics Greece 2010a,b & 2012); Nonresponding/drawn.

	21	GR21	8.37
	22	GR22	15.32
	23	GR23	4.59
	24	GR24	4.36
	25	GR25	6.59
	30	GR30	17.05
	41	GR41	5.59
	42	GR42	7.28
	43	GR43	10.98
		TOTAL	11.74
gr10	11	GR11	17.89
	12	GR12	29.27
	13	GR13	11.07
	14	GR14	13.57
	21	GR21	12.75
	22	GR22	36.11
	23	GR23	7.75
	24	GR24	15.23
	25	GR25	13.20
	30	GR30	48.79
	41	GR41	9.20
	42	GR42	27.97
	43	GR43	22.38
		TOTAL	29.16
gt06	1	Guatemala	10.57
	2	El Progreso	1.67
	3	Sacatepéquez	1.00
	4	Chimaltenango	7.52
	5	Escuintla	0.31
	6	Santa Rosa	0.00
	7	Sololá	2.58
	8	Totonicapán	8.93
	9	Quezaltenango	6.18
	10	Suchitepéquez	0.00
	11	Retalhuleu	0.00
	12	San Marcos	3.70
	13	Huehuetenango	8.54
	14	Quiché	5.74
	15	Baja Verapaz	5.00
	16	Alta Verapaz	13.01
	17	Petén	0.00
	18	Izabal	5.00
	19	Zacapa	5.99
	20	Chiquimula	4.09
	21	Jalapa	0.63
	22	Jutiapa	0.00
		TOTAL	3.67
hu07	10	Közép-	9.97
		Magyarország	
	21	Közép-Dunántúl	18.96
	22	Nyugat-Dunántúl	11.04
	23	Dél-Dunántúl	18.47
	31	Észak-	19.69
		Magyarország	
	32	Észak-Alföld	18.67
	33	Dél-Alföld	18.34
		TOTAL	15.56
hu12	10	Central Hungary	25.67
	21	Central	17.89
		Transdanubia	
	22	Western Hungary	23.76
	23	South. Transdanubia	29.14
	31	Northern Hungary	19.29
	32	Northern Great	30.07
		Plain	
	33	Southern Great	16.48
		Plain	
		TOTAL	23.71
il10	11	Jerusalem	19.15

Encuesta Nacional de Condiciones de Vida (GT-INE 2006),  
hhds, non-selected vs. hhds drawn

Tárki Háztartás Monitor,  
HMS, 2007 (Tarki 2007);  
(Refused+failed)/drawn;  
stratified at NUTS2

HMS 2012 (Tarki 2013);  
Nonresponding/drawn;  
stratified at NUTS2

HMS

21	Zefat, Kineret, Golan	8.28
23	North: Yizre'el	8.51
24	North: Akko	7.30
31	Haifa: Haifa	8.20
32	Haifa: Hadera	5.30
41	Center: Sharon	9.86
42	Center: Petah Tiqwa	11.97
43	Center: Ramla	14.44
44	Center: Rehovot	11.85
51	Tel aviv: Tel aviv	14.55
52	Tel aviv: Ramat-gan	15.01
53	Tel aviv: Holon	17.44
61	South: Ashqelon	12.17
62	South: Be'er Sheva	11.49
70	Yehuda & Shomron	9.39
	TOTAL	12.12
ie07	11 Border region	18.22
	12 Midlands region	18.22
	13 West region	18.22
	21 Dublin region	25.43
	22 Mid-east region	25.43
	23 Mid-west region	25.43
	24 South-east region	25.43
	25 South-west region	25.43
	TOTAL	23.75
ie10	11 Border region	16.12
	12 Midlands region	16.12
	13 West region	16.12
	21 Dublin region	20.78
	22 Mid-east region	20.78
	23 Mid-west region	20.78
	24 South-east region	20.78
	25 South-west region	20.78
	TOTAL	19.58
is04	0 Densely populated	25.66
	1 Thinly populated	23.35
	TOTAL	24.87
is07	0 Densely populated	25.66
	1 Thinly populated	23.35
	TOTAL	26.18
is10	0 Densely populated	25.66
	1 Thinly populated	23.35
	TOTAL	24.83
it08	1 Piemonte	24.60
	2 Valle d' Aosta	30.80
	3 Lombardia	60.50
	4 Trentino	25.30
	5 Veneto	56.00
	6 Friuli	40.70
	7 Liguria	55.50
	8 Emil. Romagna	28.80
	9 Toscana	43.70
	10 Umbria	51.40
	11 Marche	26.90
	12 Lazio	51.30
	13 Abruzzi	38.30
	14 Molise	15.90
	15 Campania	31.10
	16 Puglia	33.50
	17 Basilicata	22.00
	18 Calabria	37.90
	19 Sicilia	54.50
	20 Sardegna	40.10
	TOTAL	43.90
it10	1 Piemonte	41.30
	2 Valle d' Aosta	35.20
	3 Lombardia	58.10

EU-SILC 2008 & 2011, NRH; Source: Central Statistics Office

EU-SILC 2005, 2008, 2011: refused, absent, unable to respond, failed interv. LIS var. rural.

Indagine sui Bilanci delle Famiglie Italiane (SHIW) 2008, 2010, 2012; total nonresponse rate. SHIW 2012; total nonresponse rate.

4	Trentino	45.70
5	Veneto	47.50
6	Friuli	50.00
7	Liguria	62.70
8	Emil. Romagna	40.40
9	Toscana	53.20
10	Umbria	52.10
11	Marche	34.90
12	Lazio	59.60
13	Abruzzi	39.50
14	Molise	30.10
15	Campania	39.30
16	Puglie	34.50
17	Basilicata	32.30
18	Calabria	42.00
19	Sicilia	48.20
20	Sardegna	31.50
	TOTAL	47.30
it12	1 Piemonte	49.90
	2 Valle d' Aosta	38.00
	3 Lombardia	55.80
	4 Trentino	28.50
	5 Veneto	48.20
	6 Friuli	55.10
	7 Liguria	59.60
	8 Emil. Romagna	44.90
	9 Toscana	49.70
	10 Umbria	42.30
	11 Marche	27.10
	12 Lazio	57.50
	13 Abruzzi	46.10
	14 Molise	18.50
	15 Campania	47.60
	16 Puglia	51.10
	17 Basilicata	27.30
	18 Calabria	20.10
	19 Sicilia	37.30
	20 Sardegna	37.50
	TOTAL	47.40
jp08	1 Hokkaido	51.78
	2 Tohoku	58.90
	3 Kanto	61.43
	4 Chubu	51.78
	5 Kinki	62.74
	6 Chugoku	56.58
	7 Shikoku	56.04
	8 Kyushu	53.74
	TOTAL	58.34
mx	1 Aguascalientes	5.85
02	2 Baja California	16.76
	3 Baja California Sur	13.33
	4 Campeche	10.10
	5 Coahuila	15.99
	6 Colima	21.08
	7 Chiapas	17.10
	8 Chihuahua	9.76
	9 Distr.Federal	14.56
	10 Durango	12.27
	11 Guanajuato	18.71
	12 Guerrero	13.17
	13 Hidalgo	9.00
	14 Jalisco	13.30
	15 Edo. de Mexico	15.34
	16 Michoacan	8.64
	17 Morelos	14.66
	18 Nayarit	11.28
	19 Nuevo Leon	8.50

Japan Household Panel Survey 2009 (PDRC 2015), quasi-nonresp. rate=1-[responding/(original+reserve hhds)]

Encuesta NacI. de Ingresos y Gastos de los Hogares, ENIGH, 2002 (INEGI 2003); Total nonresponse rate=1-accepted/drawn

20	Oaxaca	15.95		24	San Luis Potosi	15.68		
21	Puebla	18.65		25	Sinaloa	16.77		
22	Queretaro	9.95		26	Sonora	16.90		
23	Quintana Roo	15.70		27	Tabasco	16.58		
24	San Luis Potosi	14.53		28	Tmaulipas	14.31		
25	Sinaloa	12.66		29	Tlaxcala	12.68		
26	Sonora	6.26		30	Veracruz	17.79		
27	Tabasco	13.00		31	Yucatan	21.07		
28	Tmaulipas	16.09		32	Zacatecas	12.56		
29	Tlaxcala	10.99			TOTAL	17.73		
30	Veracruz	12.68		mx	1	Aguascalientes	10.11	
31	Yucatan	15.80		10	2	Baja California	7.69	
32	Zacatecas	12.11			3	Baja California Sur	9.67	
	TOTAL	13.53			4	Campeche	8.17	
mx	1	Aguascalientes	11.14		5	Coahuila	10.05	
04	2	Baja California	12.46		6	Colima	15.06	
	3	Baja California Sur	9.74		7	Chiapas	9.37	
	4	Campeche	14.17		8	Chihuahua	10.87	
	5	Coahuila	12.95		9	Distr.Federal	8.11	
	6	Colima	17.93		10	Durango	15.76	
	7	Chiapas	12.71		11	Guanajuato	8.65	
	8	Chihuahua	15.47		12	Guerrero	10.33	
	9	Distr.Federal	8.66		13	Hidalgo	8.88	
	10	Durango	12.76		14	Jalisco	11.06	
	11	Guanajuato	9.53		15	Edo. de Mexico	10.31	
	12	Guerrero	15.46		16	Michoacan	8.14	
	13	Hidalgo	6.72		17	Morelos	12.00	
	14	Jalisco	16.13		18	Nayarit	5.12	
	15	Edo. de Mexico	12.60		19	Nuevo Leon	12.18	
	16	Michoacan	12.90		20	Oaxaca	6.66	
	17	Morelos	12.09		21	Puebla	12.89	
	18	Nayarit	11.03		22	Queretaro	10.32	
	19	Nuevo Leon	11.24		23	Quintana Roo	10.71	
	20	Oaxaca	16.69		24	San Luis Potosi	10.61	
	21	Puebla	3.29		25	Sinaloa	6.34	
	22	Queretaro	12.92		26	Sonora	4.98	
	23	Quintana Roo	14.71		27	Tabasco	12.25	
	24	San Luis Potosi	9.28		28	Tmaulipas	12.26	
	25	Sinaloa	12.26		29	Tlaxcala	7.09	
	26	Sonora	9.42		30	Veracruz	10.20	
	27	Tabasco	6.50		31	Yucatan	14.44	
	28	Tmaulipas	15.77		32	Zacatecas	5.56	
	29	Tlaxcala	7.13			TOTAL	10.05	
	30	Veracruz	13.45		mx	1	Aguascalientes	11.92
	31	Yucatan	9.85		12	2	Baja California	10.70
	32	Zacatecas	13.97			3	Baja California Sur	13.16
	TOTAL	11.68			4	Campeche	6.82	
mx	1	Aguascalientes	18.45		5	Coahuila	12.50	
08	2	Baja California	19.49		6	Colima	9.84	
	3	Baja California Sur	17.93		7	Chiapas	5.43	
	4	Campeche	17.38		8	Chihuahua	11.33	
	5	Coahuila	17.72		9	Distr.Federal	9.65	
	6	Colima	12.05		10	Durango	11.40	
	7	Chiapas	12.19		11	Guanajuato	12.79	
	8	Chihuahua	14.15		12	Guerrero	16.72	
	9	Distr.Federal	16.38		13	Hidalgo	9.68	
	10	Durango	17.22		14	Jalisco	17.73	
	11	Guanajuato	16.81		15	Edo. de Mexico	8.82	
	12	Guerrero	18.64		16	Michoacan	8.28	
	13	Hidalgo	19.63		17	Morelos	13.40	
	14	Jalisco	24.24		18	Nayarit	7.26	
	15	Edo. de Mexico	18.90		19	Nuevo Leon	19.80	
	16	Michoacan	15.75		20	Oaxaca	8.79	
	17	Morelos	12.80		21	Puebla	9.24	
	18	Nayarit	19.67		22	Queretaro	9.09	
	19	Nuevo Leon	18.90		23	Quintana Roo	10.36	
	20	Oaxaca	11.04		24	San Luis Potosi	15.71	
	21	Puebla	12.47		25	Sinaloa	9.84	
	22	Queretaro	19.59		26	Sonora	14.71	
	23	Quintana Roo	20.05		27	Tabasco	14.61	

ENIGH 2004; Total nonresponse rate= 1-accepted/(originally selected + additional hhds)

ENIGH 2008 (INEGI 2009), Total nonresponse rate=1-accepted/drawn

ENIGH 2010 (INEGI 2011), Total nonresponse rate=1-accepted/drawn

ENIGH 2012 (INEGI 2013), Total nonresponse rate=1-accepted/drawn

28	Tmaulipas	21.36	
29	Tlaxcala	10.78	
30	Veracruz	12.62	
31	Yucatan	13.03	
32	Zacatecas	10.90	
	TOTAL	11.83	
pe04	1	Costa norte	18.89
	2	Costa centro	11.20
	3	Costa sur	22.06
	4	Sierra norte	10.20
	5	Sierra centro	9.45
	6	Sierra sur	19.73
	7	Selva	11.55
	8	Lima metropolitana	40.77
		TOTAL	17.74
pe07	1	Costa norte	11.41
	2	Costa centro	14.23
	3	Costa sur	23.92
	4	Sierra norte	14.83
	5	Sierra centro	13.91
	6	Sierra sur	19.13
	7	Selva	15.40
	8	Lima metropolitana	37.70
		TOTAL	18.46
pe10	1	Costa norte	19.77
	2	Costa centro	14.63
	3	Costa sur	18.17
	4	Sierra norte	14.50
	5	Sierra centro	11.52
	6	Sierra sur	22.03
	7	Selva	12.69
	8	Lima metropolitana	41.32
		TOTAL	19.11
pe13	1	Costa norte	18.18
	2	Costa centro	15.52
	3	Costa sur	24.87
	4	Sierra norte	6.89
	5	Sierra centro	5.98
	6	Sierra sur	15.18
	7	Selva	6.65
	8	Lima metropolitana	59.50
		TOTAL	18.37
rs06	1	Beogradski	11.00
	2	Vojvodine	7.00
	3	Central Serbia	1.10
		TOTAL	5.00
rs10	1	Beogradski	3.10
	2	Vojvodine	12.30
	3	Central Serbia	0.70
		TOTAL	4.50
rs13	1	Beogradski	4.90
	2	Vojvodine	14.70
	3	Šumadije i Zapadne	1.80
	4	Južne i Istočne	0.10
		TOTAL	5.90
ru07	100	Moscow and St. Pet.	81.90
	200	North & North West	51.50
	300	Centr.,C. Black-Earth	48.20
	400	Volgo-Vyatski, Basin	43.30
	500	North Caucasian	44.90
	600	Ural	43.00
	700	Western Siberian	41.60

Encuesta Nacional de Hogares Sobre Condiciones de Vida y Pobreza (ENVAHO) 2004, 2007, 2010 & 2013 (INEI 2008, 2010, 2011, 2014); nonresponding/drawn.

Hhd Budget Survey, contains substitutions up to 5 hhds per enumerated area, with 5-presampled hhds (+5 original hhds). Gross sample: 4800 + 4800 reserves; Unknown region\_c in LIS.

RLMS 2007, non-stratified at regional level; total cross-section nonresp. rt.

800	E. Siberian & Far E.	53.80	
	TOTAL	56.80	
ru10	100	Moscow and St. Pet.	79.80
	200	North & North West	37.70
	300	Centr.,C. Black-Earth	51.60
	400	Volga-Vaytski, Basin	47.70
	500	North Caucasian	54.50
	600	Ural	48.90
	700	Western Siberian	48.20
	800	E. Siberian & Far E.	54.40
		TOTAL	56.70
ru13	100	Moscow and St. Pet.	83.30
	200	North & North West	56.70
	300	Centr.,C. Black-Earth	54.20
	400	Volga-Vaytski, Basin	56.10
	500	North Caucasian	56.80
	600	Ural	45.60
	700	Western Siberian	53.10
	800	E. Siberian & Far E.	46.20
		TOTAL	61.00
se00	110	Stockholms	32.57
	121	Uppsala	22.57
	122	Södermanlands	21.90
	123	Östergötlands	22.35
	124	Örebro	24.91
	125	Västmanlands	22.30
	211	Jönköpings	28.47
	212	Kronobergs	26.37
	213	Kalmar	35.74
	214	Gotlands	31.02
	221	Blekinge	28.71
	224	Skåne	25.44
	231	Hallands	19.62
	232	Västra Götalands	20.76
	311	Värmlands	22.61
	312	Dalarnas	20.45
	313	Gävleborgs	21.24
	321	Västernorrlands	18.76
	322	Jämtlands	20.52
	331	Västerbottens	21.71
	332	Norrbottnens	23.93
		TOTAL	26.40
se05	110	Stockholms	37.42
	121	Uppsala	32.76
	122	Södermanlands	37.89
	123	Östergötlands	28.75
	124	Örebro	28.92
	125	Västmanlands	29.69
	211	Jönköpings	29.04
	212	Kronobergs	25.00
	213	Kalmar	26.52
	214	Gotlands	33.27
	221	Blekinge	31.56
	224	Skåne	31.56
	231	Hallands	36.55
	232	Västra Götalands	30.52
	311	Värmlands	40.38
	312	Dalarnas	35.22
	313	Gävleborgs	25.25
	321	Västernorrlands	23.04
	322	Jämtlands	26.67
	331	Västerbottens	30.57
	332	Norrbottnens	29.56
		TOTAL	32.75
sk04	1	Bratislava	16.64

RLMS 2010, 2013; non-stratified at regional level; total cross-sectional & longitudinal nonresponse rate

Income Distribution Survey (HINK) 2000, 2005; Source: Household's finances (HEK), Statistic Sweden; total nonresponse rate

	2	Trnava, Trenč., Nitra	13.44	
	3	Žilina, B.Bystrica	13.78	
	4	Prešov, Košice	15.34	
		TOTAL	14.44	
sk07	1	Bratislava	14.95	
	2	Trnava, Trenč., Nitra	13.85	
	3	Žilina, B.Bystrica	10.24	
	4	Prešov, Košice	10.43	
		TOTAL	12.09	
sk10	1	Bratislava	24.06	
	2	Trnava, Trenč., Nitra	9.96	
	3	Žilina, B.Bystrica	7.94	
	4	Prešov, Košice	7.91	
		TOTAL	10.36	
uk04	1	UKC North East	29.40	
	2	UKE Yorks & Humb.	33.96	
	3	UKD NW. & Mersey	29.97	
	4	UKF East Midlands	31.11	
	5	UKG W. Midlands	33.87	
	6	UKH Eastern	35.31	
	7	UKI London	35.48	
	8	UKJ South East	33.83	
	9	UKK South West	30.07	
	10	UKL Wales	29.26	
	11	UKM Scotland	29.34	
	12	UKN N. Ireland	33.69	
		TOTAL	32.40	
uk07	1	UKC North East	29.15	
	2	UKE Yorks & Humb.	30.40	
	3	UKD NW. & Mersey	31.87	
	4	UKF East Midlands	31.39	
	5	UKG W. Midlands	35.15	
	6	UKH Eastern	37.07	
	7	UKI London	34.18	
	8	UKJ South East	36.33	
	9	UKK South West	33.78	
	10	UKL Wales	30.26	
	11	UKM Scotland	26.53	
	12	UKN N. Ireland	34.78	
		TOTAL	32.49	
uk10	1	UKC North East	37.00	
	2	UKE Yorks & Humb.	39.00	
	3	UKD NW. & Mersey	39.00	
	4	UKF East Midlands	40.00	
	5	UKG W. Midlands	41.00	
	6	UKH Eastern	41.00	
	7	UKI London	49.00	
	8	UKJ South East	43.00	
	9	UKK South West	44.00	
	10	UKL Wales	36.00	
	11	UKM Scotland	38.00	
	12	UKN N. Ireland	39.00	
		TOTAL	41.00	
uk12	1	UKC North East	37.00	
	2	UKE Yorks & Humb.	38.00	
	3	UKD NW. & Mersey	37.00	
	4	UKF East Midlands	37.00	
	5	UKG W. Midlands	41.00	
	6	UKH Eastern	40.00	

EU-SILC 2011, LIS tabulation, DB135=2

Family Resources Survey 2004-05, 2007-08, 2010-11, 2012-13 (ONS 2005, 2008, 2013; DWP 2014); total nonresponse (incomplete, refused and failed)

	7	UKI London	51.00	
	8	UKJ South East	39.00	
	9	UKK South West	40.00	
	10	UKL Wales	37.00	
	11	UKM Scotland	41.00	
	12	UKN N. Ireland	39.00	
		TOTAL	40.00	
uy04	1	Montevideo	5.24	
	2	Artigas	3.94	
	3	Canelones	4.50	
	4	Cerro Largo	0.56	
	5	Colonia	4.37	
	6	Durazno	6.48	
	7	Flores	3.89	
	8	Florida	1.85	
	9	Lavalleja	0.79	
	10	Maldonado	6.14	
	11	Paysandú	13.03	
	12	Río Negro	2.78	
	13	Rivera	1.07	
	14	Rocha	4.29	
	15	Salto	5.98	
	16	San José	3.14	
	17	Soriano	1.94	
	18	Tacuarembó	4.17	
	19	Treinta y Tres	2.31	
		TOTAL	4.85	
uy07	1	Montevideo	2.67	
	2	Artigas	0.29	
	3	Canelones	2.55	
	4	Cerro Largo	0.08	
	5	Colonia	0.76	
	6	Durazno	0.54	
	7	Flores	0.38	
	8	Florida	0.15	
	9	Lavalleja	0.99	
	10	Maldonado	5.50	
	11	Paysandú	0.14	
	12	Río Negro	5.13	
	13	Rivera	0.65	
	14	Rocha	0.43	
	15	Salto	0.14	
	16	San José	0.34	
	17	Soriano	0.52	
	18	Tacuarembó	1.00	
	19	Treinta y Tres	0.12	
		TOTAL	1.93	
uy10	1	Montevideo	5.87	
	2	Artigas	5.43	
	3	Canelones	10.61	
	4	Cerro Largo	11.95	
	5	Colonia	13.13	
	6	Durazno	6.49	
	7	Flores	8.42	
	8	Florida	8.27	
	9	Lavalleja	13.18	
	10	Maldonado	21.88	
	11	Paysandú	7.28	
	12	Río Negro	14.24	
	13	Rivera	10.94	
	14	Rocha	26.79	
	15	Salto	6.23	
	16	San José	3.38	
	17	Soriano	6.23	
	18	Tacuarembó	7.24	
	19	Treinta y Tres	11.37	
		TOTAL	9.62	
uy13	1	Montevideo	7.49	
	2	Artigas	9.57	
	3	Canelones	19.45	

Encuesta Continua de Hogares (ECH) 2004; only locations with over 5,000 residents; Selected/theoretical pop. drawn; nonresponse due to non-contact, non-belonging to universe

ECH 2007, 2010, 2013; Selected/theoretical pop. drawn; nonresponse due to non-contact, non-belonging to universe



Table A2. State nonresponse rates, type A, for US Current Population Survey (%)

	Reference yr	79	86	91	94	97	00	04	07	10	13
<i>region_c</i>	(CPS yr)	(80)	(87)	(92)	(95)	(98)	(01)	(05)	(08)	(11)	(14)
11	Maine	1.83	2.64	2.36	2.05	4.28	7.40	8.79	8.46	9.76	15.87
12	N. Hampshire	3.17	2.42	4.15	4.67	7.25	12.12	10.12	8.63	9.89	15.25
13	Vermont	2.82	3.47	2.58	3.74	4.84	10.67	12.23	8.51	10.48	15.37
14	Massachusetts	4.46	6.02	5.99	3.66	8.29	8.15	11.43	9.88	11.00	15.07
15	Rhode Island	2.42	4.98	4.65	4.87	7.86	10.20	7.90	12.66	11.09	15.06
16	Connecticut	3.47	6.65	5.56	1.79	6.23	7.09	12.37	12.51	11.03	16.00
21	New York	5.26	8.54	6.67	5.92	11.58	10.80	13.18	11.71	10.24	14.01
22	New Jersey	5.35	8.06	8.07	4.77	10.19	11.77	12.75	7.72	9.18	12.77
23	Pennsylvania	4.36	4.68	4.16	4.08	8.58	9.56	9.03	6.97	8.57	10.47
31	Ohio	4.61	4.59	4.82	2.86	7.10	6.95	8.51	8.53	6.88	12.37
32	Indiana	3.11	3.33	2.54	2.09	5.93	7.94	4.06	5.12	5.65	9.36
33	Illinois	6.83	4.03	4.97	4.29	8.04	9.26	6.08	8.23	5.29	8.45
34	Michigan	4.28	6.65	4.56	3.97	8.58	8.37	7.39	6.81	8.69	9.21
35	Wisconsin	3.42	2.52	3.38	2.04	8.06	6.75	5.26	4.53	4.93	6.34
41	Minnesota	2.83	1.69	3.17	2.43	4.17	6.72	9.03	6.44	6.10	10.23
42	Iowa	3.02	4.02	3.58	1.05	4.93	5.65	6.95	4.38	4.04	7.69
43	Missouri	3.42	3.58	3.11	2.22	7.48	8.07	6.89	8.80	7.40	11.19
44	North Dakota	2.67	3.21	4.44	3.59	5.47	5.38	5.24	5.32	6.35	9.84
45	South Dakota	3.18	5.25	5.88	3.94	6.24	4.49	7.37	6.06	4.54	6.73
46	Nebraska	3.60	8.02	4.70	3.18	7.12	5.43	8.05	8.51	7.73	12.29
47	Kansas	3.69	4.58	4.63	2.07	6.62	7.89	7.13	6.09	7.22	9.55
51	Delaware	5.51	8.08	2.88	5.69	6.97	8.33	8.60	9.79	8.97	9.37
52	Maryland	4.03	4.17	4.37	3.72	11.11	13.27	15.11	11.06	11.70	16.24
53	DC	6.64	10.39	8.69	13.61	12.52	18.62	17.39	12.02	10.50	13.61
54	Virginia	3.69	5.66	5.15	4.63	6.16	7.38	6.42	4.74	6.07	9.62
55	West Virginia	3.20	3.05	1.64	1.58	5.30	4.34	5.47	3.43	5.48	7.14
56	N. Carolina	3.48	3.90	4.62	4.29	5.89	7.30	8.88	5.91	10.66	8.63
57	S. Carolina	2.80	2.27	3.69	3.00	6.79	6.53	8.35	5.09	8.23	6.23
58	Georgia	3.76	5.63	4.50	4.02	3.89	6.78	7.29	6.69	6.98	7.25
59	Florida	3.79	5.45	4.80	5.28	7.31	4.85	6.58	4.69	4.99	5.81
61	Kentucky	3.66	2.85	2.73	2.43	8.09	7.75	6.30	6.63	6.46	9.19
62	Tennessee	3.34	4.03	3.00	3.17	7.20	6.25	8.00	8.45	8.34	10.65
63	Alabama	2.92	3.96	3.00	1.09	2.03	1.89	3.50	4.55	4.59	6.59
64	Mississippi	2.71	4.74	3.91	3.75	5.19	4.88	8.85	6.03	7.37	9.56
71	Arkansas	2.35	4.38	4.08	3.38	5.37	6.82	4.07	5.77	7.15	6.13
72	Louisiana	2.59	3.23	4.16	3.33	4.59	3.70	6.09	6.99	9.61	8.64
73	Oklahoma	3.73	6.61	4.79	5.81	5.46	6.90	6.80	4.67	5.65	7.45
74	Texas	3.89	5.46	5.28	4.51	7.00	7.83	10.10	8.66	9.37	12.03
81	Montana	5.34	3.30	5.66	3.30	6.69	3.98	6.56	5.93	4.91	5.83
82	Idaho	3.87	3.06	3.55	3.13	4.40	3.52	6.48	6.57	7.20	8.06
83	Wyoming	3.90	5.49	3.07	2.50	6.12	6.52	7.88	7.88	8.23	11.00
84	Colorado	3.82	5.09	5.86	8.70	8.99	9.95	11.01	8.19	7.07	9.20
85	New Mexico	3.23	5.48	4.43	6.57	8.90	7.56	11.42	9.10	8.63	10.90
86	Arizona	4.52	5.60	7.77	6.52	9.07	8.61	10.32	7.88	8.63	11.91
87	Utah	4.45	4.29	5.54	2.92	6.86	3.99	5.30	7.26	7.69	9.16
88	Nevada	5.90	5.53	4.43	5.90	9.30	7.94	8.80	8.14	11.02	9.97
91	Washington	4.66	6.91	5.76	6.00	9.63	8.01	9.15	7.64	8.50	9.44
92	Oregon	5.37	3.65	4.13	4.35	6.64	8.51	9.44	7.40	10.40	11.57
93	California	6.04	6.51	6.32	7.15	9.81	9.79	11.05	8.42	8.65	7.38
94	Alaska	8.69	7.99	6.18	2.87	11.41	8.72	11.28	8.05	7.43	13.27
95	Hawaii	5.82	7.64	3.35	3.63	11.70	11.16	11.94	9.75	8.43	7.85
	TOTAL	4.30	5.43	5.02	4.53	7.79	8.03	9.01	7.82	8.12	10.26

Notes: LIS survey identifier uses income reference year, which typically lags behind the official survey year by one. If *dname* is YY (e.g., us13), the official US CPS wave is YY+1 (US CPS 2014).

US CPS data are from CPS Annual Social and Economic Supplement (ASEC), called March Supplement until 2002. us13 survey uses original 'traditional' CPS 2014 data as opposed to redesigned data. In US-CPS for 1970 and 1975 nonresponding units are not reported on. Nonresponse rates are computed out of responders + type A nonresponders. Type A includes refusals to be interviewed and failed interviews because nobody was at home. Omitted from the calculation are types B and C nonresponders, that is vacant or demolished housing units, and group quarters including boarding houses, transient residences (hotels), tents, trailer sites, barracks and other institutional residences.

The CPS categorizes households into 254 adjustment cells. These consist of areas within the same metropolitan statistical area (MSA) or MSAs of similar size and within the same state. MSAs are further split into central and noncentral city cells, and non-MSA areas are split into

urban and rural cells. The sampling weight of nonrespondents is redistributed to the other households in the adjustment cell.

### *Nonresponse rates in EU-SILC surveys*

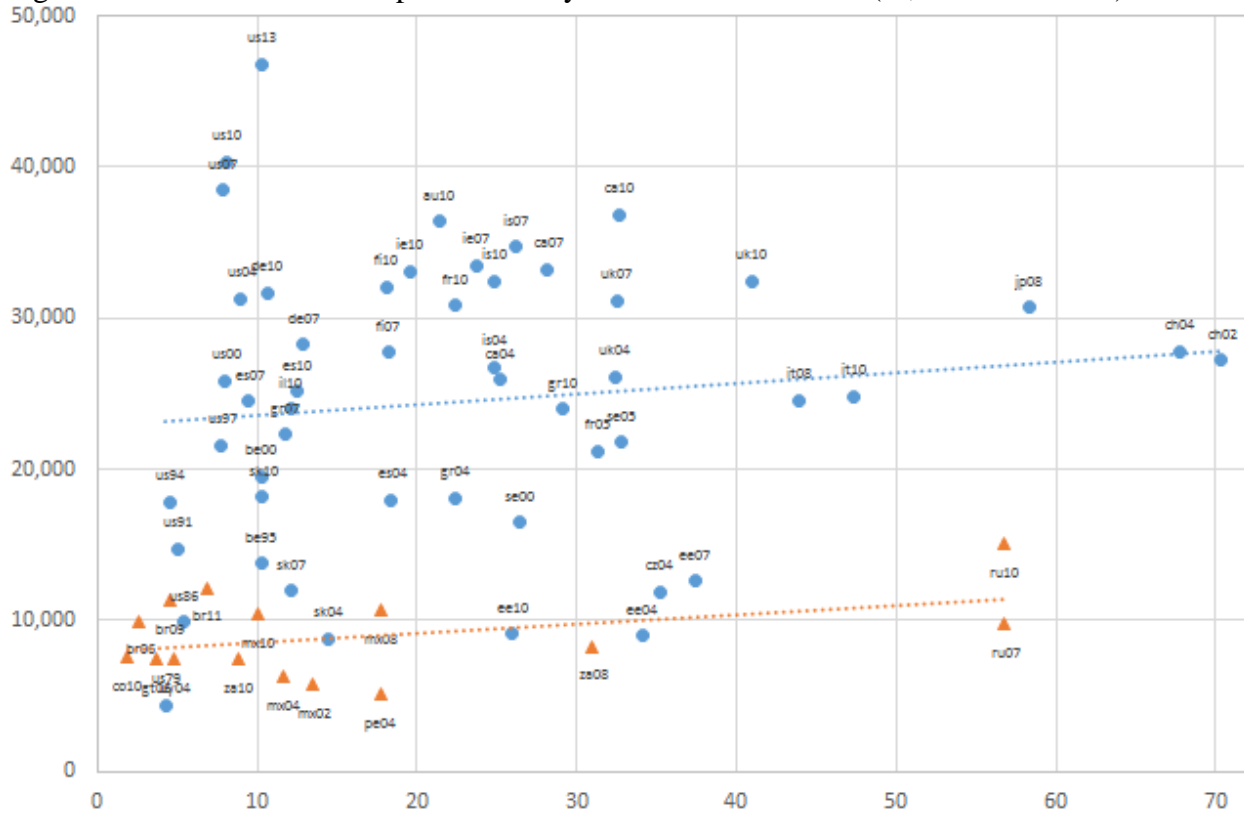
Household nonresponse rates in EU-SILC surveys are computed using Eurostat notation as follows:

$$NRh = 1 - \frac{\sum 1(DB120 = 11)}{\underbrace{\sum 1(DB120 \neq \emptyset) - \sum 1(DB120 = 23)}_{\text{Address contact rate}}} \frac{\sum 1(DB135 = 1)}{\underbrace{\sum 1(DB130 \neq \emptyset)}_{\text{Rate of complete interviews accepted}}},$$

where  $1(\cdot)$  is a binary indicator function, DB120 is the record of contact at the address, DB130 is the household questionnaire result and DB135 is the household interview acceptance result. Addresses that could not be located or accessed ( $DB120 \leq 22$ ) are accounted for in the address contact rate, while non-existing, non-residential, non-occupied and non-principal residence addresses ( $DB120=23$ ) are omitted. Rate of complete interviews accepted is the accepted interviews (i.e., at least one personal interview in household accepted) among all households completing, refusing to cooperate, temporarily absent, or unable to respond due to illness, incapacity, language or other problems.

EU-SILC surveys in some member states use information from national registers for some income components (Jantti *et al.* 2013). After these data were introduced to the French EU-SILC survey, Gini coefficient was found to increase from 39 to 44, suggesting that underestimating of top incomes had been a problem (Burrigand 2013). The use of national registers clearly affects nonresponse rates.

Figure A1. Nationwide nonresponse rates by national income level (% , 2005 US\$ PPP)

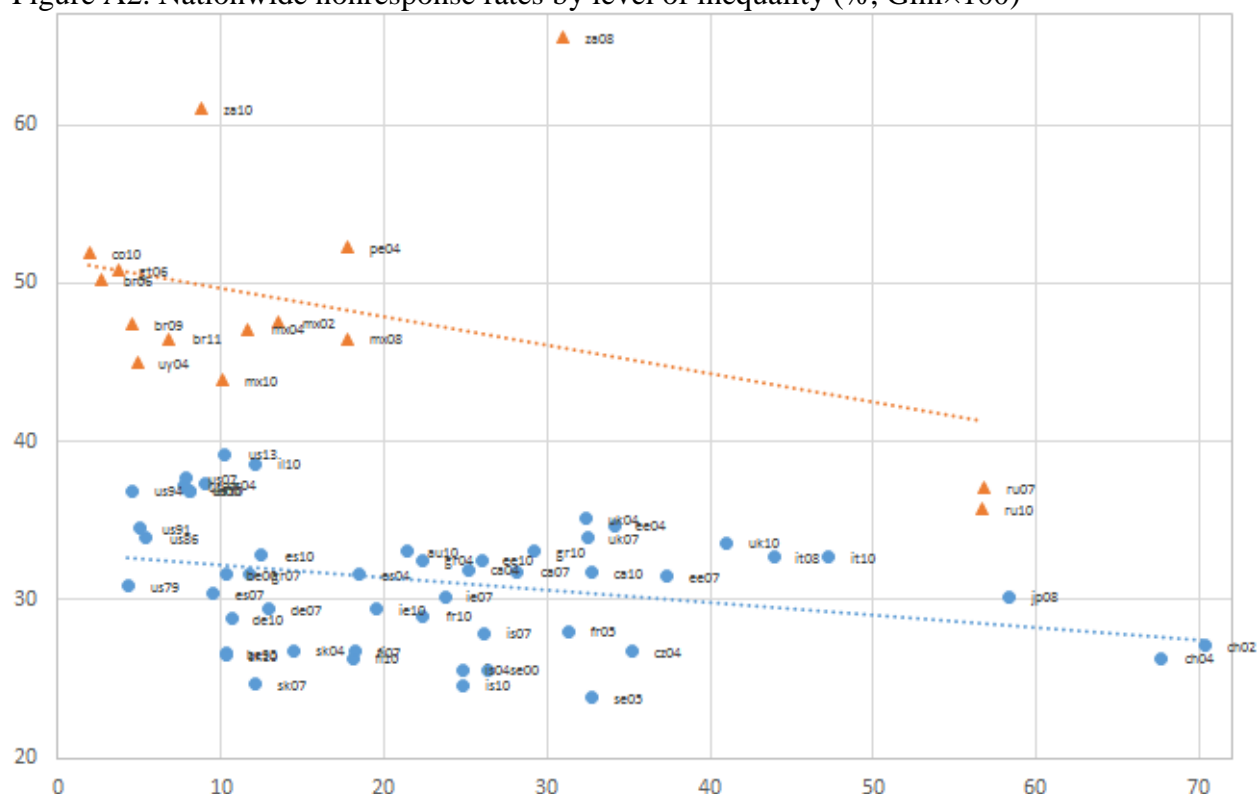


Note: Nonresponse rates on x-axis; mean national incomes on y-axis.

Blue circles – high income countries; orange triangles – middle income countries.

Interestingly, linear relationship has the same slope for both middle and high income countries.

Figure A2. Nationwide nonresponse rates by level of inequality (% , Gini×100)



Note: Nonresponse rates on x-axis; Gini coefficients (×100) on y-axis.

Blue circles – high income countries; orange triangles – middle income countries.

#### *Survey documentation and sources of nonresponse-rate data*

*be95*: Dewilde, Caroline, Marynissen, Rudy, Decoster, Kristof, and Jacobs, Therese (1999), *Methodebericht, Golf 5, 1996, Steunpunt Gezinsdemografisch Panel, Universiteit Antwerpen.*

*be00*: Ottoy, Winfried, Verstreken, Martine, Marynissen, Rudy, and Mortelmans, Dimitri (2004), *Methodebericht, Golf 10, 2001, Steunpunt Gezinsdemografisch Panel, Universiteit Antwerpen.*

*br06*: IBGE (2006), *Notas Tecnicas, Pesquisa Nacional por Amostra de Domicilios, v.27, 2006.*

*br09*: IBGE (2009), *Notas Tecnicas, Sintese de Indicadores 2009, Pesquisa Nacional por Amostra de Domicilios.*

*br11*: IBGE (2011), *Notas Tecnicas, Pesquisa Nacional por Amostra de Domicilios, v.31, 2011.*

*ca04*: Duddek, Chris, and Storey, Susan (2006), *Data Quality in the 2004 Survey of Labour and Income Dynamics (SLID), Income Research Paper Series, Catalogue no. 75F0002MIE, Vol. 5, Ottawa: Statistics Canada.*

*ca07*: Bastien, Jean-Francois (2009), *Data Quality for the 2007 Survey of Labour and Income Dynamics (SLID), Ottawa: Statistics Canada.*

*co10*: Departamento Administrativo Nacional de Estadística (DANE, 2011), Tasa de respuesta y no respuesta de acuerdo a información de resumen de cobertura enviado por cada ciudad, GEIH Marco 2005 total nacional - Etapa 01101 - Enero 2011.

*co13*: Departamento Administrativo Nacional de Estadística (DANE, 2011), Tasa de respuesta y no respuesta de acuerdo a información de resumen de cobertura enviado por cada ciudad, GEIH Marco 2005 total nacional - Etapa 01301 - Enero 2013.

*cz04*: Czech Statistical Office (CZSO, 2007), EU-SILC 2005 operation, Final quality report, Czech Republic, 20.Dec.2007.

*cz07*: Czech Statistical Office (CZSO, 2010), EU-SILC 2008 operation, Final quality report, Czech Republic, Dec.2010.

*cz10*: Czech Statistical Office (CZSO, 2012), EU-SILC 2011 operation, Intermediate quality report, Czech Republic, July 2012.

*do07*: Oficina Nacional de Estadística (ONE, 2008), Encuesta Nacional de Ingresos y Gastos de los Hogares, ENIGH, 2007: Resultados Generales.

*eg12*: Assaad, Ragui, and Krafft, Caroline (2013), The Egypt labor market panel survey: introducing the 2012 round, IZA Journal of Labor & Development 2(1):1-30.

*ee04, ee07*: Statistics Estonia (2010), The Estonian Social Survey: Methodological report, Tallinn March 2010.

*es04*: Instituto Nacional de Estadística (INE, 2005), Evaluación de la falta de respuesta en la encuesta de condiciones de vida 2005.

*es07*: Instituto Nacional de Estadística (ES-INE, 2008), Evaluación de la falta de respuesta en la encuesta de condiciones de vida 2008.

*es10*: Instituto Nacional de Estadística (ES-INE, 2011), Evaluación de la falta de respuesta en la encuesta de condiciones de vida 2011.

*es13*: Instituto Nacional de Estadística (ES-INE, 2014), Evaluación de la falta de respuesta en la encuesta de condiciones de vida 2014.

*gr04*: National Statistical Service of Greece (Statistics Greece, 2010a), Statistics on Income and Living Conditions (EU-SILC) 2007 longitudinal component 2004-2007, Final quality report, Piraeus, February 2010.

*gr07*: National Statistical Service of Greece (Statistics Greece, 2010b), Statistics on Income and Living Conditions (EU-SILC) 2008, Intermediate quality report, Piraeus, March 2010.

*gr10*: National Statistical Service of Greece (Statistics Greece, 2012), Statistics on Income and Living Conditions (EU-SILC) 2011, Intermediate quality report, Piraeus, December 2012.

*gt06*: Instituto Nacional de Estadística (GT-INE, 2006), Encuesta Nacional de Condiciones de Vida, ENCOVI 2006: Diseño de la muestra, Guatemala, February 2006.

*hu07*: Tarki (2008), Koz, Teher, Elosztas: Tarki monitor jelentések 2008, Budapest, March 2008.

*hu12*: Tarki (2013), Egyenlit lenség és polarizálódás a magyar társadalomban: Tarki monitor jelentések 2012, Budapest, April 2013.

*it04*: Faiella, Ivan, and Romina Gambacorta (2007), The Weighting Process in the SHIW, Temi di discussione del Servizio Studi, number 636, Bank of Italy, June 2007; and studies cited therein.

*jp08*: Panel Data Research Center (PDRC, 2015), Survey methodology: Keio Household Panel Survey (KHPS) and Japan Household Panel Survey Data (JHPS), unpublished manuscript, Keio University.

*mx02*: Instituto Nacional de Estadística y Geografía (INEGI, 2003), Encuesta Nacional de Ingresos y Gastos de los Hogares 2002, Metodología: Selección de la muestra, original data CD rom.

*mx04*: Instituto Nacional de Estadística y Geografía (INEGI, 2005), Encuesta Nacional de Ingresos y Gastos de los Hogares 2004, Publicación bienal. Primera edición: La Estructura, Distribución y Monto de los Ingresos de los Hogares.

*mx08*: Instituto Nacional de Estadística y Geografía (INEGI, 2009), Encuesta Nacional de Ingresos y Gastos de los Hogares 2008: Diseño muestral.

*mx10*: Instituto Nacional de Estadística y Geografía (INEGI, 2011), Encuesta Nacional de Ingresos y Gastos de los Hogares 2010: Diseño muestral.

*mx12*: Instituto Nacional de Estadística y Geografía (INEGI, 2013), Encuesta Nacional de Ingresos y Gastos de los Hogares 2012: Diseño muestral.

*pe04*: Instituto Nacional de Estadística e Informática (INEI, 2010), Encuesta Nacional de Hogares Sobre Condiciones de Vida y Pobreza – ENAHO 2004: Ficha Técnica.

*pe07*: Instituto Nacional de Estadística e Informática (INEI, 2008), Encuesta Nacional de Hogares Sobre Condiciones de Vida y Pobreza – ENAHO 2007: Ficha Técnica.

*pe10*: Instituto Nacional de Estadística e Informática (INEI, 2011), Encuesta Nacional de Hogares - Condiciones de Vida y Pobreza, Enero a Diciembre 2010: Metodología y Cuestionario, June 2011.

*pe13*: Instituto Nacional de Estadística e Informática (INEI, 2014), Encuesta Nacional de Hogares Sobre Condiciones de Vida y Pobreza – ENAHO 2013: Ficha Técnica.

*sk04*: Statistical Office of the Slovak Republic (ŠÚSR, 2006), Intermediate quality report - Statistics on Income and Living Conditions (EU SILC 2005), Bratislava, August 2006.

*sk07*: Statistical Office of the Slovak Republic (ŠÚSR, 2009), Intermediate quality report - Statistics on Income and Living Conditions (EU SILC 2008), Bratislava, September 2009.

*uk04*: Office for National Statistics (ONS, 2005), Family Resources Survey - Annual Technical report 2004/5, Appendices 13-17, document code tcm77-154289, ONS: London, UK.

*uk07*: Office for National Statistics (ONS, 2008), Family Resources Survey - Annual Technical report 2007/8, Appendices 16-20, document code tcm77-156301, ONS: Newport, UK.

*uk10*: Department for Work and Pensions, United Kingdom (DWP, 2013), Family Resources Survey, United Kingdom 2010/11, June 2012, [www.gov.uk/government/statistics/family-resources-survey-201011](http://www.gov.uk/government/statistics/family-resources-survey-201011).

*uk12*: Department for Work and Pensions, United Kingdom (DWP, 2014), Family Resources Survey, United Kingdom 2012/13, July 2014, <https://www.gov.uk/government/statistics/family-resources-survey-2012-to-2013>.

*za08*: Leibbrandt, Murray, Woolard, Ingrid, and de Villiers, Louise (2009), Methodology: Report on NIDS Wave 1, Technical Paper no. 1, July 2009.

*za10*: Brown, Michael, Daniels, Reza C., de Villiers, Louise, Leibbrandt, Murray, and Woolard, Ingrid (2014), National Income Dynamics Study wave 2 user manual, Version 2.1, 28 February 2014.

*za12*: de Villiers, Louise, Brown, Michael, Woolard, Ingrid, Daniels, Reza C., and Leibbrandt, Murray (2013), National Income Dynamics Study wave 3 user manual, Version 1.1, 9 September 2013.

*us86-us13*: Data sources: NBER CPS Supplements, [www.nber.org/data/cps.html](http://www.nber.org/data/cps.html); Roth, Jean (2015), Reading Current Population Survey (CPS) Data with SAS, SPSS, or Stata, [www.nber.org/data/cps\\_progs.html](http://www.nber.org/data/cps_progs.html), accessed on 23 October 2015.

## **Appendix 2: Delineation of world regions**

### *Europe*

Albania, Andorra, Armenia, Austria, Azerbaijan, Belarus, Belgium, Bosnia and Herzegovina, Bulgaria, Croatia, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Georgia, Germany, Greece, Hungary, Iceland, Ireland, Italy, Kazakhstan, Latvia, Liechtenstein, Lithuania,

Luxembourg, Macedonia, Malta, Moldova, Monaco, Montenegro, Netherlands, Norway, Poland, Portugal, Romania, Russia (5 regions: Moscow and St. Petersburg, Northern and North Western, Central and Central Black-Earth, Volgo-Vyatski and Volga Basin, North Caucasian regions), San Marino, Serbia, Slovakia, Slovenia, Spain, Sweden, Switzerland, Ukraine, UK. (Turkey is excluded.)

In the LIS database, the following surveys are used: *be00, cz04, ee10, fi10, fr10, de10, gr10, is10, ie10, it10, ru10, sk10, es10, se05, ch04, uk10*. Russia is included in this count despite the fact that 75% of its territory, 22% of population, and 3 of its 8 regions are in Asia. In the analysis for Europe, only Russia's 5 solely-European regions accounting for 78% of Russia's population – Moscow & St. Petersburg, Northern & North Western, Central & Central Black-Earth, Volgo-Vyatski & Volga Basin, and North Caucasian regions – are included. Ural region, the majority of which lies in Asia, is excluded. Only 78% of Russia's population is counted toward European population.

When justifying the use of the LIS sample for worldwide analysis, we label Russia a central Eurasian country, because Russian households are thought to be similar economically and behaviorally to those in surrounding central Eurasian countries on both sides of the Ural mountain range. This labeling has no impact on the actual analysis of worldwide inequality.

#### *European Single Market + European city states*

Andorra, Austria, Belgium, Bulgaria, Croatia, Cyprus, Czech Rep., Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Latvia, Liechtenstein, Lithuania, Luxembourg, Malta, Monaco, Netherlands, Norway, Poland, Portugal, Romania, San Marino, Slovakia, Slovenia, Spain, Sweden, Switzerland, UK.

In the LIS database, all European countries except for Russia are used, namely: *be00, cz04, ee10, fi10, fr10, de10, gr10, is10, ie10, it10, sk10, es10, se05, ch04, uk10*.

#### *North America*

Canada, Mexico, US (all available from LIS).

#### *Latin America (excl. Mexico and Haiti)*

Antigua and Barbuda, Argentina, Bahamas, Barbados, Belize, Bolivia, Brazil, Chile, Colombia, Costa Rica, Cuba, Dominica, Dominican Republic, Ecuador, El Salvador, French Guiana, Grenada, Guatemala, Guyana, Honduras, Jamaica, Nicaragua, Panama, Paraguay, Peru, St. Kitts and Nevis, St. Lucia, St. Vincent and The Grenadines, Suriname, Trinidad and Tobago, Uruguay, Venezuela.

In the LIS database, the following surveys are used: *br11, co10, fr10, gt06, pe04, uy04*. From the *fr10* survey, only French Guiana (DR9C Guyane, LIS *region\_c* 11) is used. Haiti is excluded as the only low income country on the continent, which is also absent from LIS database. Inference to Haiti is thus impossible.

### *Lower-middle income countries*

Armenia, Bangladesh, Bhutan, Bolivia, Cabo Verde, Cameroon, Congo Rep., Cote d'Ivoire, Djibouti, Egypt, El Salvador, Georgia, Ghana, Guatemala, Guyana, Honduras, India, Indonesia, Kenya, Kiribati, Kosovo, Kyrgyz Rep., Laos, Lesotho, Mauritania, Micronesia, Moldova, Morocco, Myanmar, Nicaragua, Nigeria, Pakistan, Papua New Guinea, Philippines, Samoa, Sao Tome and Principe, Senegal, Solomon Islands, Sri Lanka, Sudan, Swaziland, Syria, Tajikistan, Timor-Leste, Ukraine, Uzbekistan, Vanuatu, Vietnam, West Bank and Gaza, Yemen, Zambia. In the LIS database, only *gt06* is available.

### *Upper-middle income countries*

Albania, Algeria, American Samoa, Angola, Azerbaijan, Belarus, Belize, Bosnia and Herzegovina, Botswana, Brazil, Bulgaria, China, Colombia, Costa Rica, Cuba, Dominica, Dominican Rep., Ecuador, Fiji, Gabon, Grenada, Iran, Iraq, Jamaica, Jordan, Kazakhstan, Lebanon, Libya, Macedonia, Malaysia, Maldives, Marshall Islands, Mauritius, Mexico, Mongolia, Montenegro, Namibia, Palau, Panama, Paraguay, Peru, Romania, Russia, Serbia, South Africa, St. Lucia, St. Vincent and the Grenadines, Suriname, Thailand, Tonga, Tunisia, Turkey, Turkmenistan, Tuvalu, Uruguay.

In the LIS database, the following surveys are used: *br11*, *co10*, *mx10*, *pe04*, *ru10*, *uy04*, *za10*. In the analysis for Latin America, French Guiana (DR9C Guyane, LIS *region\_c* 11, in *fr10* survey) is reclassified as an upper-middle income region. Compared to World Bank's (2015c) classification, Russia and Uruguay are classified here as upper-middle (rather than high) income countries in agreement with LIS's classification and with the countries' status in 2010 and 2004, respectively.

### *High income countries*

Andorra, Antigua and Barbuda, Argentina, Aruba, Australia, Austria, Bahamas, Bahrain, Barbados, Belgium, Bermuda, Brunei Darussalam, Canada, Cayman Islands, Channel Islands, Chile, Croatia, Curacao, Cyprus, Czech Rep., Denmark, Equatorial Guinea, Estonia, Faeroe Islands, Finland, France, French Polynesia, Germany, Greece, Greenland, Guam, Hong Kong, Hungary, Iceland, Ireland, Isle of Man, Israel, Italy, Japan, Korea, Kuwait, Latvia, Liechtenstein, Lithuania, Luxembourg, Macao, Malta, Monaco, Netherlands, New Caledonia, New Zealand, Northern Mariana Islands, Norway, Oman, Poland, Portugal, Puerto Rico, Qatar, San Marino, Saudi Arabia, Seychelles, Singapore, Sint Maarten (Netherlands), Slovakia, Slovenia, Spain, St. Kitts and Nevis, St. Martin (France), Sweden, Switzerland, Trinidad and Tobago, Turks and Caicos Islands, UAE, UK, USA, Venezuela, Virgin Islands.

In the LIS database, the following surveys are used: *au10*, *be00*, *ca10*, *cz04*, *ee10*, *fi10*, *fr10*, *de10*, *gr10*, *ie10*, *is10*, *il10*, *it10*, *jp08*, *sk10*, *es10*, *se05*, *ch04*, *uk10*, *us10*.

Source of world regions: own analysis by authors.

Source of income-bracket groups of countries: World Bank (2015c).





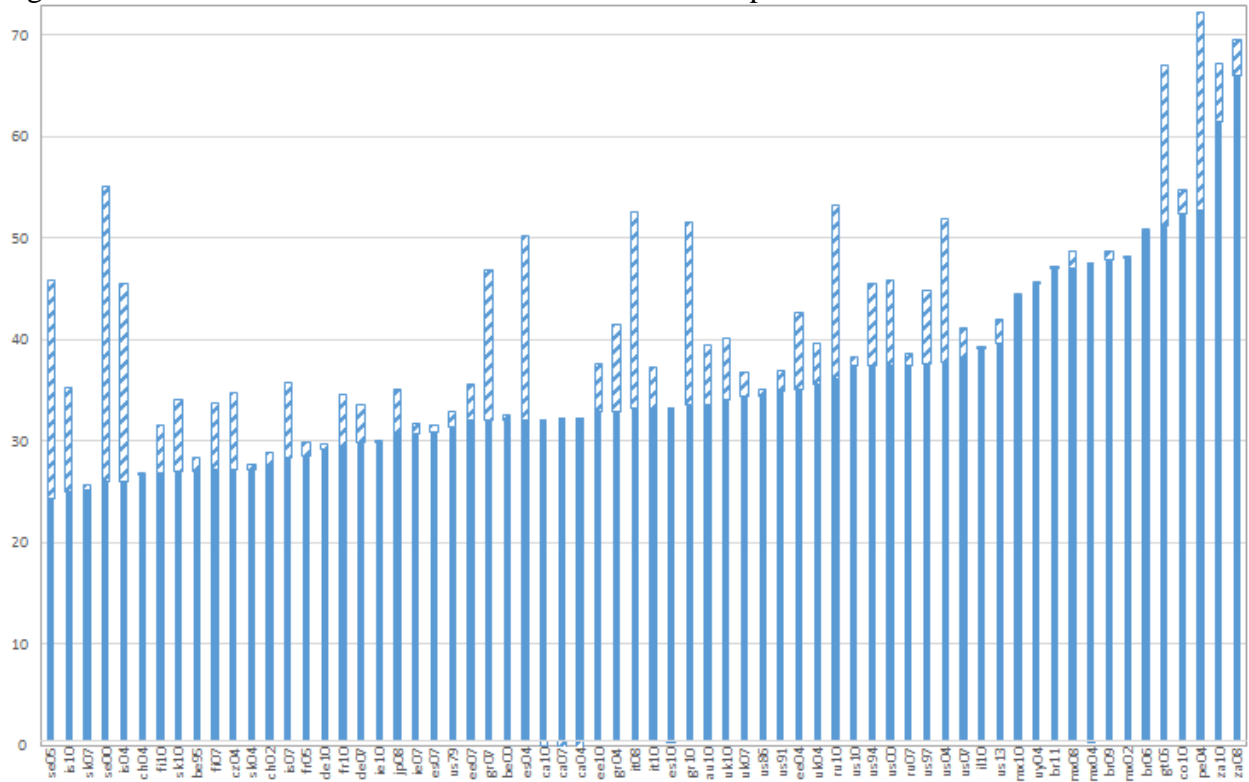
Table A5. Additional inequality measures for world regions: richest-household income shares

	Uncorrected income distributions					Unit-nonresponse corrected distributions				
	Top 0.5% share	Top 1% share	Top 5% share	Top 10% share	Top 20% share	Top 0.5% share	Top 1% share	Top 5% share	Top 10% share	Top 20% share
EU Single Market	3.64	5.67	16.75	26.91	42.84	12.82	16.04	28.95	39.09	53.78
EU	3.73	5.87	17.42	27.97	44.41	19.02	22.83	35.83	46.20	60.27
North America	4.15	6.53	19.17	30.56	48.14	5.80	8.62	22.20	33.81	51.27
Latin America	6.17	9.64	25.66	37.86	54.34	29.05	33.79	48.80	58.15	69.74
World (28) <sup>a</sup>	5.33	8.36	24.02	37.14	55.39	16.70	20.93	37.29	50.58	67.21
World (34) <sup>b</sup>	5.36	8.40	24.13	37.28	55.57	20.15	24.33	40.43	52.97	68.77

<sup>a</sup> 28 countries: *au10, be00, br11, ca10, co10, cz04, ee10, fi10, fr10, de10, gr10, gt06, is10, ie10, il10, it10, jp08, mx10, pe04, ru10, sk10, za10, es10, se05, ch04, uk10, us10* and *uy04*. Excluding low-income countries, for lack of data. Extrapolation to high, upper-middle and lower-middle countries done by augmenting household weights by author-computed subsampling rates of 70.838, 23.429 and 0.509%, respectively.

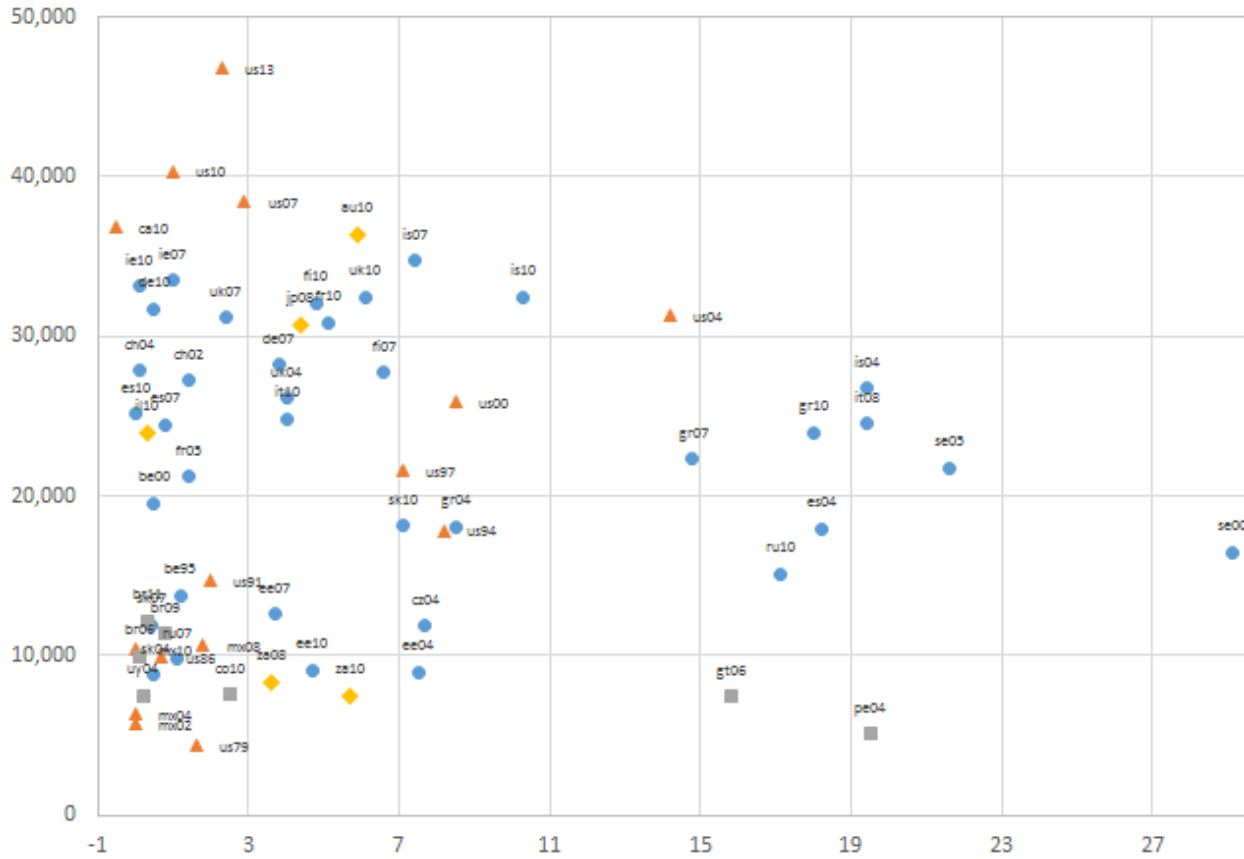
<sup>b</sup> 34 countries: *at04, kr06, lu10, nl10, pl10* and *ro97* added. For the first 28, subnational regional nonresponse rates are used. For the latter 6, national nonresponse rates are used – surveys are not disaggregated geographically. Excluding low-income countries, for lack of data. Extrapolation to high, upper-middle and lower-middle countries done by augmenting household weights by author-computed subsampling rates of 80.571, 24.186 and 0.509%, respectively.

Figure A3. Corrections of Gini coefficients for unit nonresponse bias



Solid bars: uncorrected Gini; dashed bars: nonresponse-bias corrections.

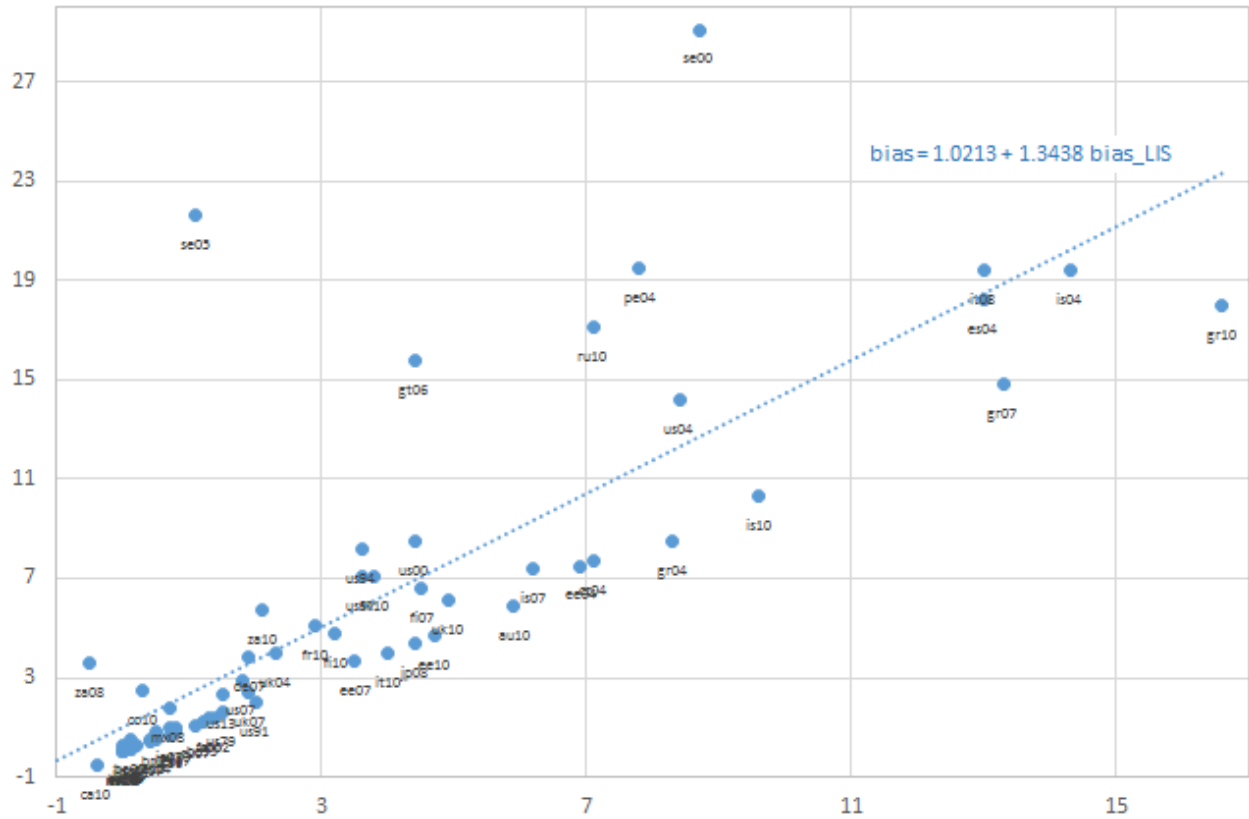
Figure A4. Unit-nonresponse bias by world region and national income level: 66 surveys



y-axis: mean national income (2005 US\$ PPP); x-axis: bias in non-winsorized income distributions.

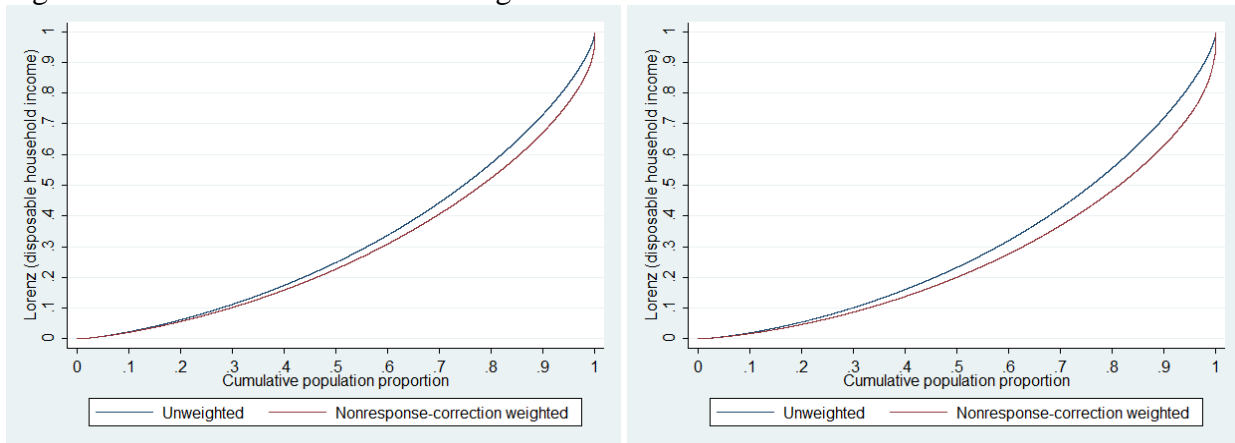
Blue circles – Europe; orange triangles – North America; grey squares - Latin America; yellow diamonds – Rest of the world.

Figure A5. Bias estimated in non-winsorized vs. LIS-winsorized income distributions: 66 surveys



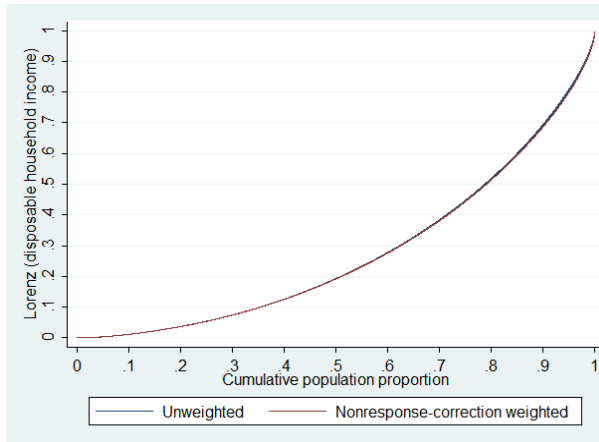
y-axis: bias in original income distributions; x-axis: bias in LIS-winsorized income distributions.

Figure A6. Lorenz curves for world regions on uncorrected and corrected income distributions

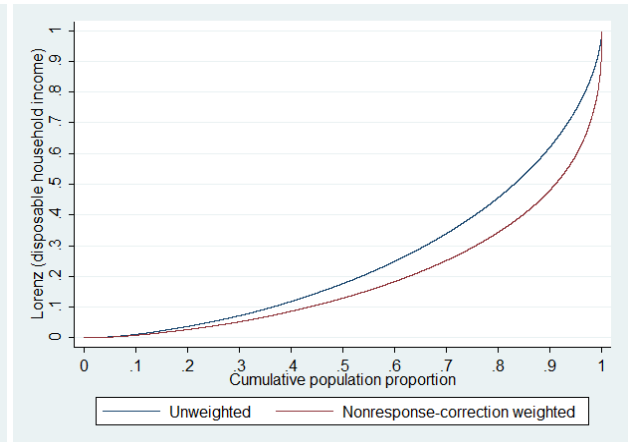


i. EU Single Market

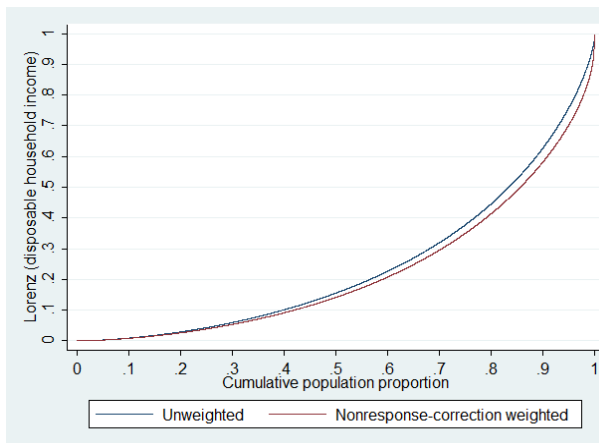
ii. EU



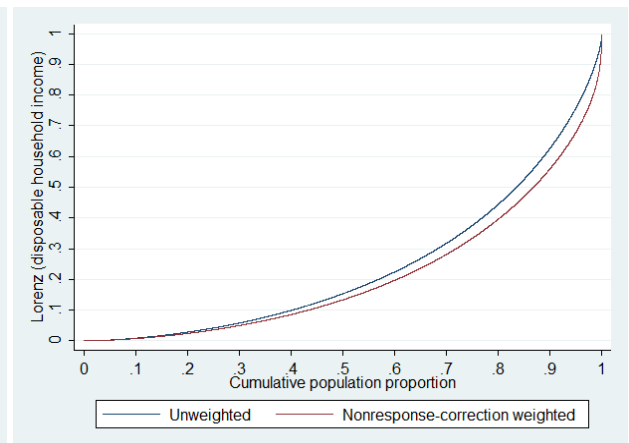
iii. North America



iv. Latin America



v. World (28)<sup>a</sup>



vi. World (34)<sup>b</sup>

<sup>a</sup> 28 countries: *au10, be00, br11, ca10, co10, cz04, ee10, fi10, fr10, de10, gr10, gt06, is10, ie10, il10, it10, jp08, mx10, pe04, ru10, sk10, za10, es10, se05, ch04, uk10, us10* and *uy04*. Excluding low-income countries, for lack of data. Extrapolation to high, upper-middle and lower-middle countries done by augmenting household weights by author-computed subsampling rates of 70.838, 23.429 and 0.509%, respectively.

<sup>b</sup> 34 countries: *at04, kr06, lu10, nl10, pl10* and *ro97* added. For the first 28, subnational regional nonresponse rates are used. For the latter 6, national nonresponse rates are used – surveys are not disaggregated geographically. Excluding low-income countries, for lack of data. Extrapolation to high, upper-middle and lower-middle countries done by augmenting household weights by author-computed subsampling rates of 80.571, 24.186 and 0.509%, respectively.