

A Comparison of Top Income Shares and Global Inequality Datasets

By James K. Galbraith and Béatrice Halbach
The University of Texas Inequality Project
August 22, 2016

Working Paper # 73

Abstract

This paper presents a comparison of coverage and values between the Top Income Shares of the World Wealth and Income Database (WWID), published by the World Inequality Lab at the Paris School of Economics, and the Estimated Household Income Inequality (EHII) data set of the University of Texas Inequality Project. The WWID and EHII have major international coverage and present measurements of the distribution of income and wealth and estimated Gini coefficients of gross household income inequality, respectively. While these two concepts are not directly comparable, arguably they should be reasonably consistent, with a high top-income share corresponding, in most cases with a high measure of income inequality. In terms of coverage, the paper shows the breadth of EHII in comparison to limited and regional coverage of the Top Income Shares. A rank-order comparison of inequality across countries shows inconsistencies between the top income shares and EHII, as well as between the WWID measures and other inequality measures published by the Luxemburg Income Studies (LIS) and the OECD.

We thank the Institute for New Economic Thinking for generous support.

I. Introduction

Galbraith *et al.* 2016 compare inequality measures published by the Luxembourg Income Studies (LIS), the Organization for Economic Co-operation and Development (OECD), the European Union's Statistics on Incomes and Living Conditions (EU-SILC) and the World Bank's World Development Indicators (WDI) with the Estimated Household Income Inequality (EHII) data set of the University of Texas Inequality Project. They illustrate the historical depth and range of EHII across countries, and its broad comparability with LIS, OECD and EU-SILC, as well as problems with the cross-country comparability of the WDI, which result largely from the use of inconsistent income concepts. Rossi 2016 presents a supplementary comparison of the Socio-Economic Database for Latin America and the Caribbean (SEDLAC) with the above-mentioned data sets. She finds that the movement of inequality in Latin America as reported by EHII corresponds broadly with trends in inequality reported by the other data sets, while offering superior coverage of the region.¹

This paper adds a comparison of coverage and values between the Top Income Shares of the World Wealth and Income Database (WWID), published by the World Inequality Lab at the Paris School of Economics, and EHII. In addition, I present a rank-order comparison of cross-country inequality between the WWID's Top Income Shares and other measures of inequality, including EHII, LIS, and OECD.

II. Top Income Shares and EHII

A. Coverage

Table 1 compares the coverage of EHII and the Top Income Shares (Top 1%, Top 0.5%, Top 0.1%) in terms of the total number of observations and the range of countries and time period covered. The Top Income Shares contain less than one third the number of observations of EHII data set, and less than a quarter of the countries. The Top Income Shares also extends further in time than EHII, with the earliest observation beginning 13 years earlier and the latest observation ending 6 years later. However, for the set of overlapping countries and years, the EHII's coverage is more comprehensive than each of the Top Income Shares.

Table 1 presents this comparison. EHII has more observations not only for the countries and years covered by EHII, but also for the countries and years covered by the WWID, except in one case – the Top 0.5% share – where the number of observations are nearly equal.

¹ In addition, a recent special issue on Cross-National Inequality Databases in the *Journal of Economic Inequality* (see references) includes reviews of several global inequality data sets, a summary of which can be found in Galbraith *et al.* 2016.

Table 1. Comparison of Coverage Across EHII and WWID

Data Set	Observation Type	Total Observations	Countries Covered	Years Covered	# WWID observations for countries and years covered by EHII	# EHII observations for countries and years covered by WWID
EHII	Gross Income Gini	3842	147	1963-2008	n.a.	n.a.
WWID	Top 1% Income Shares	1032	28	1950-2014	973	989
WWID	Top 0.5% Income Shares	1027	30	1950-2014	970	1025
WWID	Top 0.1% Income Shares	1117	34	1950-2014	1063	1156

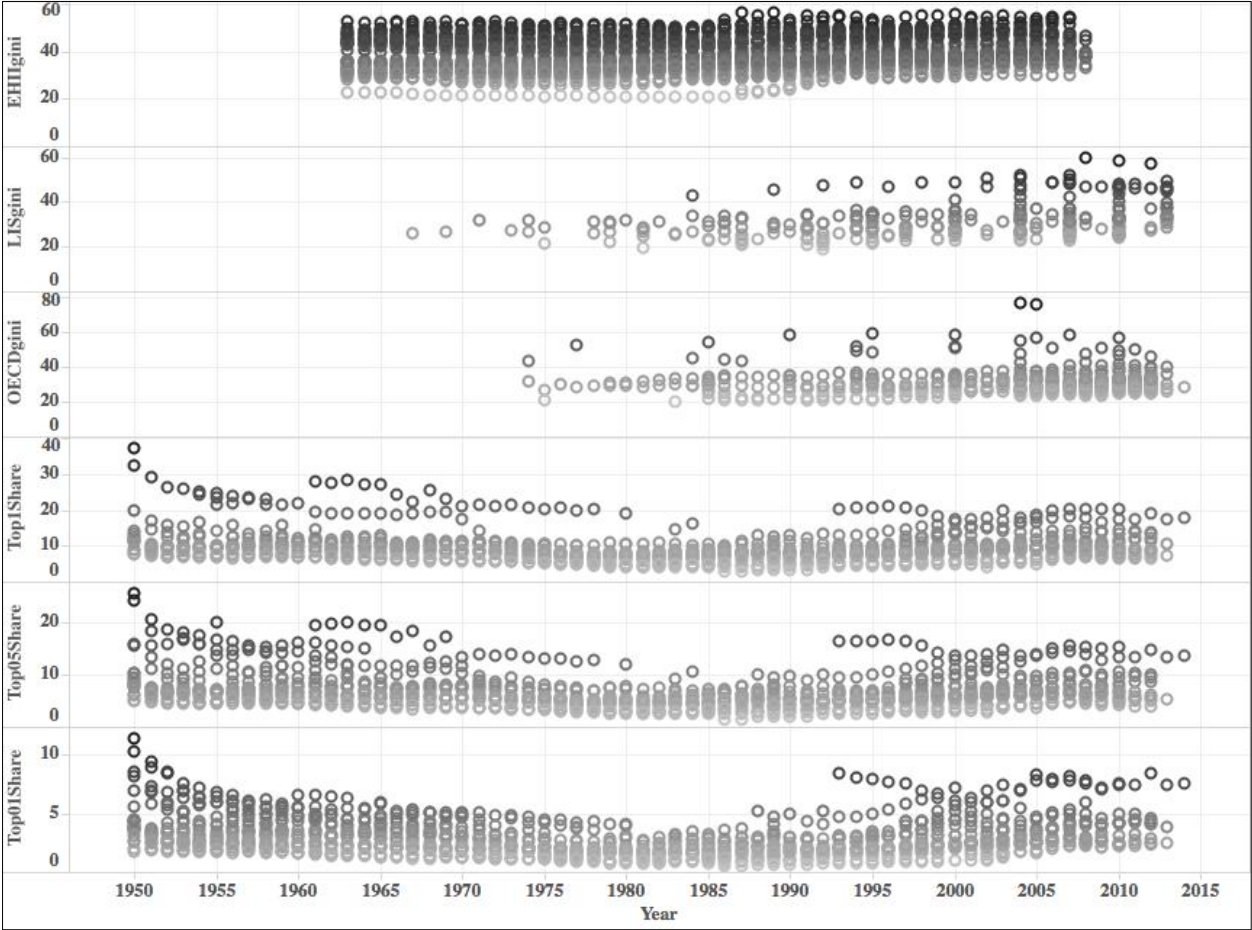
Table 2 compares the coverage of the EHII and WWID across four regions: the Americas, Eurasia (Europe and Central Asia), Asia and Oceania, and Africa and the Middle East. Among the observations of top income shares, approximately 18% are concentrated in the Americas, 36% in Eurasia, 32% in Asia and Oceania and 14% in Africa and the Middle East. For EHII, the corresponding numbers are 19.5%, 32.4%, 18% and 30.1%. While the concentration is similar for the Americas and in Eurasia, the numbers show that the WWID observations are more concentrated in Asia and Oceania than in Africa and the Middle East, and vice versa for EHII.

Table 2. EHII and WWID Coverage by Regions

Regions	Americas	Eurasia	Asia and Oceania	Africa and Middle East	Total
Top 1% Income Shares, Observations 1950-2014	155	414	359	104	1032
Number of Countries (Top 1%)	5	10	9	4	28
Top 0.5% Income Shares, Observations 1950-2014	153	414	312	148	1027
Number of Countries (Top 0.5%)	5	10	7	8	30
Top 0.1% Income Shares, Observations 1950-2014	153	422	344	198	1117
Number of Countries (Top 0.1%)	5	10	9	10	34
EHII Observations 1963-2008	750	1245	692	1155	3842
Number of Countries (EHII)	29	44	24	50	147

Figure 1 highlights the density of observations and countries in EHII, and extends the comparison of coverage to include the LIS and OECD inequality measures, which are focused mainly on middle- and high-income countries.

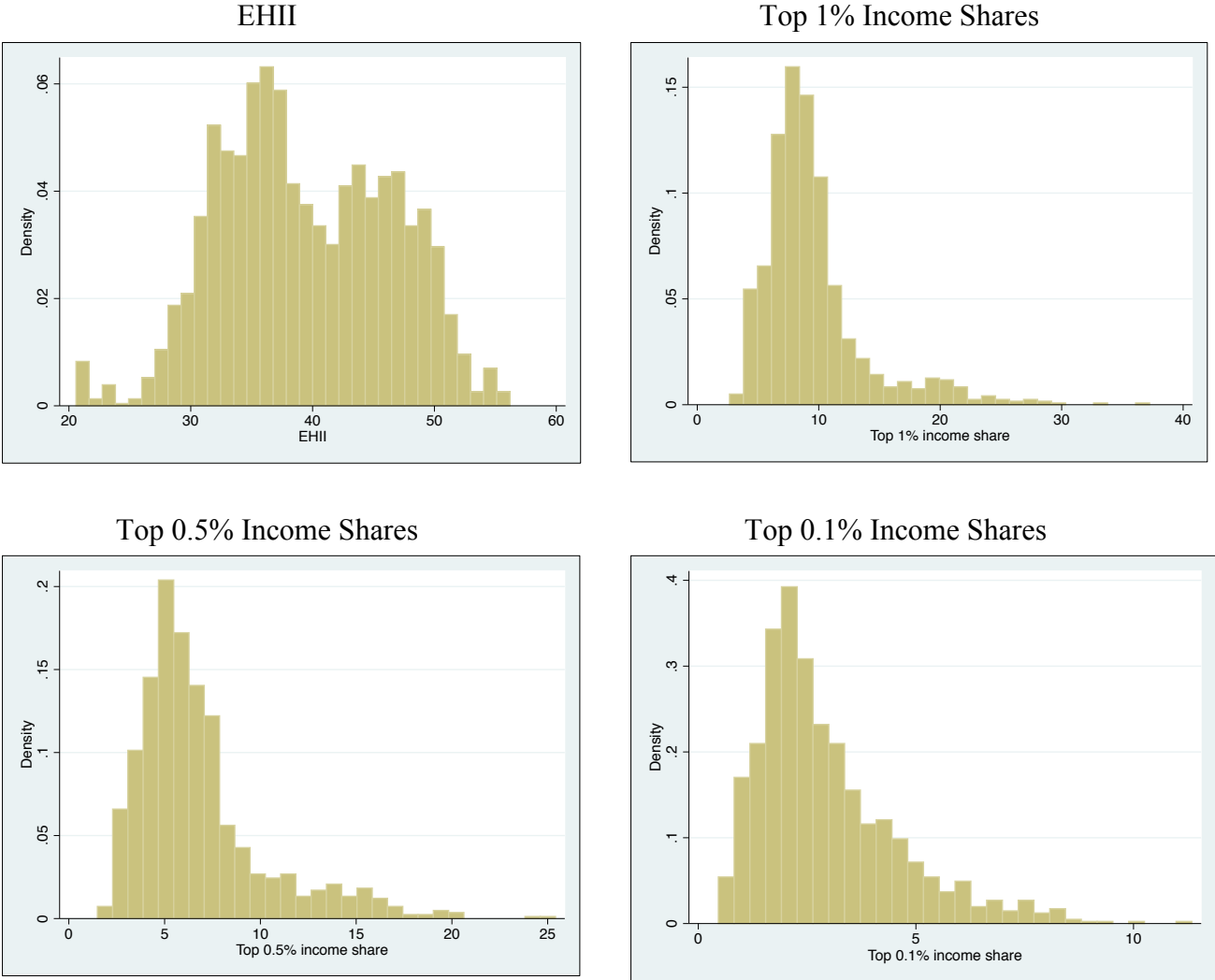
Figure 1. Coverage Across Inequality Data Sets



B. Strength of Association Between Inequality Measures

Spearman’s correlation coefficient measures the strength of association between two ranked variables. It can be used when the assumptions for using Pearson’s correlation coefficient are not met. In the case of the EHII and Top Income Shares, several of these assumptions are violated. First, none of the variables is normally distributed. As shown in Figure 2, EHII follows a bi-modal distribution, while the top 1%, top 0.5% and top 0.1% income shares have right skew.

Figure 2. Non-Normal Distribution of EHII, Top Income Shares



In addition, the EHII measures and Top Income Shares do not have a linear relationship, as required for Pearson’s Coefficient. Spearman’s coefficient requires a monotonic relationship between the two variables, a less restrictive assumption. Figure 3 shows scatter plots for the EHII and Top Income Shares, which indicate an exponential relationship between the EHII measures and the Top Income Shares.

Figure 3. Non-Linear Relationship between EHII and Top Income Shares



Table 3 presents the Spearman rank-order correlation coefficient between EHII and the WWID top share measures across regions, along with the number of overlapping observations and countries for each region. For comparison purposes, Table 4 presents the same results for EHII, LIS and OECD measures.

The Spearman correlation coefficients across the total number of observations between EHII and the top 1%, 0.5% and 0.1% income shares are 0.51, 0.58, and 0.56, respectively. In comparison, Spearman’s Coefficient between the EHII and LIS measures and between the EHII and OECD measures is .75 and .79, respectively. This indicates a stronger association in terms of country rankings across these data sets than between EHII and the Top Income Shares.

Spearman’s coefficient between EHII and the Top Shares also differs substantially across regions. In general, the strength of association is highest in Africa and the Middle East, followed

by the Americas, then Asia and Oceania, and finally with Eurasia showing the lowest correlation, between .42 and .52 for different income-share measures. However, because Spearman's coefficient measures the extent to which the ranking of countries differs depending on the measure of inequality used, it is highly sensitive to the number of countries included in the calculation. Thus, Spearman's Coefficient between EHII and the Top 1% Income Share for Africa and the Middle East is 1.0, meaning that the countries included in the calculation are ranked in the same order of inequality according to both measures. However, this does not necessarily indicate a very strong association between the two measures, given that the calculation is based on only 2 countries in this region. In general, the coefficient declines as countries are added to the comparison, a fact which is not wholly reassuring.

Table 3. Spearman's Correlation Coefficient, EHII and Top Income Shares Across Overlapping Countries and Years

Region	Top 1%	# Obs	# Countries	Top 0.5%	# Obs	# Countries	Top 0.1%	# Obs	# Countries
Americas	0.9894	99	4	0.9914	97	4	0.9914	97	4
Eurasia	0.4258	270	10	0.4248	272	10	0.5221	267	10
Asia & Oceania	0.639	234	9	0.6696	202	7	0.5661	220	9
Africa & Middle East	1	54	2	0.814	69	5	0.6375	84	7
Total	0.5087	657	25	0.5752	640	26	0.5569	668	30

Table 4 presents a comparison of EHII rank-orderings with those of the LIS and the OECD data sets – restricted to the overlapping years and observations. Note that the consistency of EHII to these data sets appears considerably higher than with the top income shares, even when as many as 24 countries are included in the ranking, as in the case of Eurasia. Of course the small overlaps for Asia/Oceania and for Africa/Middle East reflect the fact that these regions are not well covered by the OECD or by LIS, and the coefficients in these cases cannot tell us very much.

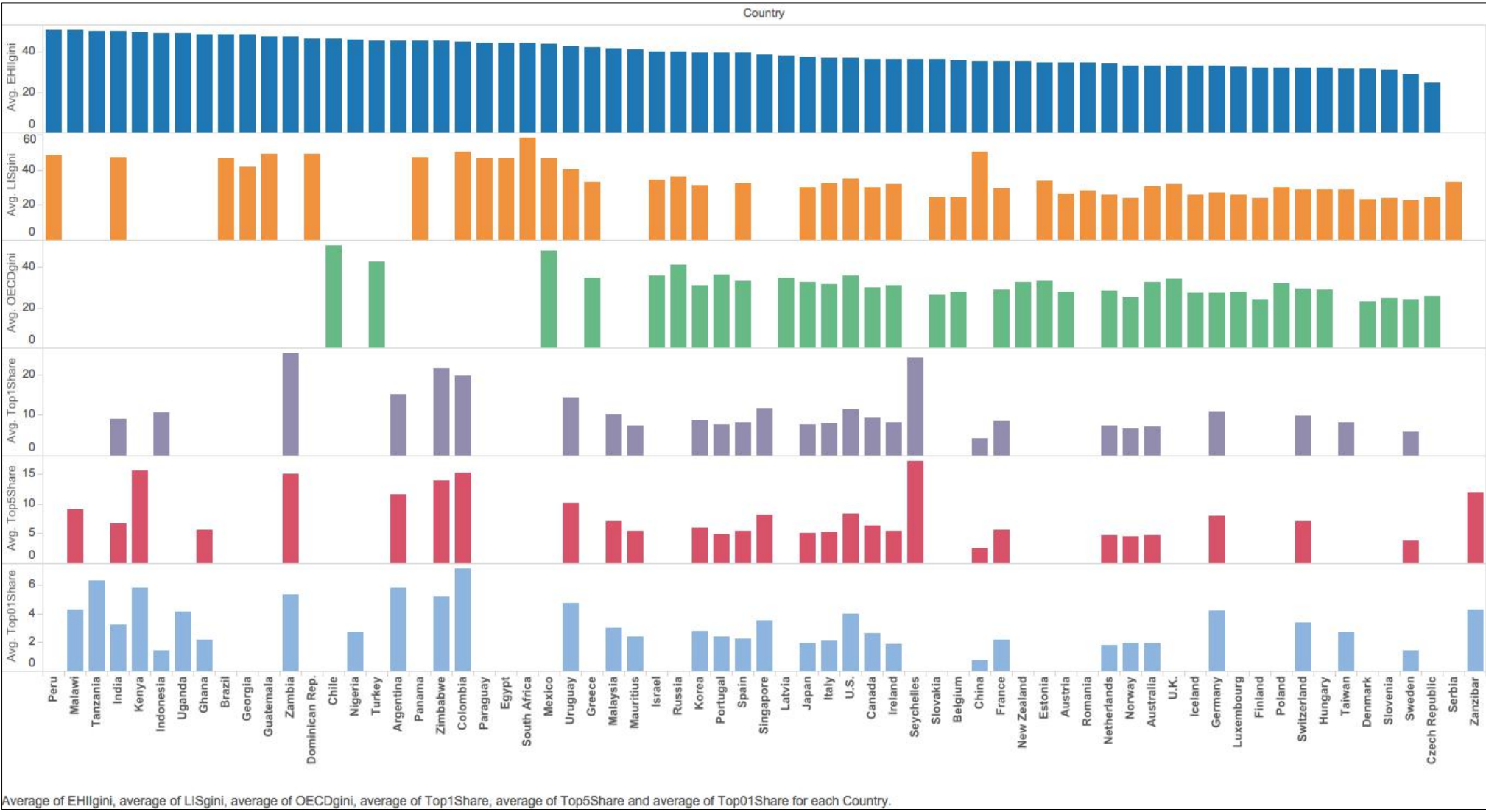
Table 4. Spearman's Correlation Coefficient, EHII, LIS and OECD Across Overlapping Countries and Years

Region	LIS	# Obs	# Countries	OECD	# Obs	# Countries
Americas	0.9703	33	7	1	63	4
Eurasia	0.6174	133	24	0.8231	178	24
Asia & Oceania	1	11	4	0.8582	10	4
Africa & Middle East	1	7	2	NA	5	1
Total	0.7459	184	37	0.7948	256	33

III. A Rank-order Comparison of Inequality Across Countries

Figure 4 compares the average measure by country for EHII, LIS, and the WWID Top Shares, for the set of overlapping countries between the three data sets. The countries are ranked in order of highest to lowest inequality according to EHII. The figure highlights the relatively close alignment in the ranking of countries between the EHII, LIS and OECD data sets, and the presence of major discrepancies between these data sets and the Top Income Shares. In addition, the figure highlights the problem of missing data in the Top Income Shares collection.

Figure 4. A Rank-order Comparison of Inequality Across Countries: EHIIgini, LIS, OECD and Top Income Shares



Average of EHIIgini, average of LISgini, average of OECDgini, average of Top1Share, average of Top5Share and average of Top01Share for each Country.

Figures 5 through 8 present the same rank-order comparisons, but this time organized by regions. This helps the reader pick out the important divergences in coverage and valuations. In particular, top-income India, Indonesia, and China seem notably low by comparison with income inequality in those countries. In the case of Germany, the WWID measures seem to come in higher than the comparison would predict.

Figure 5. A Rank-order Comparison of Inequality Across Countries and by Region: EHI, LIS, and Shares: Africa and the Middle East

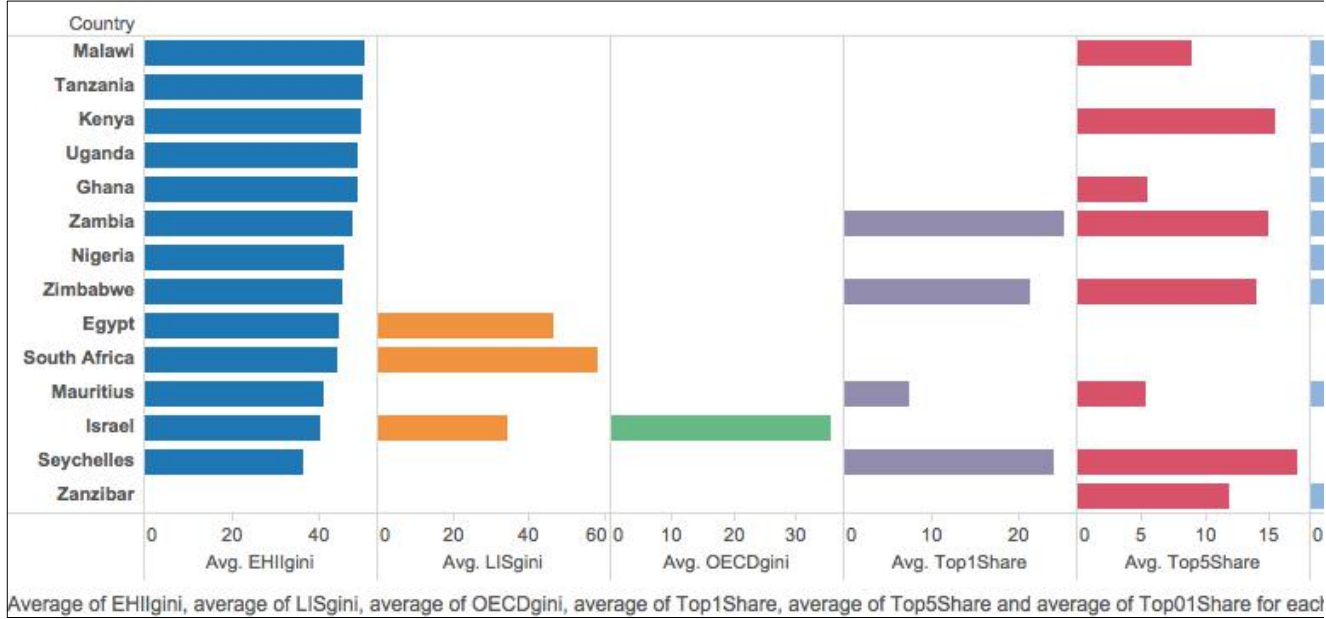


Figure 6. A Rank-order Comparison of Inequality Across Countries and by Region: EHI, LIS, and Shares: Asia and Oceania

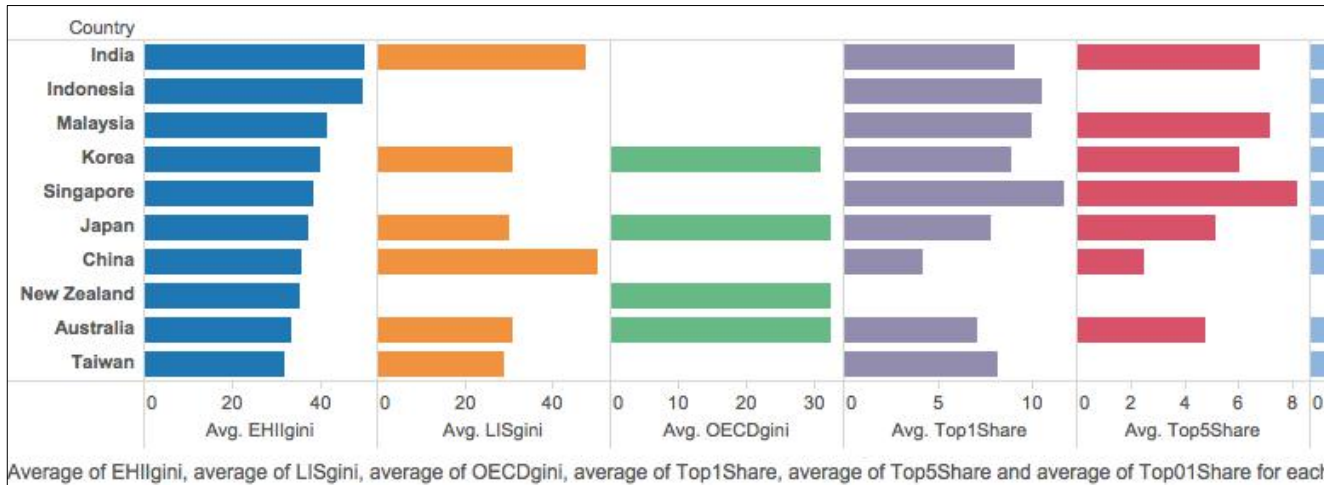


Figure 7. A Rank-order Comparison of Inequality Across Countries and by Region: EHI, LIS, and Top Income Shares: Europe and Central Asia

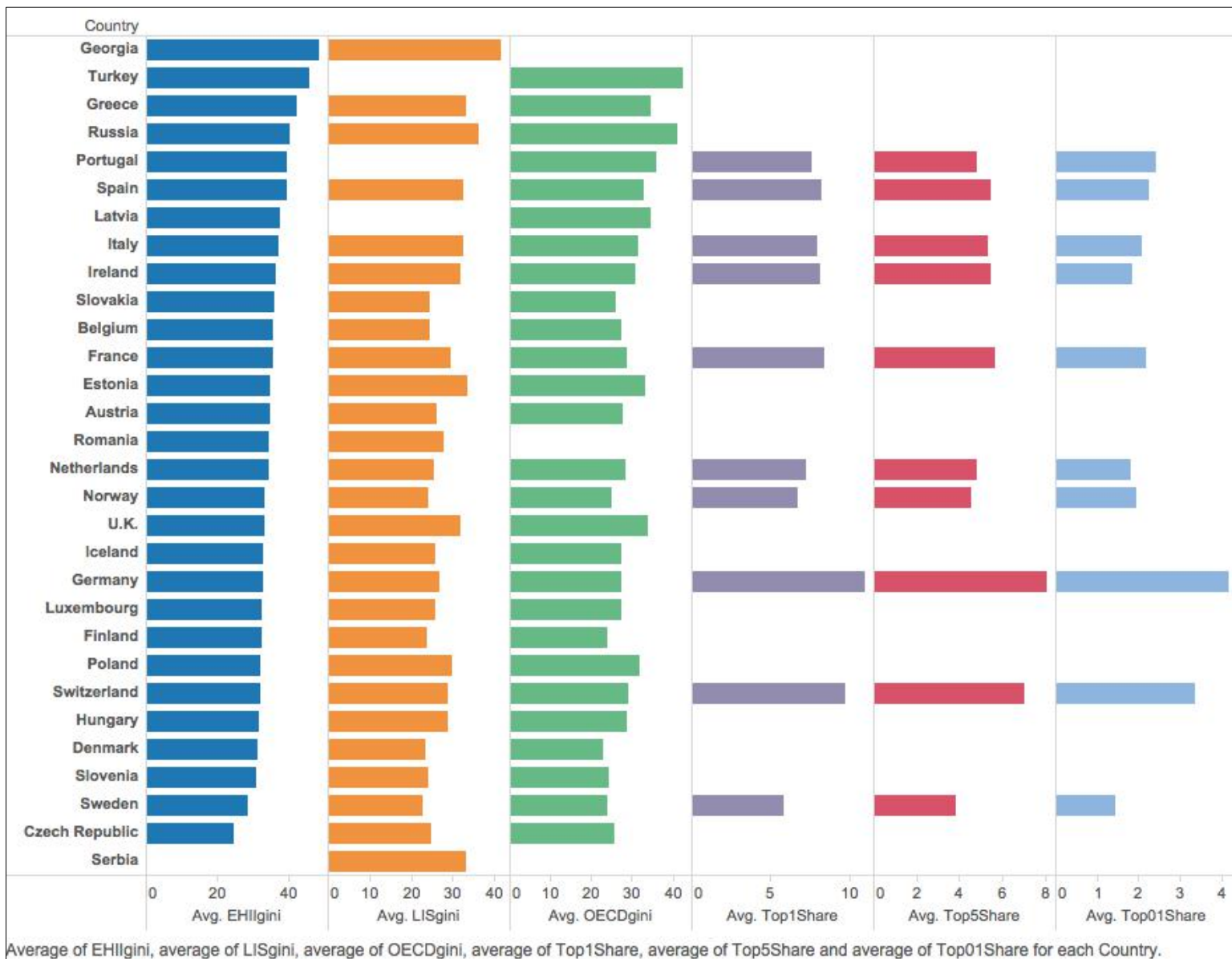
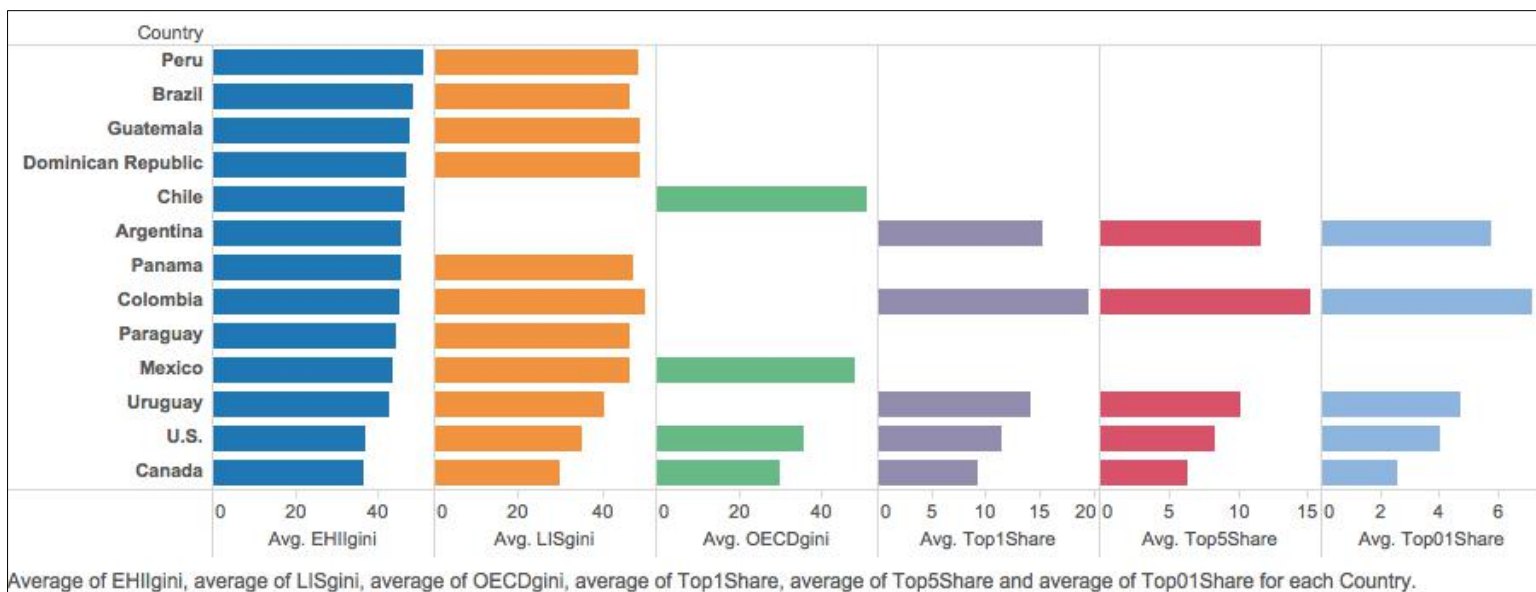


Figure 8. A Rank-order Comparison of Inequality Across Countries and by Region: EHI, LIS, and Top Income Shares: North and South America



IV. Conclusion

The creation of consistent global data sets for the purpose of comparing economic inequality across countries and through time is a work in progress, and multiple approaches in terms of data and method are to be welcomed. EHII, LIS, the OECD and the WWID each take distinctive approaches. EHII draws on measures mainly of inequality in industrial pay and estimates household income inequality from that, using a simple model. LIS works by combining and standardizing the micro-data sets available in different countries. The OECD draws on national data sources and estimates for the wealthy countries who form that organization. The WWID is drawn from tax records. And since the concept of “top income share” is not the same as the concept of “household income inequality” there is no reason why different measurements should align exactly.

Nevertheless, income inequality is a seamless phenomenon. It stands to reason that in general – if not in every case – a country with a high share of income concentrated in the top echelons should also have a high degree of income inequality – and conversely for those with low top-income shares. So the two sets of measures should be reasonably consistent.

We find here that, so far as comparisons are possible owing to the limited coverage of the LIS, OECD and WWID, that the first two are broadly consistent with EHII while the third – the WWID – is less so, and also less consistent with the OECD and LIS. There are many reasons why this might be the case, especially insofar as tax laws (and therefore the definition of income) differ between countries and change over time as tax laws evolve. Further, there is the problem of tax evasion – a problem whose severity may differ between countries, according to the rate structure imposed in the income tax and the efficiency of income-reporting and of tax enforcement.

These considerations suggest that the use of the WWID measures for the purpose of comparing countries in terms of their general degree of equality or inequality, should be done, if at all, with caution.

References

- Galbraith, James K., Béatrice Halbach, Jaehye Choi, Aleksandra Malinowska, and Wenjie Zhang. "A Comparison of Major World Inequality Data Sets: LIS, OECD, EU-SILC, WDI and EHII." *Research in Labor Economics, Income Inequality Around the World*, 44 (2016): 1–48.
- Rossi, Delfina (2016). "A comparison of Latin America inequality data sets." Working Paper 72. The University of Texas Inequality Project. Lyndon B. Johnson School of Public Affairs. The University of Texas at Austin.

Additional Literature on Cross-National Inequality Data Sets

- Atkinson, A.B., and A. Brandolini. "On Data: A Case Study of the Evolution of Income Inequality across Time and across Countries." *Cambridge Journal of Economics* 33 (2009): 381–404. doi:10.1093/cje/bel013.
- . "Promise and Pitfalls in the Use of 'secondary' Data-Sets: Income Inequality in OECD Countries as a Case Study." *Journal of Economic Literature* 34, no. 3 (2001): 771–99.
- Bourguignon, François. "Appraising Income Inequality Databases in Latin America." *Journal of Economic Inequality*, Special Issue on Cross-National Inequality Databases (August 20, 2015): 1–22.
- Brandolini, A. "The World Top Incomes Database - WTID: An Assessment." In *Appraising World Income Inequality Databases*. Helsinki, Finland, 2014.
- Gasparini, Leonardo, and Leopoldo Tornarolli. "A Review of the OECD Income Distribution Database." *Journal of Economic Inequality*, Special Issue on Cross-National Inequality Databases (March 18, 2015): 1–24.
- Jenkins, Stephen. "World Income Inequality Databases: An Assessment of WIID and SWIID." *Journal of Economic Inequality*, Special Issue on Cross-National Inequality Databases (August 22, 2015): 1–43. doi:10.1007/s10888-015-9305-3.
- Ravaillon, Martin. "The Luxembourg Income Study." *Journal of Economic Inequality*, Special Issue on Cross-National Inequality Databases (May 7, 2014): 1–21.
- Smeeding, Timothy, and Jonathan P. Latner. "PovcalNet, WDI and 'All the Ginis': A Critical Review." *Journal of Economic Inequality*, Special Issue on Cross-National Inequality Databases (July 22, 2015): 1–26. doi:10.1007/s10888-015-9312-4.