Issue No. 13 (March 2020)

Inequality Matters

Quarterly updates on inequality research, LIS micro data releases, and other developments at LIS

MESSAGE FROM THE EDITOR

Dear readers,

LIS is grateful to STATBEL who made it possible to close the data gap for Belgium in the LIS Database – now available are five more datasets covering the period from 2004 to 2016. In addition, we are happy to announce the inclusion of a new country to the LIS Database. The new dataset Palestine - PS17 contains information on incomes and expenditures. For more data announcements for Canada (LIS), Czech Republic (LIS) , and Italy (LIS & LWS) see our 'data news' section

We are looking forward to release a new interactive visualisation tool - ***Data Access Research Tool*** **(DART).** DART provides unrestricted access to explore income and wealth inequality around the world. The innovative feature is its richness of inequality measures disaggregated by different social strata.

We recommend you to read through our latest inequality research articles. Arthur B. Kennickell develops illustrative examples applied to the Survey of Consumer Finances to highlight some of the problems in making comparisons of wealth inequality measures when there are specific defects in the measurement of the upper tail of the distribution. By using a novel approach to distributional analysis Marco Ranaldi analyses the dynamics of the capital share of income and how it affects inter-personal income inequality. Carlos Gradín showcases common inequality measures (Gini & Mean Log Deviation) decomposed by different social strata.

Enjoy reading!

Jörg Neugschwender

IN THIS ISSUE

Estimates of Wealth Inequality and Right Tail Coverage: An Illustration of Oversampling in the Survey of Consumer Finances

Arthur B. Kennickell, (Stone Center on Socio-Economic Inequality, Graduate Center, City University of New York (CUNY))

This article is an abridgement of Kennickell, A. (2019). 'The tail that wags: differences in effective right tail coverage and estimates of wealth inequality'. The Journal of Economic Inequality , vol. 17, pp. 443-459, https://doi.org/10.1007/s10888-019-09424-8.

Introduction

In household surveys, it is rare that every sample member is willing to participate. A given segment of an observed distribution may be over- or under-represented in a survey relative to the population, either because of random variation in the sampling process or because of differences across the spectrum of survey sample members in their willingness to participate in the survey. If differences in willingness of sample members to participate are not statistically independent with respect to the analytical dimension(s) of interest, then the measured distribution will differ from what would be estimated from the full sample and many classes of estimates made on such data would be biased.

This article focuses on the sensitivity of survey-based estimates of wealth inequality to the quality of the measurement of the upper tail of the distribution. Such distortions in the measurement of the upper tail of many economic distributions may be especially problematic, because this tail is often highly skewed, as in the case of income or wealth—and in the absence of external bounding information, the distribution is open-ended. In the case of wealth, even within the group of individuals captured in the Forbes list of the 400 wealthiest individuals in the U.S., there is very substantial variation; for example, the minimum wealth to qualify for membership in the list in 2013 was $1.3 billion while the maximum holding among the group was $72 billion.[1] Relative to the $81,200 median U.S. household wealth in 2013, $72 billion is extraordinarily remote. Indeed, the total wealth of the wealthiest few members of the Forbes list possessed more net worth than the least wealthy half of all U.S. households together, as measured in the Survey of Consumer Finances (SCF).

The article presents two of the illustrative examples in Kennickell (2019) to highlight some of the problems in making comparisons of wealth inequality measures when there are specific defects in the measurement of the upper tail of the distribution. For motivation, the article first presents an example based on two sets of time series estimates of wealth shares from the 2013 SCF: one computed from the full SCF sample, including a component that oversamples wealthy households, and the other computed without that additional sample. Next, the 2013 SCF is used to simulate an assortment of distortions in the upper tail of the wealth distribution that might be present in a survey. A final section concludes and outlines a potential research program for improving comparability of wealth measurement across surveys and within waves of a given survey.

Alternative Measures Using Simulated Populations

In Kennickell (2019), we test five experimental samples in order to test the extent of effective coverage of the upper tail of the wealth distribution. These variations span a plausible range of problems in measuring the upper wealth tail in the case of most surveys without strong controls over that population through the sample design or weighting adjustments. The simulations operate by altering the weights for some observations in the 2013 SCF combined sample to impose patterns of "non-observation" on the upper tail of the distribution of household wealth. Table 1 summarizes the five experimental scenarios. For further technical information and illustration see Kennickell (2019).

By design, none of the experimental samples involves oversampling in the upper tail of the wealth distribution, unlike the combined SCF sample. To give an indication of how much difference the oversampling alone makes in the precision of the estimated inequality measures considered, a set of random replicates for the full combined sample was constructed ignoring the structure imposed by the oversampling. Thus by construction, estimates in this case differ in expectation from the combined sample estimates only in the width of the confidence intervals.

Table 1: Specification of experimental samples

Item	Experiment 1	Experiment 2	Experiment 3	Experiment 4	Experiment 5
Population addressed	Wealthiest 1%	Wealthiest 1%	Wealthiest 1%	Wealthiest 1%	Wealthiest 5%
Average weight reduction	50%	50%	90%	100%	90%
Pattern of decay	Flat	Linear	Exponential	Flat	Exponential

Illustrative Example: Survey of Consumer Finances

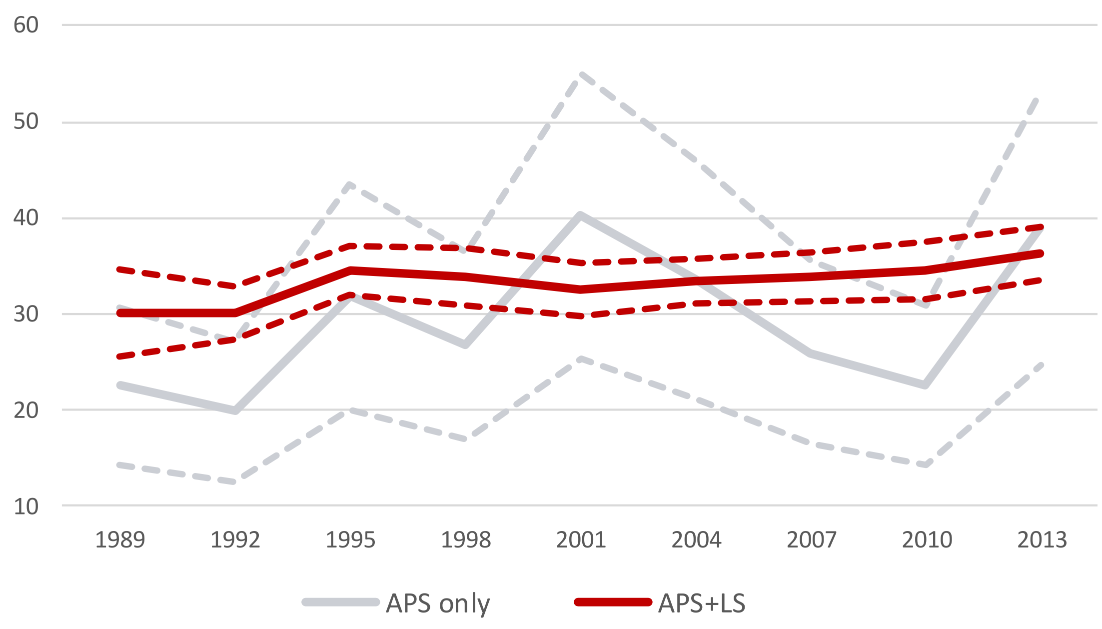
The SCF is particularly helpful for illustrating the effects of differences in the effective coverage of the upper tail of the household wealth distribution on estimates of wealth concentration. The SCF is based on a dual-frame sample design, including both an area-probability sample (APS) and a list sample (LS).[2] Households in the APS are selected with equal probability and stratified to yield a sample with a balanced geographic distribution. The LS is designed specifically to strongly oversample wealthy households. In the final sample, the APS and the LS are combined for analysis through the construction of weights that maximize the strengths of each sample.

The APS provides robust national representation of broadly-distributed characteristics. But an equal-probability sample contains, on average, only 1% of its observations among the wealthiest 1%. For example, an APS of 7,900 sample elements (as in the 2013 SCF) would, on average and assuming all cases are in scope and agree to participate, include only 79 elements to represent the wealthiest 1%. The use of random sampling also implies that the number of such wealthy cases selected would vary; for example, there is a 95% probability that the number of cases in this group would differ from the average figure by no more than about 14.

Hence, the LS is specifically intended to supplement the small expected number of wealthy respondents obtained in the APS, to identify wealth-related nonresponse and to support meaningful adjustments for such differential nonresponse.[3] The sample is based on statistical records derived from individual income tax returns by the Statistics of Income Division of the U.S. Internal Revenue Service. A combination of models is used to project capital income and other characteristics observed in those records to an approximation of wealth, a "wealth index", which is used to stratify the sample.

Because the LS is based ultimately on information from personal income tax returns, the unit of observation—the "tax unit"—does not necessarily align with the household concepts of the APS. In practice, there appears to be only a very small difference at the top of the income or wealth distribution, but differences are substantially greater differences at lower levels. The LS can include only people who filed individual tax returns, and many low-income households do not file returns. In addition, experience indicates that households with relatively lower income or wealth are more likely to have important secondary filers, especially spouses who file their tax returns separately. Thus, the LS focuses very heavily on the top 1% of the distribution of the wealth index and includes only a relatively small measure of other cases to facilitate the integration of the two samples. To highlight the importance of the LS in inequality measurement, Figure 1 shows the estimated wealth share (and its 95 percent confidence interval) of the wealthiest 1% for the surveys conducted between 1989 and 2013, using the combined APS and LS samples, and the same estimates made using only the APS.

Fig. 1: Wealth share of the wealthiest 1% and associated confidence intervals; combined APS and LS and APS alone; SCF, 1989–2013



For each of the samples, Fig. 2 shows estimates of the wealth share for the wealthiest ten, five and 1%, along with the associated confidence intervals. Although the estimates of the share of the wealthiest 10% from the experimental samples show the smallest absolute and proportional bias among the share measures shown, the difference between the value estimated using the combined APS and LS samples and the estimated value for the fourth experiment is about 12 percentage points. As one might expect in light of the results presented earlier in this paper, the range of estimates for the wealthiest 1% is much wider: the estimated value for the fourth experiment is less than half the estimated value for the combined sample. The estimated confidence intervals for the all the experiments except the first give a misleading impression of the reliability of the share estimates. Thus, these results suggest that comparisons of such straightforward estimates of wealth shares for the upper tail across time or across surveys are unlikely to be informative, except where there is minimal nonresponse and a sufficiently large sample, or where there is a strong control on the measurement of that tail.

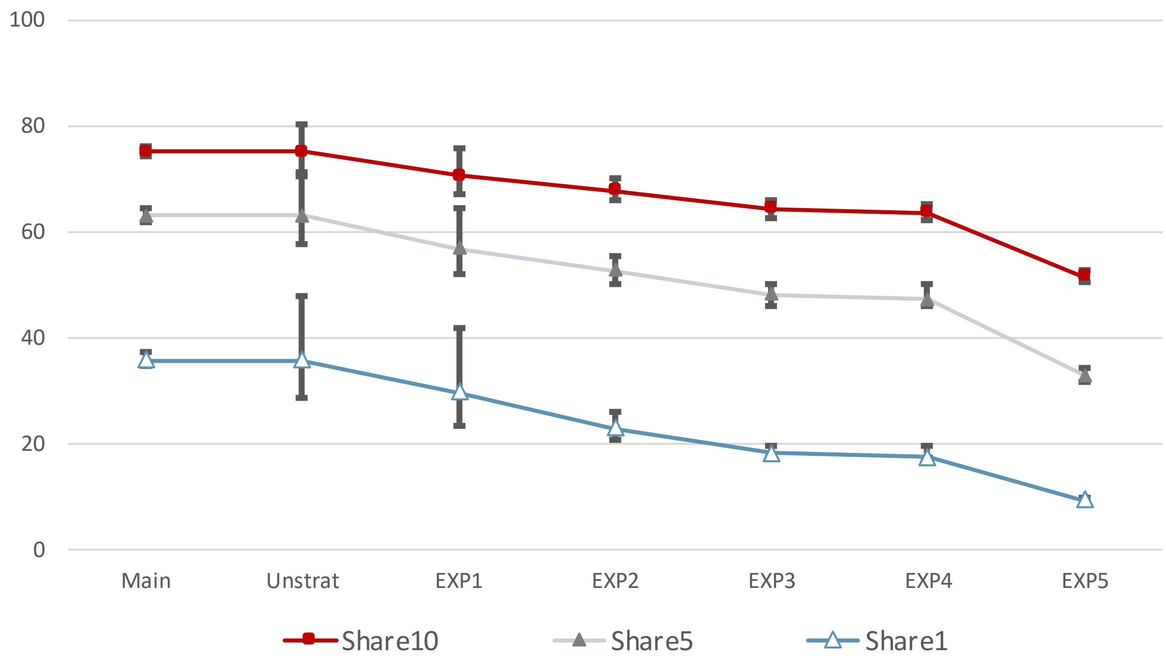
Conclusions and a Way Forward

Kennickell (2019) explored the sensitivity of a variety of indicators of the distribution of wealth. Results for one of those indicators considered here—wealth shares of various percentile groups—strongly indicate that in the absence of effective controls on the measurement of the upper tail of the wealth distribution, great caution should be the rule in the interpretation of commonly used measures of wealth distribution from a given survey, comparison of such measures across the waves of the survey, and perhaps even more strongly, comparison across independently designed and managed surveys. Among the inequality measures considered in Kennickell (2019), only the ratio the 95th to the 25th percentile of wealth and the ratio of the 90th to the 25th percentile of wealth appear to be reasonably reliably informative when estimated from surveys with biases in the measurement of the top of the distribution.

Without access to reliable data with information on characteristics closely related to wealth, such as income tax data, it is very difficult at best to develop a sample that provides sufficiently effective coverage of the upper tail of the wealth distribution to yield purely survey-based and reasonably stable estimates of wealth concentration. Although the SCF appears to do very well in addressing the relevant measurement concerns, even it shows some signs of deviation at the highest levels of the wealth distribution, and the sampling error in estimates disproportionately influenced by that tail, though not enormous, is also not negligible (see Kennickell, 2017). There is room for improvement in all wealth surveys.

Improvement can be made by using administrative data for sampling or for weighting adjustment, as in the SCF, the Encuesta Financiera de las Familias and the Enquête Patrimoine, or in the case of countries with reliable wealth register data, by replacing some or all of the survey measures. For example, Saez and Zucman (2016) take a related approach of estimating wealth entirely from administrative data on income and related measures. But these approaches are not possible for every survey of wealth. In other cases, some degree of modeling may be helpful in improving the measurement of the upper tail. Using only the observed data in the Austrian implementation of the Household Finance and Consumption Survey for 2010, Eckerstorfer et al. (2016) estimated a Pareto distribution for the upper tail of using information from a range of data from about the 70th to 99th percentiles of the observed data and they “recover” substantial additional wealth. When administrative data are not available for direct use, it may still be possible to obtain estimates of distributional curvature for wealth or a proxy from such data and apply those estimates to adjust the survey weights. Other external estimates, such as “rich lists” may also be used to adjust the weighting of observed data or to “impute” unobserved wealth. Vermeulen (2016, 2018) used Forbes and similar data on wealthy individuals in conjunction with survey data for several countries to estimate Pareto distributions to

Fig. 2: Share of total wealth held by the wealthiest 10%, 5%, and 1%; combined APS and LS, unstratified combined sample and experiments 1–5; SCF, 2013



describe the augmented data. Bach et al. perform a similar exercise with a set of surveys that perform oversampling and find incorporating such external information make only a small difference in those cases. Even national accounting data on wealth may be useful in developing improved estimates, for example, by estimating Pareto distributions conditional on the total implied wealth equalling the aggregate value. In my view, finding the most robust approach that is applicable to many surveys should be the highest priority for research in this direction.

1 *See https://www.forbes.com/forbes-400/list/9/#version:static (accessed May 2019) for the most recent such information.*

2 *See Kennickell (2017) for a detailed discussion of the SCF samples and for references to other technical research and information about the survey.*

3 *The survey explicitly omits individuals who appear on the Forbes list of the 400 wealthiest people in the US.*

References

Bach, S., Thiemann, A., and Zucco, A. (2015), The Top Tail of the Wealth Distribution in Germany, France, Spain, and Greece. DIW Discussion Papers, No. 1502.

Eckerstorfer, P., Halak, J., Kapeller, J., Schütz, B., Springholz, F., Windauer, R. (2016), Correction for the missing rich: an approach to wealth survey data. *Rev. Income Wealth*. 62, 605–627.

Kennickell, A.B. (2019), The tail that wags: differences in effective right tail coverage and estimates of wealth inequality, *Journal of Economic Inequality*, 17: 443. https://doi.org/10.1007/s10888-019-09424-8

Kennickell, A.B. (2017), Lining up: survey and administrative data estimates of wealth concentration. *Statistical Journal of the International Association for Official Statistics*, 33, 59–79.

Saez, E., Zucman, G. (2016), Wealth inequality in the United States since 1913: evidence from capitalized income tax data. *Q. J. Econ*. 131, 519–578.

Vermeulen, P. (2016), Estimating the top tail of the wealth distribution. *Am. Econ. Rev. Pap. Proc*. 106(5), 646–650.

Vermeulen, P. (2018), How fat is the top tail of the wealth distribution. *Review of Income and Wealth*. 64, 357–387.

Income Composition Inequality: A Novel Approach to Distributional Analysis

Marco Ranaldi , (Stone Center on Socio-Economic Inequality, Graduate Center, City University of New York (CUNY))

One of the major findings of Piketty's *Capital in the XXI Century* (Piketty, 2014) is the rise in the capital share of income in many advanced economies. From a technical perspective, the capital share of income is a simple ratio of capital to total income in a given economy. It is thus a number that ranges between zero and one: it is equal to one when an economy's national income is entirely composed of profits and to zero when it is entirely composed of wages. From a conceptual perspective, both the level and trend of a country's capital share of income provide us with crucial information on the same country's economic structure. The level can indeed be regarded as a measure of the bargaining power different social groups with conflicting interests (such as "workers" and "capitalists") have. At the same time, its longer term dynamics are driven by important global phenomena such as technological progress and financialization.

In this article I will address a simple question: how do the dynamics of the capital share of income affect inter-personal income inequality? Or in other words: how does the distribution of capital and labor in national income, also called *functional income distribution*, relate to the distribution of income among individuals?

Many scholars have studied this issue, especially during the last few decades. In his book, Thomas Piketty assumes the variation of these two variables, - variation in the share of capital income and in income inequality, - to be positively associated. This assumption is motivated by the fact that capital incomes tend to be mainly concentrated in the hands of those at the top of the income ladder.

From an econometric perspective, Bengtsson and Waldenström (2018) find evidence of a strong positive link between the functional and personal distribution of income, a link which has grown stronger over the past century. In contrast, Francese and Mulas-Granados (2015), based on an analysis that covers 93 countries between 1970 and 2013, find that the distribution of income between labor and capital has not been a major factor in explaining changes in income inequality.

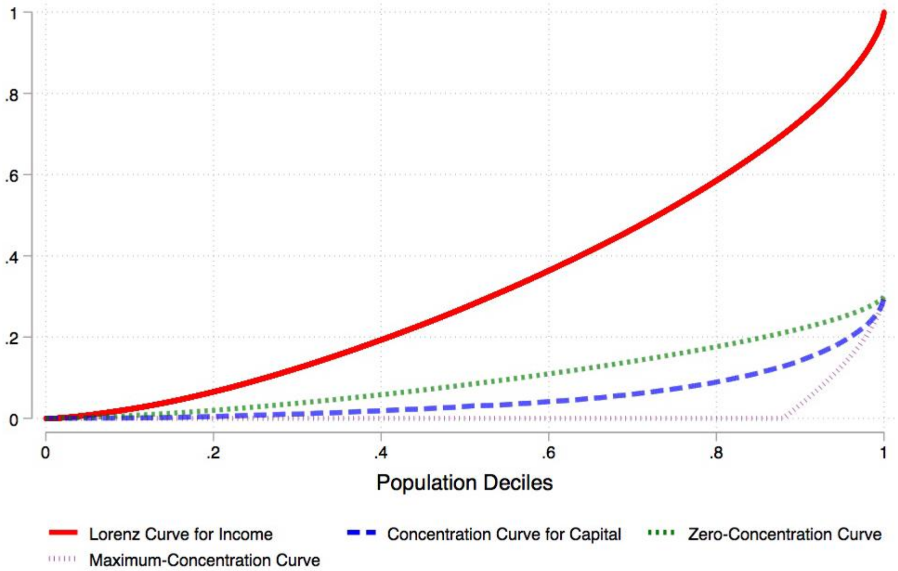
The conflicting nature of these results highlights that this relationship is not "as simple and unambiguous as it may seems", to use Milanovic's own words (Milanovic, 2017). In a recent article I argue that, in order to study this relationship, we need to introduce a novel inequality concept that I termed *income composition inequality*. Furthermore, in the same article I constructed a new statistic called the *Income-Factor Concentration* (IFC) index, which serves for its measurement (Ranaldi, 2020).

In this piece, I will introduce both the novel concept of income composition inequality and the IFC index. Furthermore, I will illustrate how this approach can be useful to carry out a novel political-economic analysis of contemporary capitalist economies. To this end, an application of this method to the Italian context will also be presented (Iacono and Ranaldi, 2018).

I will start by introducing the concept of income composition inequality with a simple example. Let us suppose a stylized society consisting of two individuals only, Adrien and Beatrice. Suppose that Adrien and Beatrice have the same level of income, which is of $1000 per month. The total income inequality in this society is therefore zero. But what can we say about the composition of Adrien and Beatrice's incomes? Suppose, for instance, that Adrien's monthly income consists entirely of profits since he receives the monthly rent from a house in Paris that he owns. Beatrice's income instead consists entirely of her salary as a teacher in the high school. To sum up, Adrien and Beatrice have the same level of income but a completely different source of income. Suppose now that both Adrien and Beatrice earn 50% of their incomes from labor and 50% from return on capital. The composition of their incomes is now exactly the same. To sum up, in the first scenario the composition of the two income sources was *unequally distributed* across Adrien and Beatrice, whilst in the second scenario it was *equally distributed* between them.

Bearing in mind the previous examples, we can now easily grasp the more formal definition of income composition inequality, as follows: if we decompose total income into two components, such as capital and labor income, then income composition inequality is the extent to which income composition is distributed unevenly across the income distribution. Inequality in income composition is *maximal* when individuals at the top and at the bottom of the income distribution separately earn two different types of income, such as capital and labor income. It is minimal when each individual has the same composition of capital and labor income.

Figure 1. Concentration Curves for Income Source - Italy, 1989



Note: The concentration curve for capital (blue line), the zero-concentration curve (green line), the Lorenz curve for income (red line) and the maximum-concentration curve (purple line) for Italy in 1989 are presented using data from the 1989 Survey on Household Income and Wealth (SHIW) carried out by the Bank of Italy. Capital income is defined as the sum of property income and the capital component of net self-employment income. Labor income is instead defined as the sum of payroll income, pensions, net transfers and the labor component of mixed income. Both the capital and labor components of self-employment income are imputed following Glyn (2011). The IFC index is defined as the ratio between the area between the zero-concentration curve and the concentration curve for capital income, and the area between the zero- and the maximum-concentration curves. The higher the numerator, the more capital income is concentrated at the top of the total income distribution and the stronger the link between the functional and personal distribution of income, as an increase in the overall capital share would increase the income of those at the top of the total income distribution, and hence increase personal inequality.

From a more technical point of view, we can say that income composition inequality links the functional and personal distribution of income. This entails a straightforward observation: if the rich earn all the capital income in an economy, then an increase in the capital share of income makes the rich richer. Consider once again the stylized society previously described, in which Adrien was the only capital income earner. If the capital share of income grew by a certain amount (holding its distribution across individuals constant), then Adrien's income would increase, and inequality across total income subsequently also.

From a more political-economic perspective, income composition inequality can provide insights into the degree of capitalism of a social system (Milanovic, 2017). Indeed, under high income composition inequality a society can be seen as an example of classical capitalism, where the rich earn the capital income and the poor the labor income. In contrast, under low income composition inequality a society can be regarded as exemplifying new capitalism, where both rich and poor have the same composition of capital and labor income.

Once agreed on the definition of income composition inequality, the next step concerns its measurement. Although there may be many different ways to measure income composition inequality, I favor one method in particular. To this end, let me first recall the notion of concentration curves, an important tool for distributional analysis, first introduced in the 1970s (see, for instance, Kakwani, 1977).

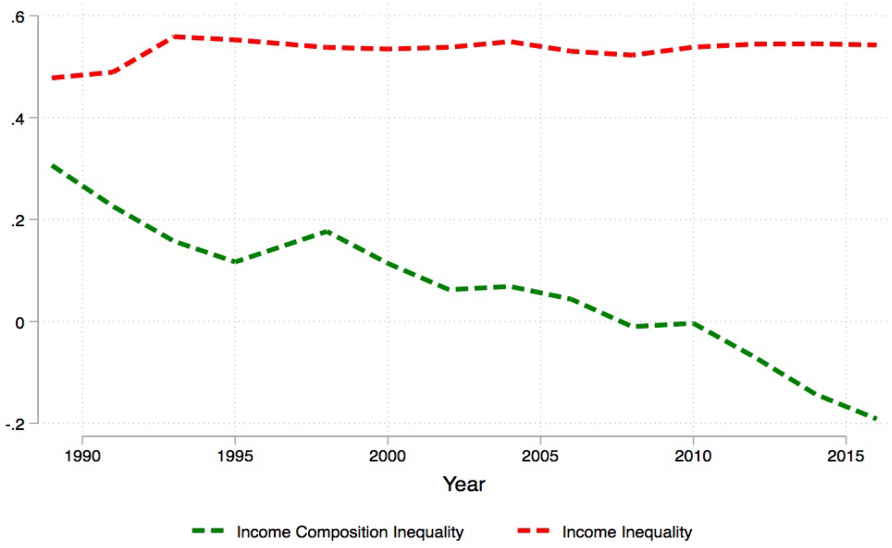
The concentration curve is a curve that cumulates the relative share of a given variable (such as capital income) across the population with individuals ranked according to another variable (such as total income). Suppose that Adrien has an income of $100, which is composed of $10 of capital income and $90 of labor income, whereas Beatrice has an income of $1000, which is composed of $900 of capital income and $100 of labor income.

To construct the concentration curve for capital income with individuals ranked according to their total income one firstly needs to rank individuals from the poorest (Adrien) to the richest (Beatrice). Then, one has to calculate their relative shares of capital income. The latter equals 10/910 for Adrien (Adrien's capital income over total capital income in the economy), and 900/910 for Beatrice. The concentration curve for capital income *cumulates* these two numbers across the income distribution. The final concentration curve will, hence, be characterized by the three distinct pairs: (0,0), (1/2, 10/910) and (1,1). If one now multiplies the concentration curve just constructed by the total capital income share in the economy (which is 910/1100 in this case) one obtains what I would call the *concentration curve for capital income*. Figure 1 shows the concentration curve for capital income (blue line) for Italy in 1989.

As can be seen, the concentration curve for capital income is almost flat up until the 4th decile. This means that the bottom 40% of the total income distribution earns almost no capital income, which is instead concentrated at the very top of the total income distribution, as the concentration curve starts growing very rapidly from the 90th percentile onwards. In correspondence with the 100th percentile the concentration curve reaches the point 0.3, which is the overall capital share of income captured by the survey.

In sum, the more "convex" the concentration curve is, the more concentrated is the capital income at the top of the total income distribution. Similarly, the more "concave" the concentration curve is, the more concentrated is the capital income at the bottom of the total income distribution. Although the concentration curves provide a clear, graphical indication of the *direction of concentration* (either at the top or at the bottom), we still do not know how to measure the extent of such a concentration.

Figure 2. Income Composition Inequality - Italy, 1989:2016



Note: The two series of income composition inequality, as measured by the IFC index, and of income inequality, as measured by the Gini coefficient, for Italy between 1989 and 2016 are constructed using the SHIW data. Capital income is considered as the sum of property income and the capital component of net self-employment income and labor income as the sum of payroll income and the labor component of net self-employment income. Total income is the sum of capital and labor income as previously defined.

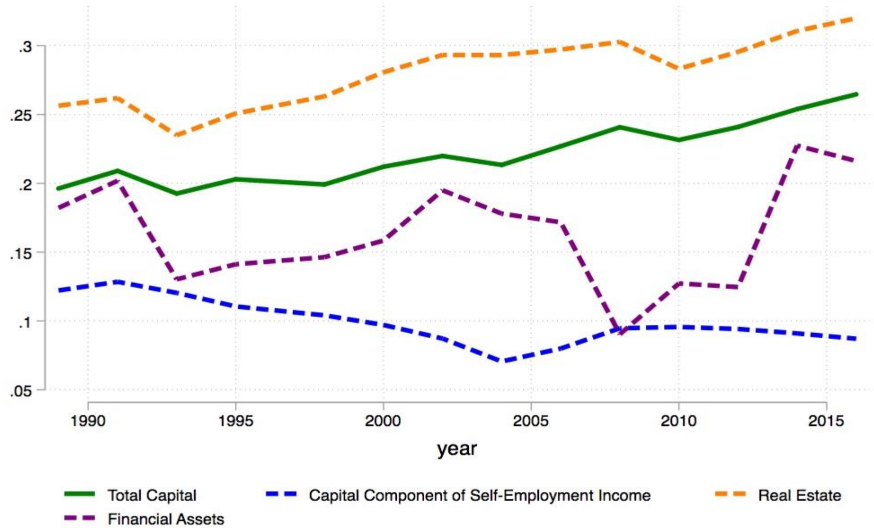
To answer this question, we need to introduce two benchmark conditions: the benchmark of zero- and of maximum-concentration of income sources, together with their related concentration curves. We say that there is zero concentration of income sources when the composition of capital and labor is the same for all individuals (i.e. Adrien and Beatrice have the same shares of capital and labor income in their total income). The curve that describes this distribution of income sources is the Lorenz curve for total income, multiplied by the capital share (green line in Figure 1). The zero-concentration curve is the benchmark of zero inequality in income composition. This curve should be seen as the equivalent of the egalitarian line used to calculate the Gini coefficient. However, differently from the egalitarian line, the zero-concentration curve changes as the Lorenz curve and the functional distribution of income change. In contrast, we have maximum concentration of income sources when the bottom *p* percent of the total income distribution earn from one single source, and the top *1-p* percent of the income distribution earn from the other source.[1] The maximum-concentration curve is hence flat up to a given threshold, and then cumulates all the capital income in the hands of the remaining fraction of the population. It is important to notice that the maximum-concentration curve can potentially describe a distribution of income sources in which capital income is concentrated at the *bottom* of the total income distribution and labor income at the top (as unlikely it may seem). The maximum-concentration curve is, hence, the benchmark of maximal inequality in income composition.

Now that the concentration curve for capital income and the benchmarks for minimal and maximal inequality in income composition have been introduced, we can define the first measure of income composition inequality. If we denote by ***A*** the area between the zero-concentration curve and the concentration curve for capital, and by ***B*** the area between the zero and the maximum concentration curves, I define the income-factor concentration (IFC) index as the ratio between ***A*** and ***B***.

The IFC index is a number ranging between 1 and -1. It is equal to 1 when capital income is concentrated at the top (and labor income at the bottom), whereas it is equal to -1 when capital income is concentrated at the bottom (and labor income at the top). Finally, it is equal to 0 when the composition of capital and labor is the same across the total income distribution. From a mathematical standpoint, it can be easily shown that the derivative of the Gini coefficient to changes in the functional income distribution is equal to the numerator of this statistic (i.e. ***A***). This makes the IFC index the first measure of the link between the functional and personal income distribution. By applying the IFC index to the case of Italy between 1989 and 2016, Roberto Iacono and I find that income composition inequality has steadily decreased in Italy over the period considered (see Figure 2).

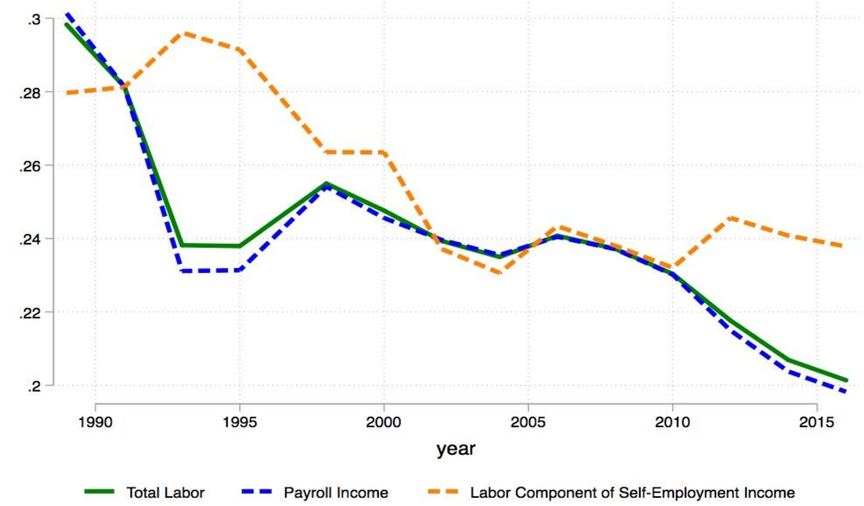
The implications of this result, which is robust to different definitions of capital and labor income, are twofold. *First*, fluctuations in the total factor shares of income are having an increasingly weaker impact on income inequality in Italy. This latter aspect is a direct consequence of the decrease in the IFC index: the lower the IFC index, the less capital income is concentrated at the top (of the total income distribution), the weaker the link between the functional and personal distribution of income. Under a weak link between the functional and personal distribution of income, an increase in the capital share has almost no effect on inter-personal inequality. Second, Italy is moving towards practising a new type of capitalism, in which individuals earn from multiple sources of income and the dichotomy between “rich capitalist” and “poor workers” does not seem to hold anymore. This fact underlines some major changes in the structure of the Italian economy. As shown in Figures 3 and 4, Italy has become more of a home-owning society (as real estate is now more equally distributed across the total income distribution). At the same time, both self-employment and payroll incomes have moved from the bottom to the top of the total income distribution, suggesting a change in the composition of the top income earners. These components are the ones responsible for the fall in income composition inequality over the last three decades in Italy.

Figure 3. Area of the Concentration Curves - Capital Income Components



Note: The series of the area of the concentration curve for capital (green line), together with the area of the concentration curves for the capital component of self-employment income (blue line), real estate (orange line) and financial assets (purple line), for Italy between 1989 and 2016 constructed using the SHIW data. When the area of the concentration curve for total capital (for instance) increases, then total capital is becoming more equally distributed across the income distribution. The reverse situation happens when the areas decreases.

Figure 4. Area of the Concentration Curves - Labor Income Components



Note: The series of the area of the concentration curve for labor (green line), together with the area of the concentration curves for payroll income (blue line) and the labor component of self-employment income (orange line), for Italy between 1989 and 2016 constructed using the SHIW data. When the area of the concentration curve for total labor income (for instance) increases, then total labor income is becoming more equally distributed across the income distribution. The reverse situation happens when the areas decreases.

To sum up, in this article I introduced a novel inequality concept that I term income composition inequality. I argued that income composition inequality links functional and personal distribution of income and allows for a novel political-economic analysis of contemporary capitalist economies à la Milanovic. Then, I introduced the Income-Factor Concentration Index, a summary statistic constructed to measure the degree of income composition inequality. Finally, I showed the evolution of income composition inequality in Italy over the last three decades and how it has steadily decreased over this period.

It is of the utmost importance to highlight that, while I use capital and labor as income components in this paper, the study of income composition inequality can be applied to analyze the joint distribution of any pairs of income (or wealth) components, such as net income and taxes, savings and consumption, or financial and non-financial assets, among others. The flexible nature of income composition inequality therefore paves the way for future research on the topic.

1 *We refer the reader to the scientific article for a better understanding of the precise choice concerning the fraction* p.

References

Bengtsson, E., Waldenström, D. (2018), Capital Shares and Income Inequality: Evidence from the Long Run. *Journal of Economic History*, Vol. 78, Issue 3, pp.712-743.

Francese, M., Mulas-Granados, C. (2015), Functional Income Distribution and Its Role in Explaining Inequality. IMF Working Paper, WP/15/244.

Iacono, R., Ranaldi, M. (2018), Sources of Inequality in Italy. Working Paper 479, ECINEQ, Society for the Study of Economic Inequality.

Kakwani, N. (1977), Application of Lorenz Curves in Economic Analysis. *Econometrica*, Vol. 45, No. 3, pp. 719-728.

Milanovic, B. (2017), I*ncreasing Capital Income Share and its Effect on Personal Income Inequality*. In H. Boushey *(ed.), After Piketty. The Agenda for Economics and Inequality*, J. Bradford DeLong, Marshall Steinbaum. Ch. 10.

Piketty, T. (2014), *The Capital in the XXI Century*. Harvard University Press.

Ranaldi, M. (2020), Income Composition Inequality. Stone Center Working Paper Series, No. 07.

Why Is Inequality in South Africa Higher than in Germany?

Carlos Gradín, (UNU-WIDER, Helsinki)

The understanding of inequality requires the analysis of changes in income distributions across countries and over time as well as the identification of its drivers. To achieve this we use different statistical tools to identify the distributional patterns and summarize the results using inequality indices. The use of decomposition analysis has been particularly popular in the field for the purpose of identifying strong statistical associations, even if the identification of causality is complex in this context. There are at least three different types of common inequality decompositions: by population groups, by income sources, and regression-based decompositions.

In this article, I will use a practical case and combine these different approaches that are usually investigated independently, based on the methods proposed in Gradín (2018, 2019), where they are explained in more detail. I will use data from the Luxembourg Income Study (LIS). Inequality will be measured for disposable income per equivalent adult.[1] I will focus on two indices that have the right properties, the Mean Log Deviation (MLD) and the Gini index.[2] I will compare inequality in one highly unequal country (South Africa, 2012) and in one low-inequality country (Germany, 2013). Indeed, South Africa is the country with the highest level of income inequality in LIS, while Germany has one of the lowest levels among the largest economies. Inequality in South Africa is higher than in Germany as shown in Figure 1, with the gap being 0.316 (Gini) and 0.560 (MLD). The aim is to shed some light on the role in driving the cross-country gap in income inequality played by i) a head's attained education[3], probably the most relevant socioeconomic characteristic, and ii) two incomes sources (net of direct taxes): 'market income', broadly defined here as incomes derived mainly from labor, old-age pensions and capital, and 'public social benefits' (other than pensions).

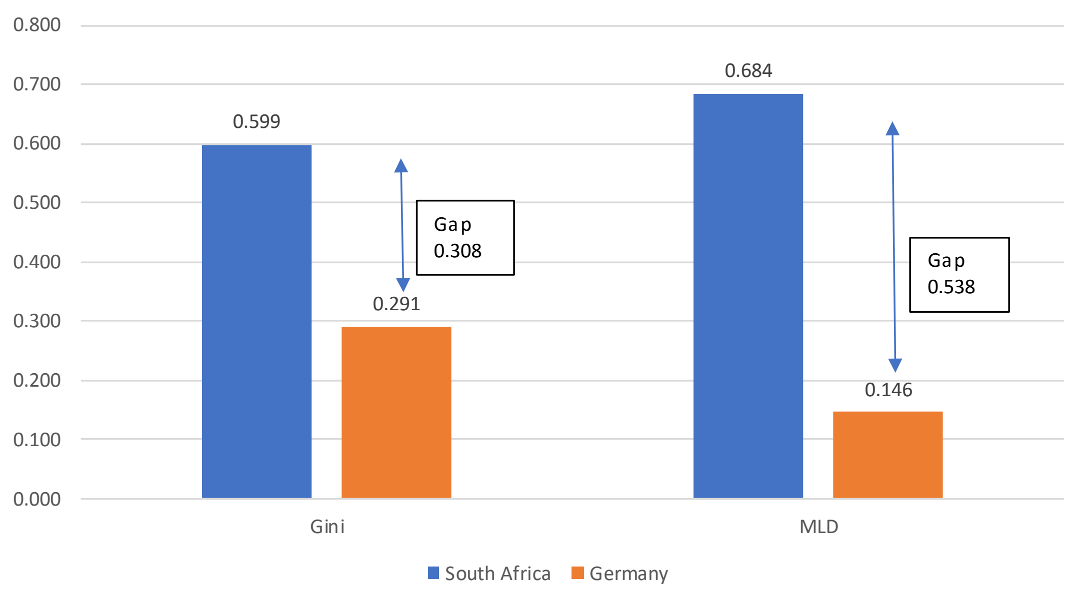
Inequality decomposition by population groups

The first decomposition type implies breaking the population of each country into groups based on one socioeconomic characteristic, such as race, region, education, etc. The most common approach implies decomposing total inequality into the contribution of inequalities between groups and within groups so as to identify how strongly inequality is associated with each particular characteristic (higher share explained by the between-group component), using an index such as the Mean Log Deviation, in which the overall level of inequality is equivalent to the sum of these two components. Inequality between groups is obtained as the level of inequality that remains after equalizing the incomes within groups in each country (assigning everyone the mean of their group). Inequality within groups is the level of inequality remaining after re-scaling individual incomes so that all groups have the same mean income in each country, which is equivalent to the sum of group inequalities, with each group weighted by its population size. [4]

It is when following the decomposition according to population groups that we know that inequality among all citizens of the world is mainly determined by differences in the average income of the country in which we live (inequality between countries), even if the within-country component is becoming more relevant over time. We also know that the urban-rural gap played an important role in the increase of inequality in China after the economic reforms, or that race, caste or ethnicity are fundamental to understand inequality in many countries, with South Africa and India standing out in this respect.

In my example, MLD bar in Figure 2, it turns out that the inequality gap between South Africa and Germany, measured by the MLD, is in part due to the striking mean income differences among educational groups in South Africa. On average the most educated group receives 15 times

Figure 1. Inequality in South Africa and Germany



Source: own calculations based on Luxembourg Income Study (LIS) Database.

the income of the least educated, compared with 2.4 times in Germany. Despite being large, however, between-group inequalities still explain only 39 per cent of the total gap. The main component, about 61 per cent, is due to cross-country differences in within-group inequality. That is, differences that occur among people with the same head's educational level (regardless what their group mean income is). Though in South Africa your level of education determines to a larger extent where you are in the income distribution (35 per cent of inequality is between educational groups, compared to 18 per cent in Germany), there is also an even larger variability within each educational group.

Inequality decomposition by income sources

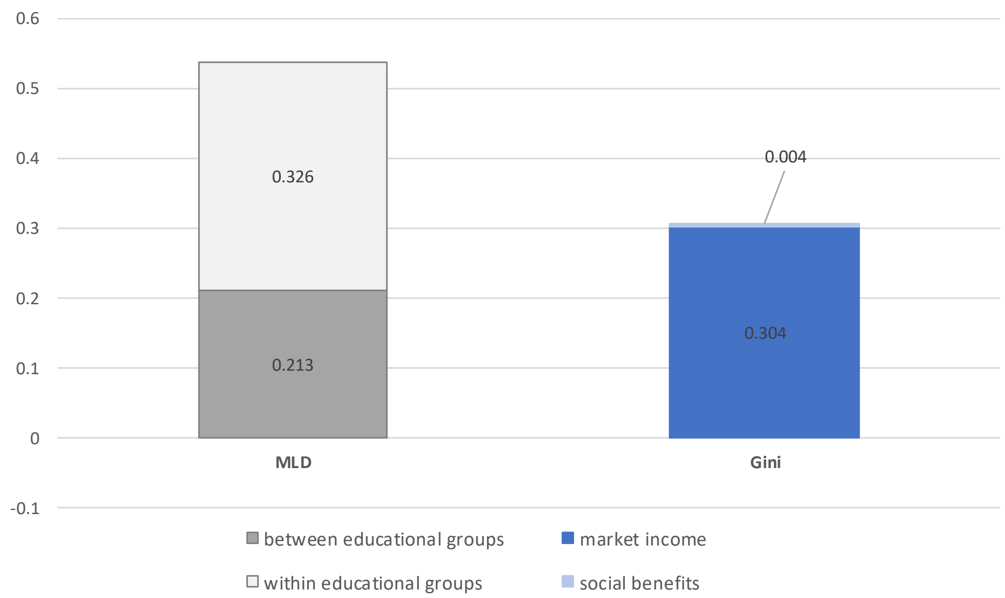
A second decomposition type is the decomposition of total inequality into the contribution of income sources (e.g. earnings, social benefits, taxes …). In its simplest and most popular version, the contribution of an income source can be measured as the change in inequality after adding that source to the other incomes (Musgrave and Thin 1948, and subsequent literature).[5] Defined in this way, a source can be progressive or regressive depending on whether it contributes to making inequality lower or higher. The analysis of income sources has allowed to find out that income inequality in most countries is mainly generated in the labor market, while it is partially offset by the effect of taxes and social benefits, but with great variability across countries based on factors such as their economic structure, inequalities in human and physical capital, labor market institutions, exposition to trade or technological change, and how redistributive the tax-benefit system is, among other things.

Therefore, the second question addressed in my example, is whether the observed country gap is the result of a more disequalizing market or of a weaker welfare state in South Africa compared with Germany. For that, I compare inequality, measured by the Gini index, before and after adding public social benefits to market incomes. Inequality decreases by a similar amount in both countries: from 0.341 to 0.291 (-0.050) in Germany, and from 0.644 to 0.599 in South Africa (-0.045). The initial gap of 0.304 Gini points only slightly increases to 0.308 (Gini bar in Figure 2). It turns out that the much higher original level of market income inequality, not the smaller reduction resulting from social benefits, is the reason why inequality is higher in South Africa.

Finally, the regression-based decomposition approach allows us to decompose the differential in inequality between two distributions into a composition effect (difference driven by the divergent distribution of characteristics) and an income structure or distribution effect (difference driven by how population groups are differently distributed across incomes). For example, it is possible that a household with a given educational level obtains the same relative income in both countries and inequality is higher in South Africa simply because there are more people with lower education (and therefore lower relative earnings). In that scenario, we could say that the inequality gap is driven by a composition effect (by educational groups). Alternatively, the two countries might have the same share of the population by educational level but differ in the income distribution of each educational group. In this case, the distribution effect would be the reason for the inequality gap that indicates to what extent educational groups are associated with more inequality in one country, i.e. some groups tend to be at the bottom and/or top of the distribution.[6] This should not be confused with the within-group inequality discussed above because groups may differ in income variability but also in their average income.

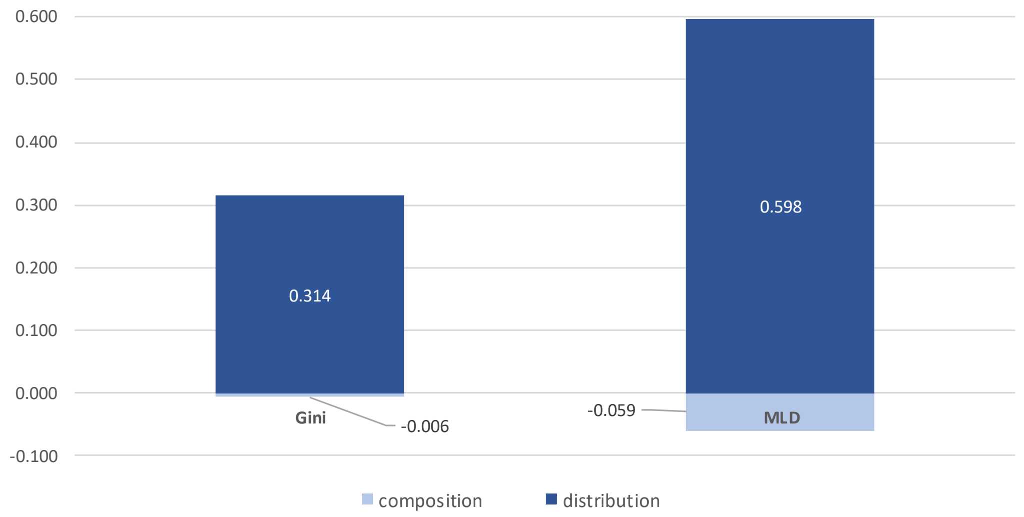
I obtain the composition effect estimating how much of the inequality gap disappears after equalizing the distribution of educational groups in both countries (they have the same proportion of people in households with higher education, for example). I do this by constructing a counterfactual (hypothetical) distribution in which I give households in Germany the educational distribution in South Africa and repeat the exercise swapping countries (by reweighting the corresponding samples). The composition effect is the average level of inequality that has been reduced in both cases. The distribution effect indicates the inequality gap that remains when both countries are compared in terms of the population shares by educational levels.

Figure 2. Decomposing the inequality gap between South Africa and Germany by income source (Gini) and population groups (MLD)



Source: own calculations based on Luxembourg Income Study (LIS) Database.

Figure 3. Decomposing the inequality gap between South Africa and Germany into composition and distribution effects



Source: own calculations based on Luxembourg Income Study (LIS) Database.

There is no doubt that the educational structure of the population in both countries is quite different, with more South Africans living in households where the head has not achieved lower secondary education (35 per cent of the population, compared to only 2 per cent in Germany) and fewer in which the head has a bachelor degree or higher education (5 per cent versus 26 per cent). Despite these striking differences, once they are removed, the inequality gap would still be similar or even higher (by 0.006 with Gini, 0.059 with MLD) as shown by the composition effects in Figure 3. Therefore, the inequality gap between the two countries cannot be described as the result of a composition effect. The entire cross-country inequality gap stems from a distribution effect, that is, from the stronger association of a head's education level with the position of households along the income scale in South Africa.

Combining different approaches

It is interesting to note that despite the strong connection and potential complementarities in the study of inequality among these three decomposition approaches, they have been investigated and used almost independently from each other. So far, I have shown three basic results: the inequality gap is i) mainly driven by inequality within educational groups (although between-group inequalities are also notable), ii) is generated before public social benefits are accounted for, and iii) is generated by the different income distribution of educational groups, not by their population shares being different in both countries.

My main point here is that these narratives can be connected (see Table 1). For example, if it is the income distribution of educational groups, not their size, that matters, these differences can arise because either

Table 1. Detailed decomposition of the cross-country gap

	Composition			Distribution			Total		
	market income	social benefits	total	market income	social benefits	total	market income	social benefits	total
Gini	*0.035*	*-0.042*	*-0.006*	*0.268*	*0.046*	*0.314*	*0.304*	*0.004*	*0.301*
less than upper secondary	0.196	-0.047	0.149	0.028	0.024	0.053	0.224	-0.023	0.350
lower secondary	0.002	-0.004	-0.002	0.036	0.005	0.042	0.038	0.002	0.037
upper secondary	-0.073	0.003	-0.070	0.140	0.005	0.146	0.067	0.009	0.006
post-secondary non-tertiary	-0.010	0.003	-0.007	0.026	0.004	0.030	0.015	0.007	0.015
bachelor or equivalent	-0.043	0.002	-0.040	0.019	0.004	0.023	-0.024	0.007	-0.057
master or equivalent	-0.036	0.000	-0.036	0.019	0.003	0.022	-0.017	0.003	-0.051
	between-group	**within-group**	**total**	**between-group**	**within-group**	**total**	**between-group**	**within-group**	**total**
MLD	*-0.044*	*-0.015*	*-0.059*	*0.257*	*0.341*	*0.598*	*0.213*	*0.326*	*0.479*
less than upper secondary	0.045	0.075	0.120	0.047	0.038	0.085	0.092	0.113	0.325
lower secondary	-0.024	0.006	-0.018	0.031	0.042	0.073	0.007	0.048	0.037
upper secondary	-0.058	-0.040	-0.098	0.055	0.204	0.259	-0.003	0.164	0.063
post-secondary non-tertiary	0.017	-0.009	0.008	0.013	0.035	0.048	0.030	0.025	0.064
bachelor or equivalent	0.007	-0.031	-0.024	0.049	0.016	0.065	0.056	-0.015	0.017
master or equivalent	-0.030	-0.016	-0.046	0.060	0.007	0.067	0.030	-0.010	-0.026

Note: Detailed composition effects include the impact of non-linearity in the relationship between group's contribution on inequality and population shares.

Source: own calculations based on Luxembourg Income Study (LIS) Database.

these groups have different average incomes (between-group inequality) or different intragroup variability (within-group inequality). That is, we can reassess the role of between-group and within-group inequalities in the scenario in which both countries have the same relative group sizes, to find out that these proportions are 43 and 57 per cent (Figure 2). Thus, removing the composition effect does not significantly alter the fact that it is within-group variability that explains higher inequality in South Africa. Similarly, one can ask whether the different income distribution of educational groups, when both countries are compared with the same educational distribution, is produced by market income or by social transfers, and then find out that after removing the composition effect it turns out that the weaker social transfers in South Africa are more relevant than suggested above in explaining the inequality gap (0.046 Gini points) but that this was partially hidden by the associated composition effect (the disproportionally larger share of South Africans with low education who benefit more from these transfers) (Figure 3). One therefore definitely needs to analyze why the labor market in South Africa generates so much inequality to understand higher inequality in South Africa (see for example, Murray et al., 2020).

In addition, Table 1 provides more details on the educational groups through which these effects are channeled because it is possible to identify the contribution of each group to overall inequality, or to any of its components (inequality by income source, between-group and within-group inequality, composition and distribution effects, or combinations of them).[7] One can see that the largest contribution to the inequality gap is associated with households with the lowest educational level but that this is mainly the result of a composition effect (this group is disproportionally larger in South Africa). When it comes to the distribution effect, however, it is the distinct income distribution of the upper-secondary group in both countries that contributes most to the gap through market income inequality and within-group inequality.

1 *The squared root of the household size.*

2 *While MLD is exactly decomposable into the sum of between-group and within-group inequalities, Gini can accommodate zero (or negative) incomes that might arise if we remove an income source from disposable income.*

3 *I have used the education attained by the spouse or the maximum level of education in the household in the few cases in which the education of the head was not available.*

4 *Other indices of the Generalized Entropy family are also additively decomposable as defined in Shorrocks (1984), but the interpretation of the two terms is more problematic. The Gini and Atkinson indices can also be decomposed but in different ways.*

5 *But one can also consider various sequences in which sources can be added, or even consider the average across all possible sequences (the Shapley decomposition). An alternative approach exploits the fact that some inequality indices are just a weighted sum of all incomes ('natural decompositions').*

6 *This exercise can be done controlling for other explanatory variables and there are different methods available based on the natural decomposition of the variance, re-weighting and/or the Recentered Influence Function (RIF). These are extensions of the Blinder-Oaxaca decomposition of average outcome differences between two distributions.*

7 *This is done based on the statistical concept of RIF (i.e. the impact on any index of marginally increasing the proportion of population at each income) proposed by Firpo et al. (2009).*

References

Firpo, S.; Fortin, N.M.; Lemieux, T. (2009). "Unconditional quantile regressions". *Econometrica*, 77, 953–973.

Gradín, C. (2018). "Quantifying the contribution of a subpopulation to inequality: An application to Mozambique". WIDER Working Paper 60/2018. *Journal of Economic Inequality* (forthcoming).

Gradín, C. (2019). "Inequality by population groups and income sources: Accounting for inequality changes in Spain during the recession". WIDER Working Paper 73/2019.

Leibbrandt, M.; Green, P.; Ranchhod, V. (2020). "South Africa: The top-end, labour markets, fiscal redistribution and the persistence of very high inequality". In C. Gradín, M. Leibbrandt, and F. Tarp (eds.) *Inequality in the development world*. Oxford University Press, forthcoming.

Musgrave, R.A. and Thin, T. (1948). "Income Tax Progression 1929–48". *Journal of Political Economy*, 56: 498–514.

Shorrocks, A.F. (1984). "Inequality Decomposition by Population Subgroups". *Econometrica*, 52(6): 1369–85.

Peru: National Household Survey "25 years"

Nancy Hidalgo Calle, (National Institute of Statistics and Information Technology, Peru)

The National Household Survey (ENAHO), executed by the National Institute of Statistics and Informatics (INEI) is a statistical research that, since 1995, has been monitoring the indicators that allow to know the evolution of poverty, well-being and living conditions of households in Peru. As of 2003, ENAHO runs continuously throughout the year.

ENAHO is the main source of information for producing official statistics of national interest, such as indicators of monetary poverty, unsatisfied basic needs (NBI), employment levels, education, as input for the elaboration of poverty maps, among others.

The design of the questionnaire is obtained as a result of consensus with users, a practice that is strongly institutionalized. The questionnaire addresses issues such as citizen participation, access to social programs, governance, corruption, democracy, informality, financial inclusion, education, health, employment and income, among others; which make it a valuable source of information for public policy makers, the academic community and in the elaboration of cross-sectional and longitudinal studies through the panel component of the housing sample.

Committed to transparency, trust and credibility in the technical quality of the information provided by ENAHO, continuous improvement processes are carried out to guarantee data quality, from the collection in the field to the validation of methodologies for the production of final statistics.

In this context, in 2007, under the auspices of the World Bank (WB), INEI convened a Specialized Advisory Committee, integrated by representatives of international organizations, national government agencies, representatives of the academic community and research centers. This Committee participates every year in monitoring, verifying and guaranteeing the quality of the survey in the measurement of poverty and other indicators; subsequently, through Supreme Resolution No. 097-2010-PCM, the Poverty Advisory Committee is constituted as an Advisory Commission for Poverty Estimation and other related indicators in the country.

In 2010, ENAHO was distinguished by the World Bank with the **Regional Prize for Statistical Innovation in Latin America and the Caribbean.** In the contest participated 177 programs from 26 countries in the categories of censuses, surveys and administrative records.

According to the great transformation processes that societies live and that justify demographic, economic and behavioral changes in a population, there is a need to have a more consistent methodology and in accordance with the reality of the country. In this context, ENAHO and the Poverty Advisory Commission worked on updating the methodology for the measurement of monetary poverty; task that consisted mainly in the readjustment of the urban /rural population structure according to the 2007 National Censuses, the identification of new consumption patterns and the evaluation of the components of expenditure, changes in caloric needs and the inclusion of new sources of information such as the National Family Budget Survey 2008-2009; changing significantly the parameters that define poverty indicators. This work was completed in March 2012 and is a current methodology.

For the collection of information in the field, ENAHO makes use of digital technology since 2010, replacing physical questionnaires with the use of the PDA; later in 2016 the integral transition of this technology to the TABLET devices was done. The migration of all data processing and data collection programs to new technologies allowed us to obtain operational advantages in the field and opened the doors for housing georeferencing, real-time fieldwork monitoring, and online delivery (using a data plan) of the information collected at the conclusion of each interview.

The monitoring of the ENAHO field operation is a fundamental process for quality assurance, and it is done with the support of the Management System for Monitoring Data Collection, a system that aims to: 1) Follow up and control the different activities of the survey, providing timely information to ensure coverage and take preventive and corrective actions, 2) Improve the quality of the information collected in the field by monitoring quality indicators and assessing monetary variables, 3) Become an instrument of consultation for the personnel in charge of the analysis of the information that facilitates the detection of biases and/or inconsistencies in the information collected from the field in a timely manner.

The National Household Survey has a long history in the generation of official statistics on the households living conditions in Peru, applying throughout its 25 years of execution, improvements and innovations to ensure the quality of information which is available to general public at (https://www.inei.gob.pe/cifras-de-pobreza/). We are happy that various surveys from the National Household Survey are also available through the harmonized Luxembourg Income Study (LIS) Database and thus can easily be analyzed in cross-national perspective. In 2009, as part of the MECOVI program, LIS included the first dataset from Peru (2004). In 2015 and 2019 four more datasets were added.

On the 25th ENAHO`s anniversary, INEI-Peru expresses its recognition to all the people behind each process, who with their effort and professionalism make possible the execution of this survey, as well as to the users whose information demands always motivates the processes of continuous improvement. INEI makes also a special recognition to all Peruvian households that open their doors and provide their information.

Data News / Data Release Schedule

Data Releases – Luxembourg Income Study (LIS)

Belgium

Five new datasets from Belgium BE04 (Wave VI), BE07 (Wave VII), BE10 (Wave VIII), BE13 (Wave IX) and BE16 (Wave X) have been added to the *LIS Database*. The datasets are based on the 2005, 2008, 2011, 2014, and 2015 waves of the Survey on Income and Living Conditions (SILC) carried out by the Belgian statistical office (STATBEL).

Canada

Five new datasets from Canada, CA12 (Wave IX), CA14 (Wave IX), CA15 (Wave X), CA16 (Wave X), and CA17 (Wave X) have been added to the *LIS Database*. The datasets are based on the Canadian Income Survey (CIS) carried out by Statistics Canada.

Czech Republic

One new dataset from the Czech Republic, CZ16 (Wave X) has been added to the *LIS Database*. The dataset is from the 2017 wave of the Survey on Income and Living Conditions (SILC) carried out by the Czech Statistical Office.

Italy

One new dataset from Italy, IT16 (Wave IX) has been added to the *LIS Database*. The dataset is derived from the 2016 wave of the Survey of Household Income and Wealth (SHIW), carried out by the Bank of Italy.

Palestine

LIS is delighted to announce the addition of Palestine to the *LIS Database*. One data point has been added, PS17 (Wave X). The dataset is based on the Household Expenditure and Consumption Survey, 2016/2017, carried out by the Palestinian Central Bureau of Statistics.

Data Releases – Luxembourg Wealth Study (LWS)

Italy

One new dataset from Italy, IT16 (Wave X) has been added to the *LWS Database*. The dataset is derived from the 2016 wave of the Survey of Household Income and Wealth (SHIW), carried out by the Bank of Italy.

Note that the SHIW data are used for the creation of the Household Finance and Consumption Survey (HFCS) data of the European Central Bank (ECB). However, LIS has used the original SHIW data to create the LWS dataset.

Data Revisions – Luxembourg Income Study (LIS)

Belgium - Variables *emp* and *lfs* are now available in datasets BE85 and BE88. Datasets BE85, BE88, BE92, BE95, BE97, and BE00 have been revised for consistency, particularly the labour market and income sections.

Canada – Variables *hc41* (actual rent), *hc4* (actual rent and utilities), *hxmort* (mortgage instalments), and *hhouscost* (housing costs) are now available in datasets CA07, CA10, and CA13. Household composition and living arrangements related variables (*partner*, *parents*, *nchildren*, and *ageyoch*) are now available for a larger universe in CA07, CA10, and CA13.

Czech Republic - An error in variable *status1* (status in employment, main job) has been corrected in datasets CZ10 and CZ13.

Germany - Variable *hourstot* is now available in DE16.

Italy - The provision of more detailed variables for taxes and social contributions (as simulated by the Bank of Italy) allowed the addition of taxes and contributions to all labour income and pension variables in IT14; as a result, income variables are now reported gross rather than net and variable *grossnet* has been revised accordingly.

Sweden - Variables *emp* and lfs are now available in SE75. Variable *lfs* was reviewed in SE81 and SE05, with impact on *emp* in SE05 only.

General Revisions (entire LIS Database)

Variables *pitotal* (total individual income) and *pi42* (unemployment benefits) have been corrected to include also amounts assigned to *pi421* (unemployment insurance) and *pi422* (unemployment assistance). The following datasets are concerned:

Austria (AT94 AT97 AT00 AT07 AT10 AT13 AT16), **Belgium** (BE85 BE88 BE92 BE95 BE97 BE00), **Brazil** (BR06 BR09 BR11 BR13), **Canada** (CA87 CA91 CA94 CA97 CA98 CA00 CA04 CA07 CA10 CA13), **Switzerland** (CH07 CH10 CH13), **Chile** (CL90 CL92 CL94 CL96 CL98 CL00 CL03 CL06 CL09 CL11 CL13), **China** (CN02), **Czech Republic** (CZ92 CZ96 CZ02 CZ04 CZ07 CZ10 CZ13), **Denmark** (DK87 DK92 DK95 DK00 DK04 DK07 DK10 DK13 DK16), **Germany** (DE78 DE83 DE84 DE87 DE89 DE91 DE94 DE95 DE98 DE00 DE01 DE02 DE03 DE04 DE05 DE06 DE07 DE08 DE09 DE10 DE11 DE12 DE13 DE14 DE15 DE16), **Estonia** (EE04 EE07 EE10 EE13), **Spain** (ES95 ES00 ES07 ES10 ES13 ES16), **Finland** (FI87 FI91 FI95 FI00 FI04 FI07 FI10 FI13 FI16), **France** (FR78 FR94 FR00 FR05 FR10), **Greece** (GR95 GR00 GR04 GR07 GR10 GR13), **Hungary** (HU91 HU94 HU99 HU05 HU07 HU09 HU12 HU15), **Ireland** (IE94 IE95 IE96 IE00 IE04 IE07 IE10), **India** (IN04 IN11), **Iceland** (IS04 IS07 IS10), **Italy** (IT95 IT98 IT00 IT04 IT08 IT10 IT14), **Lithuania** (LT10 LT13), **Luxembourg** (LU85 LU91 LU94 LU97 LU00 LU04 LU07 LU10 LU13), **Mexico** (MX08 MX10 MX12 MX14 MX16 MX18), **Netherlands** (NL87 NL90 NL93 NL99 NL04 NL07 NL10 NL13), **Norway** (NO86 NO95 NO00 NO04 NO07 NO10 NO13), **Poland** (PL95 PL07 PL10 PL13 PL16), **Serbia**(RS06 RS10 RS13 RS16), **Russia** (RU00 RU04 RU07 RU10 RU11 RU13 RU14 RU15 RU16), **Sweden** (SE67 SE75 SE92 SE95 SE00 SE05), **Slovenia** (SI97 SI99 SI04 SI07 SI10 SI12 SI15), **Slovakia** (SK92 SK04 SK07 SK10 SK13), **United Kingdom** (UK69 UK74 UK79 UK86 UK91 UK94 UK95 UK99 UK04 UK07 UK10 UK13 UK16), **United States** (US74 US79 US86 US91 US94 US97 US00 US04 US07 US10 US13 US16), **Uruguay** (UY04 UY07 UY10 UY13 UY16), **Vietnam** (VN11 VN13), and **South Africa** (ZA08 ZA10 ZA12).

Variables *gross1*/*net1* (gross/net hourly wage, main job) have been converted to current currency. The following datasets are concerned:

> **Austria** (AT94 AT97 AT00), **Belgium** (BE95 BE97 BE00), **Czech Republic** (CZ96), **Germany** (DE81 DE84 DE87 DE89 DE91 DE94 DE95 DE98 DE00 DE01), **Estonia** (EE07 EE10), **Spain** (ES95 ES00), **France** (FR94), **Greece** (GR95 GR00), **Ireland** (IE94 IE95 IE96 IE00), **Italy** (IT87 IT89 IT91 IT93 IT95 IT98 IT00), **Luxembourg** (LU91 LU94 LU97 LU00), **Netherlands** (NL83 NL93 NL99), and **Slovakia** (SK04 SK07).

Variables *educlev* (highest completed education levels) and *educ* (education – 3-category recode) have been revised for consistency. As a result, variable *edyrs* (years of education) was adjusted accordingly.

The following datasets concern *educ* and *educlev*:

> **Austria** (AT04 AT07 AT10), **Canada** (CA71 CA75 CA81 CA87 CA91 CA94 CA97 CA98 CA00 CA04 CA07 CA10), **Czech Republic** (CZ92 CZ96 CZ02 CZ04 CZ07 CZ10 CZ13), **France** (FR78), **Hungary** (HU91 HU94), **Italy** (IT95 IT98 IT00 IT04 IT08 IT10 IT14), **Norway** (NO86 NO91 NO95), and **Poland** (PL86).

The following datasets concern *educlev* only:

> **Austria** (AT87 AT94 AT95 AT97 AT00 AT13 AT16), **Belgium** (BE97), **Germany** (DE13 DE14 DE15 DE16), **Denmark** (DK87 DK92 DK95 DK00 DK04), **Greece** (GR07 G10 GR13), **Hungary** (HU99 HU05 HU07 HU09 HU12 HU15), and **Poland** (PL92 PL95 PL99 PL04 PL07 PL10 PL13 PL16).

In addition, *edyrs* is now available in **Norway** and **Poland**.

Data Revisions – Luxembourg Wealth Study (LWS)

Germany - Variable *hourstot* is now available in DE17.

Italy - The provision of more detailed variables for taxes and social contributions (as simulated by the Bank of Italy) allowed the addition of taxes and contributions to all labour income and pension variables in IT14; as a result, income variables are now reported gross rather than net and variable *grossnet* has been revised accordingly.

Variable *bocd1_c* (constraints in debt repayment) is now available in IT08, IT10, and IT14.

Luxembourg - Replicate weights have been added to LU10 and LU14.

Japan - Variables *hxmort* (mortgage instalments), *hxloan* (installment for other loans), and *hhouscost* (housing costs) have been corrected in JP04.

Spain - Replicate weights have been added to ES02, ES05, ES11, and ES14.

Canada - Integrated net worth is now available in CA99, CA05, CA12, and CA16.

General Revisions (entire LWS Database)

Variables *pitotal* (total individual income) and *pi42* (unemployment benefits) have been corrected to include also amounts assigned to *pi421* (unemployment insurance) and *pi422* (unemployment assistance). The following datasets are concerned:

> **Austria** (AT11 AT14), **Germany** (DE12 DE07 DE02 DE17), **Greece** (GR09 GR14), **Italy** (IT14 IT10 IT08 IT04 IT00 IT95), **Luxembourg** (LU10 LU14), **Norway** (NO13 NO10), **Sweden** (SE05 SE02), **Slovenia** (SI14), **Slovakia** (SK10 SK14), and **United Kingdom** (UK11 UK09 UK07).

Variables *gross1*/*net1* (gross/net hourly wage, main job) have been converted to current currency. The following datasets are concerned:

> **Italy** (IT95 IT00).

Variables *educlev* (highest completed education levels) and *educ* (education – 3-category recode) have been revised for consistency. As a result, variable *edyrs* (years of education) was adjusted accordingly.

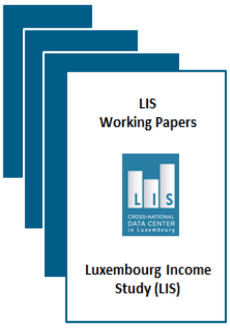
The following datasets concern *educ* and *educlev*:

> **Canada** (CA16 CA12 CA05 CA99), and **Italy** (IT95 IT00 IT04 IT08 IT10 IT14).

LIS/LWS Data Release Schedule

	Summer 2020	Autumn 2020
LIS Database		
Estonia	EE16	
Greece	GR16	
Ireland		IE13/16
Laos	LA02/07/12	
Latvia	LV13/LV16	
Lithuania	LT16	
Luxembourg	LU16	
Norway	NO16	
Panama		PA16
South Africa	ZA15/17	
South Korea	KR14/16	
Switzerland	CH16	
Vietnam		VN93/98/02/04/06/08/10
United Kingdom		UK17/15/14/12/11/09/08
United States	US17/18	US15/14/12/11/09/08
LWS Database		
Chile		CL07/12/14/17
Norway	NO16	
South Africa	ZA15/17	
United Kingdom	UK13/15	

Working Papers & Publications



Focus on 'Differences across Place and Time in Household Expenditure Patterns: Implications for the Estimation of Equivalence Scales' **LIS WP No.781 by *Angela Daley* (School of Economics, University of Maine), Thesia Garner (Bureau of Labor Statistics, US Department of Labor), Shelley Phipps (Department of Economics, Dalhousie University), Eva Sierminska (Luxembourg Institute of Socio-Economic Research)**

When comparing economic well-being using income or expenditures, equivalence scales are often used to adjust for differences in characteristics that affect needs. For example, a family of two is assumed to need more income than a single person, but not twice as much due to the economies of scale in consumption. There are different types of equivalence scales that yield different estimates of economies of scale, and thus different estimates of economic well-being. However, a common equivalence scale is often used when comparing economic well-being across countries and across time. In this study, we ask whether it is appropriate to use a common equivalence scale across countries and time if consumption expenditure patterns differ? Based on an Engel methodology, we estimate equivalence scales for a diverse set of countries (Canada, France, Israel, Poland, South Africa, Switzerland, Taiwan, United States) in different time periods (1999-2012). Our data come from Statistics Canada (Survey of Household Spending), the Bureau of Labor Statistics (Consumer Expenditure Survey) and Luxembourg Income Study Data Center (an archive of harmonized survey data across countries). We estimate relative needs by looking at the shapes of equivalence curves for households of different sizes, as well as smoothed single-parameter estimates, for three necessity bundles: (1) food; (2) food, housing and clothing; (3) food, housing, clothing and health care. We find that equivalence scales differ across bundles; for most countries, economies of scale are larger when considering necessities other than food. Moreover, we find considerable differences in economies of scale across countries for all bundles. For example, based on the third necessity bundle, a family of two needs between 27.9 percent (Canada) and 59.8 percent (Israel) more income than a single person. The average across countries is 43.1 percent. We also find that economies of scale have increased over time, and our single-parameter estimates imply larger economies of scale than the widely accepted 'square root of household size' equivalence scale. The latter corresponds to a value of 0.5, which is greater than our single parameter estimates that range from 0.25 in South Africa to 0.47 in Israel. Taken together, these findings suggest that using a common equivalence scale to compare economic well-being across countries and time is misleading. Specifically, if economies of scale are understated (as is the case when using the 'square root of household size'), the relative poverty experienced by larger versus smaller families is being overstated.

LIS working papers series

LIS working papers series - No. 781
Differences across Place and Time in Household Expenditure Patterns: Implications for the Estimation of Equivalence Scales
by Angela Daley, Thesia Garner, Shelley Phipps, Eva Sierminska

LIS working papers series - No. 782
Income Growth and Preferences for Redistribution: The Role of Absolute and Relative Economic Experiences
by David Weisstanner

LIS working papers series - No. 783
Experience and Perception of Social Mobility - a Cross-Country Test of the Self-Serving Bias
by Nina Weber

LIS working papers series - No. 784
The Contribution of the Spatial Dimension to Inequality: A Counterfactual Analysis for OECD Countries
by Luis Ayala, Javier Martín-Román, Juan Vicente

LIS working papers series - No. 785
Consumption Taxes and Income Inequality. An International Perspective with Microsimulation
by Julien Blasco, Elvire Guillaud, Michaël Zemmour

LWS working papers series

LWS working papers series - No. 30
Housing, Wealth Accumulation and Wealth Distribution: Evidence and Stylized Facts
by Orsetta Causa, Nicolas Woloszko, David Leite

Published in the OECD Economics Department Working Papers, no. 1588 (2019), OECD Publishing, Paris, https://doi.org/10.1787/86954c10-en

News, Events and Updates

The New LIS Data Access Research Tool (DART). Stay Tuned!

LIS is looking forward to release a new interactive visualization tool (DART). DART is a powerful web-based data access tool populated with various national indicators on income and wealth, across countries and over time, based on the *LIS Databases*. DART's innovative feature is its richness of inequality measures disaggregated by different social strata.

With the launch of DART, LIS foresees to serve a broader base of users who will be able to create summaries of indicators, tailored to their interests and needs.

Stay tuned for its official launch on April 15th.

The Comparative Welfare States Dataset, 2020

The Comparative Welfare States Dataset assembled by David Brady, Evelyne Huber, and John D. Stephens has been updated. Most variables now have data up to 2016 or 2017. The data cover earnings and income distribution, social spending and welfare state institutions, labor force and labor market institutions, demographic data, macroeconomic data, research and development spending, product market regulation, and political variables like voter turnout and partisan distribution of votes, seats, and cabinet share. They are available for 22 post-industrial countries and go back to 1960 when possible.

The 2020 version including Codebook, Dataset (in excel format), and Dataset (in Stata format), are now available for download.

The earlier version of 2014 including Codebook, Dataset (in excel format), and Dataset (in Stata format), is still available.

Application for the LIS Summer Workshop 2020 Is now Open!

The LIS Summer Workshop will be held on 06-10 July, and is organised with the University of Luxembourg and LISER, the workshop has been recently named the Summer Workshop on Inequality and Poverty Measurement. This workshop is a one-week intensive course designed to introduce researchers in the social sciences to comparative research on income and wealth distribution, employment and social policy, using the harmonised Luxembourg Income Study (LIS) and the Luxembourg Wealth Study (LWS) Databases. Attendees will be trained to use both databases independently and will have the opportunity to:

- Acquire advanced knowledge about methods used in inequality research
- Gain skills related to the study of comparative inequality
- Learn in detail about the LIS and LWS data and develop ties with LIS' large international network.

Researchers and doctoral students from various social science disciplines are invited to apply. To apply, kindly fill the online application form available: here.

Application deadline : April 15, 2020

Acceptance announcement : by April 22, 2020

For more information, please visit our webpage.

(LIS)²ER Stream at ESPAnet 2020, Leuven

In context of the new (LIS)²ER project, Daniele Checchi, Petra Sauer, and Philippe Van Kerm host a stream on "Methodologies for comparative social policy analysis" at ESPAnet 2020 in Leuven.

Different countries pursue different policy goals with alternative policy instruments, and government turnover leads to changes in policy objectives and implementations over time within the same country. However, while there is a large literature describing patterns of inequality which takes a cross-country analysis of time variation, there is much less research on variations in policy packages (welfare policies, tax policies, labour market regulation, educational policies) and on their impact of inequality and poverty. One main reason lies in the absence of appropriate, consistently defined and comparable indicators of the policy stance with respect to specific dimensions. Take the United States as a point of comparison. A wealth of research exploits variations across States and over time to assess the impact of policy decisions on a wide range of dimensions; Hoynes and Patel's (Journal of Human Resources, 2018) recent analysis of the Earned Income Tax Credit impact on inequality and poverty reduction is only one of many examples. Such research design is largely unequalled elsewhere around the globe. This is all the more regrettable given the increasingly recognized 'American exceptionalism' in policy preferences and income distributions. There is a crucial need for analysis of policy impacts in different demographic, economic, and institutional environments.

This stream invites papers which take novel approaches to comparative social policy analysis, using different methodologies and datasets to tackle the task to make "measurements" of policy frameworks amenable to empirical research. We particularly appreciate studies which contribute to our understanding of (the evolution) of different welfare models around the globe, and provide insights into which policy packages work to fight poverty inequality. We also welcome research which analyses the (causal) impact of policy changes onto several other social dimensions, such as education, labour market participation, employment, household formation, health or well-being.

The deadline for abstract submission is 15 April 2020. For further practical information on the call for abstracts please click here.

Visiting Scholars

During this quarter, LIS welcomed four visiting scholars who came to work onsite with the *LIS Databases* in the framework of the InGRID2 project; namely Roberto Iacono, Elisa Palagi, Anders Villadsen, and Pedro Salas Rojo.

Roberto Iacono is Associate Professor at the Norwegian University of Science and Technology (NTNU) in Trondheim, Norway. Elisa Palagi is a Ph.D. student in Economics at the Sant' Anna School of Advanced Studiesin Pisa, Italy. They are currently working on a joint research project related to income composition inequality in the Nordic countries, using the LIS database. This project aims at understanding how inequality in the composition of incomes, with a focus on labour and capital income, is evolving in countries that are otherwise regarded as egalitarian with respect to pre-tax income inequality.

During his stay in February, Professor Anders Villadsen (Aarhus University) worked on a project comparing the immigrant wage gap in the public and private sectors across a range of countries. With this project, he aims to explore the hypothesis that public employment, compared with private employment, is more inclusive of immigrants and the degree to which such a tendency varies across different countries with different organizations of the public sector.

In March, LIS welcomed Pedro Salas-Rojo, a PhD student at Complutense University of Madrid. During his stay, Pedro was measuring the impact of inheritances and parental education on several wealth definitions, by using Machine Learning algorithms. He is particularly interested in discerning by how much those factors affect the current wealth distribution, and how they affect the probability of acquiring new assets. During his stay, he expanded his sample of countries to the UK, US, Italy and, to a lesser extent, Canada and Austria, in addition to Spain.

Stone Center – Launched Working Paper Series

On February 20, the Stone Center launched a new Working Paper Series. The papers are authored or co-authored by scholars who work with the Stone Center, including the center's core faculty and postdoctoral scholars, Affiliated Scholars, and PhD students. These papers are data-driven, interdisciplinary, methodologically diverse, and policy-oriented, addressing a broad array of questions about inequalities throughout the world. Papers include works-in-progress and pre-publication versions of articles. Many of them will be published subsequently in journals, or in authored and edited volumes. (Papers that use the LIS/LWS data will first be added to the LIS Working Paper Series.)

In addition to appearing on the Stone Center's website, the papers are archived at SocArXiv, an online server for the social sciences, dedicated to the proliferation of open science.

Stone Center – Co-hosted NBER/CRIW Conference

On March 5-6, the Stone Center – with several other institutions – co-hosted the: "Conference on Measuring and Understanding the Distribution and Intra/Inter-Generational Mobility of Income and Wealth". Ten people associated with the Stone Center participated: Janet Gornick, Branko Milanovic, Salvatore Morelli, Marco Ranaldi, Charlotte Bartels, Yonatan Berman, Nathaniel Johnson, Joseph Van Der Naald, Ercio Andrés Muñoz Saavedra, and Arthur Kennickell. Stone Center Visiting Scholar (spring 2020) Stephen Jenkins also attended. In addition, two long-time LIS Board Members presented: Pirmin Fessler and Richard Tonkin.

Stay tuned for news about the edited conference volume!

Editor: Jörg Neugschwender
Layout and Design: Heba Omar & Jörg Neugschwender
Photo credit: Cover @ ibreakstock/fotolia.com