A comparison of major world inequality data sets:

LIS, OECD, SILC, WDI and EHII

By

James K. Galbraith, Jaehee Choi, Béatrice Halbach, Aleksandra Malinowska, and Wenjie Zhang

The University of Texas Inequality Project

Lyndon B. Johnson School of Public Affairs

The University of Texas at Austin

March 7, 2015

UTIP Working Paper 69

Abstract

We present a comparison of coverage and values for five inequality data sets that have world-wide or major international coverage and independent measurements that are intended to present consistent coefficients that can be compared directly across countries and time. The comparison data sets are those published by the Luxembourg Income Studies (LIS), the OECD, the European Union's Statistics on Incomes and Living Conditions (SILC) and the World Bank's World Development Indicators (WDI).The baseline comparison is with our own Estimated Household Income Inequality (EHII) data set of the University of Texas Inequality Project. The comparison shows the historical depth and range of EHII and its broad compatibility with LIS, OECD and SILC, as well as problems with using the WDI for any cross-country comparative purpose. The comparison excludes the large World Incomes Inequality Database (WIID) of UNU-WIDER and the Standardized World Income Inequality Database (SWIID) of Frederick Solt; the former is a bibliographic collection and the latter is based on imputations drawn, in part, from EHII and the other sources used here.

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

Corresponding author: Galbraith@mail.utexas.edu Comments Welcome.

The state of world inequality data

Since the landmark publication by the World Bank in 1996 of the Klaus Deininger - Lyn Squire (DS) data set of world-wide inequality measures, comparative, time-series and panel studies of economic inequality have become a significant field of economic research. But the ambitions of researchers have often run ahead of the quality, consistency and coverage of the data, so that many empirical questions remain open to dispute. This situation has in turn spurred new efforts to develop better and more consistent comparative measures of income inequalities.

The data sets now available are of five broad types. There are, first, large bibliographic data sets, of which the preeminent example is the World Income Inequality Database (WIID) of the World Institute for Development Economics Research (WIDER) of the United Nations University (UNU) at Helsinki. WIID is the successor to DS, and is a diverse collection of coefficients, chronicling the struggle to measure inequality around the world over the past six decades, with careful documentation as to the concepts and sources of information. But the WIID is not, itself, a data set of comparative measures. It is rather a source, from which such measures may be extracted, according to the preferences and criteria of the researcher.

The opposite approach consists of synthetic measures, represented at large scale by the Standardized World Income Inequality Database (SWIID) prepared by Frederick Solt at the University of Iowa. SWIID contains some seven thousand Gini coefficients each for market and disposable income, covering the world almost comprehensively. But the numbers in the SWIID, while consistent, are not actually measures. They are in many cases imputations, based on relationships across time or between countries, so as to fill in gaps in the statistical record. The imputations are in turn based partly on other data sources, including those examined here.

Original, consistent measures are to be found in two significant data sets: the Luxembourg Income Studies (LIS) summary tables, and the Statistics on Income and Living Conditions (SILC) data set of the European Union. LIS is based on an intricate process of international harmonization of existing data sets; SILC is based on European surveys. Both are limited in coverage, in the case of LIS because of the demanding preparation required before each number is published, and because the underlying sources are of higher quality in the richer countries. In the case of SILC the surveys are restricted to the European countries.

A fourth type of data set consists of measures supplied to international agencies, mainly (if not exclusively) by the statistical services of their member states. The World Development Indicators (WDI) of the World Bank have achieved wide use as a standard source of world-wide Gini coefficients, in part because of the authority of the Bank and the easy access afforded in the WDI to inequality measures alongside other indicators of economic and social performance. Meanwhile the OECD has presented a table of inequality measures, concentrated on the OECD member countries, which has also achieved wide recognition for similar reasons.

The final approach to be mentioned here is that of the University of Texas Inequality Project, which in 2005 introduced the Estimated Household Income Inequality (EHII) data set (Galbraith and Kum, 2005). EHII is a panel of estimated Gini coefficients, based on a table of measures and a simple model. The table of measures is called UTIP-UNIDO, consisting of the between-groups components of Theil's T statistics calculated across industrial categories from the Industrial Statistics of the United Nations Industrial Development Organization. UTIP-UNIDO was introduced in 1999 (Galbraith, Lu and Darity, 1999), and was updated most recently by Amin Shams as reported in Galbraith *et al.* 2014.

The calculations behind EHII are based on a regression that shows the very close relationship between inequalities of industrial pay and household income inequalities, as measured in 430 overlapping country-year observations in the original DS data set. Controls specify whether the original DS measure represents inequality of household or persons, of income or expenditure, and whether it is gross or net of tax.[1] Once these are taken into account, the coefficients are very stable and it is possible to use them to produce a large table of estimated Gini coefficients on a consistent gross household income inequality basis, with 3872 observations covering 149 countries from 1963 to 2008.[2]

Thus, although EHII is not a direct measure of income inequality, unlike the SWIID every observation reported is based on a direct measurement for that country in that year. The advantage of the EHII approach is thus dense coverage without loss of degrees of freedom for statistical purposes. In our 2014 paper we show that for at least 40 countries, the EHII measures are mostly plausible as estimates, since they track the movement of other measures well and tend to lie quite close to the (relatively few) direct measures of gross income inequality that exist. However, there are so many different measures of each type of inequality in most countries, with such differences of concept and coverage, that systematic evaluation of data quality against the whole literature is impractical. Comparison to the full literature remains a matter of eyeballs on the page.

The question we take up here is: how does EHII compare to other data sets that each purport – or have been widely taken – to present tables of comparable inequality measures? These are the OECD, the WDI, the SILC, and the LIS. We exclude the WIID on the ground that summary measures of its diverse contents are not very meaningful, and we exclude the SWIID because it is based on multiple imputations and derived, in part, from EHII and the other data sets. The exclusions are not meant to imply criticism. The WIID is invaluable as a resource, and the SWIID appears to be largely consistent with EHII so far as we have observed on a case-by-case basis, allowing for the fact that SWIID estimates market and net income inequality, but not gross income-inequality.

Criteria of comparison

There are two criteria that can be deployed to compare data sets of this type. The first is *coverage*. Given the scope of the data set, in terms of countries and years, how many actual observations are to be found and therefore how dense and complete are the measures? This is a relatively straightforward thing to measure, but there are subtleties, including the question of balance across countries and time. Other things equal, a data set that spans a matrix of countries and years in a fairly even way is preferable to one that has an over-representation of observations in some countries and a dearth in others.

The second major criterion is *accuracy*. Here there is a problem: there is no objective standard of accuracy in this field. As a matter of principle, we cannot simply compare two Gini coefficients for the same country-year observation and declare one to be more accurate than another. We can of course rely on the properties of the data set and our confidence in the underlying techniques, or in the authority of the publisher, but these are often subjective and risky judgments. What we can observe, is the *consistency* of measures across data sets. Where two data sets compiled by different techniques and from different sources broadly arrive at similar measures, our confidence in the joint accuracy of the two data sets will rise. Where overlapping measures diverge, we should be inclined to caution.

1 The ratio of manufacturing employment to population is the one other economic variable in the EHII model.

2 For present purposes, we dropped the estimate for Macau, reducing the country count to 148 and the observation count to 3842.

Standardizing to a common concept

Comparing inequality measures, even with matched country-year observations has a major pitfall. Different data sets may measure different inequality concepts. The prevailing data sets from certain sources (the LIS, SILC and OECD data sets are examples) have concentrated on providing measures of net or disposable household income inequality. This measure will differ from gross household income by the extent to which direct taxes have progressive effect. And that will vary substantially from one country to the next, depending on political history, economic system, and level of development.

In advanced countries gross income inequality is substantially higher than disposable income inequality, but by how much? Wang and Caminada (2011) give estimates of fiscal redistribution based on LIS data, but we find that they do not effectively bridge the gap between EHII and disposable-income inequality estimates in the three data sets. So, for simplicity, we simply adjust the center of the color scheme in the comparison tables by the mean difference between the EHII and the comparison values for the observations that overlap exactly. This enables a reasonable visual assessment of the consistency of the data set pairs; specifically it preserves rank-order differences across countries and discrepancies in particular years.

Then, to free an analysis of differences from bias due to the difference in concepts, we simply add that mean difference to all values in the disposable-income based data set; this value is about 6.7 Gini points for LIS, 6.9 Gini points for the OECD and about 7.5 Gini points for the SILC. This is obviously an artificial procedure; it does not prove that gross income inequality exceeds net income inequality by these amounts in any particular case. Still, the striking consistency across data sets of these mean differences in these mostly-wealthy countries with similar welfare states – despite differences in the specific country-year observations that are matched with EHII – is a reassuring sign.[3]

The WDI poses yet another issue, since it uses a variety of concepts, including both consumption and income measures of inequality, without attempting to standardize them. Differences in consumption and income measures of inequality can be very large. Still, outside the OECD most estimates of fiscal redistribution (gross to net) are very small, and the mean difference between the WDI as a whole and EHII for some 846 matched observations is only about 1.7 Gini points. Mean differences between the WDI and EHII for the various regions are larger, and are reported in Table 2.

Coverage

Table 1 presents coverage ratios for the data sets under study. For each data set, we give the following information: total number of countries in the data set; range of years covered; total number of country-year observations. Then we provide the following pairwise comparisons to EHII: total number of EHII observations in the same range of years for the countries covered; total number of EHII observations for the countries covered, over the full range of years in the EHII data.

3 There are clearly in these data sets some countries – for instance Turkey and the Baltics – where fiscal redistribution is minor and the mean adjustment is too high, but making a better one would require having reliable calibration specific to the country, for which there is no reliable source.

Table 1. Comparison of coverage across data sources

Data Set	Total Observations	Countries Covered	Years Covered	Observations through 2008, countries covered by both EHII and the comparison set.	EHII Observations Matched by Countries and Years Covered	EHII Observations Matched by Countries, 1963-2008
LIS	235	41	1967-2013	206	1319	1415
WDI	1110	149	1978-2013	846	2676	3793
OECD	382	34	1983-2012	286	711	1266
SILC	443	33	1995-2013	288	371	1118
EHII	3872	148	1963-2008	n.a.	n.a.	n.a.

EHII has a strong coverage advantage in both the span of countries and depth of time. Only LIS has comparable historical reach, with some observations as far back as 1967, but with a sixth of the observations over a quarter of the countries. Only the WDI has a comparable breadth of countries,[4] but the WDI starts fifteen years later and has barely a third of EHII's observations even within the years that it covers. The remaining two data sets are much smaller in time and country coverage, as well as number of observations.

A significant advantage of the four comparison data sets is the presence of observations for the most recent five-year interval, 2009-2013. This stems from the ability of these data sets to include new surveys as they are published, and of course to draw them from the same sources. It also reflects the increasing availability of surveys in recent years, as more resources have been devoted to measuring household income inequalities. A pending update to EHII will narrow this gap, to 2010 or 2011, depending on the source data.

In all cases, the EHII all-years coverage, going back to 1963, for the countries covered by the other data sources exceeds their own coverage by large factors. In all cases except the SILC, the EHII coverage is much greater even for the period after the other data set begins. The SILC exception arises because SILC begins only in 1995 and extends to 2013, whereas the EHII data set reaches only to 2008; thus a large fraction of SILC's observations lie outside of EHII's current range. In the other cases, the EHII coverage for the same countries and years is on the order of two to six times as dense.

Ranking the values

Tables in the appendix, prepared using Tableau software, present the comparisons across data sets. In each case, we present a matrix of countries and years, ranked from left to right by the average Gini coefficient in the EHII data set, in ascending order. A color legend tracks the difference in inequality levels, and the evolution of inequality over time in each country, on a consistent scale from blue to red.

4 However they are not the same countries; the WDI includes 28 countries not covered by EHII, most of them small island states, while omitting an almost-equal number of countries that EHII covers.

For the EHII measures, the legend is centered on a Gini score of 40 and so standardized across comparisons to facilitate visual reference. For the other data sets except WDI, the center of the color band is shifted to the left by the mean difference between that data group and EHII in matched observations; the effect is to compensate for the differences in concept between disposable and gross measures of inequality.

The choice of reference case for ranking countries is arbitrary, and may be prejudicial at first glance, since there is a clear gradient from low to high for the reference case, while discrepancies show up in the comparison. This is not meant to imply that the reference case is necessarily correct and that the comparison case is necessarily in error. But it is useful in order to isolate the differences and in some cases, anomalies.

Findings

EHII-LIS.

It is important to note that the Luxembourg Income Study is primarily a database service for researchers in comparative micro-economics. It is not specifically oriented to the production of inequality statistics. However LIS does produce a table of summary Gini coefficients, carefully adjusted to consistency on various concepts, and these have achieved wide acceptance and use.

Tables A1 and A2 present matched coverage of the LIS series for disposable income inequality and the EHII estimates, with the center of the legend for the LIS table shifted down by the mean difference in the overlapping observations, which is 6.7 Gini points. The most obvious point is the far greater coverage of EHII. However, it is also clear that the two data sets are highly compatible; they share rankings and also trends to a high degree. Only China appears far out of place in the cross-country rankings, but this could be because China is ranked on its average EHII value, including low values from earlier years, whereas there is only one LIS value for China, from among the peak years in Chinese inequality.

After adjusting the color scheme, LIS shows higher *relative* values for the US than EHII – a point of difference likely due to the fact that incomes based on capital values are very important in US data, and are not picked up in EHII at all. Since EHII is estimated from pay statistics rather than tax records or income surveys, this suggests that EHII is not based on a model appropriate to capture non-earnings sources of income. However, this appears to be a problem very specific to the United States. Apart from the UK in the OECD comparison below, the same cause does not appear to produce observed discrepancies for other countries. Galbraith *et al*. (2014) explore this issue in depth.

LIS is lower for Greece than EHII – an unexplained divergence – and has an anomalous one-year discrepancy for France. It has one value for Japan that is considerably lower than the EHII estimate, even after adjusting for the mean degree of fiscal redistribution. Otherwise the two data sets broadly coincide. A notable instance of agreement is the high (singleton) LIS value for India – far out of step with the WDI Gini coefficients for that country, but very close to the EHII estimates.

EHII-OECD

Tables A3 and A4 compare the EHII and OECD Gini values for the countries covered in the OECD data set, which is again a collection of measures of disposable income-inequality. Again the major difference is in coverage, which is dense in EHII going back to 1963, but very spotty in the OECD's

own collection before 2004. Again we adjust for the conceptual difference by shifting the center of the legend for the OECD table by the mean difference for overlapping values, in this case 6.9 Gini points.

Once this adjustment is made, the two data sets are largely consistent. As does LIS, the OECD shows higher values for the United States than EHII, and also for the UK, which the other major economy with large amounts of well-recorded capital-asset-based incomes. Korea is one country for which the OECD reports a lower average inequality ranking than EHII, however OECD data for Korea are all very recent, and the ranking of that country in the EHII data set is influenced by high values earlier in its history.

EHII-SILC

The SILC is a data set of disposable income inequalities, and therefore the raw Gini values are comprehensively lower than those in EHII or any other data focused on gross household income. Our comparison is therefore again with the SILC values after adjusting the center of the legend, and therefore the break between blue and red, for the mean difference between the two data sets; in this case the adjustment is 7.5 Gini points. Tables A5 and A6 compare the coverage and values with EHII.

The advantage of SILC, as noted previously, is excellent coverage over the most recent years. The disadvantage is lack of historical depth and of course the narrow focus on the EU and its near neighbors alone. Discrepancies against EHII within this range of countries appear to be subtle; both data sets show Scandinavia and Eastern Europe to be on the low side, and Southern Europe and especially Turkey to be high.

EHII-WDI

The most challenging comparison is with the World Development Indicators inequality measures published by the World Bank. It is not entirely clear what purpose these indicators are intended to serve, since they include a hotch-potch of income and expenditure, and gross and net inequality coefficients. Nevertheless, they are widely-cited for comparative purposes, and often taken as authoritative.

Unlike LIS, the OECD or SILC, the WDI have global reach, covering 149 countries with 1110 total observations over the years 1978 to 2013, including 107 observations for 28 countries not covered in EHII (many small island states, Laos, Vietnam, Palestine, DR Congo, Belarus, Turkmenistan, Tajikistan, Uzbekistan, and Montenegro), and 178 observations for the years 2009-2013. However, also unlike the three other data sets, the WDI does not offer any consistency in the conceptual basis of its coverage: it includes both expenditure and income-inequality measures, and among the income-inequality measures the difference between net and gross is not clearly specified. Therefore it seems to us that an adjustment of the kind made for the other three data sets is not appropriate in this case.

There are 1003 WDI observations for the 147 countries also covered by EHII, not necessarily exactly overlapping in years covered, and 932 observations in the same countries over the span of 1978 – 2008. EHII has 2676 observations for this period, including observations in 26 countries that have no observations reported in the WDI: Afghanistan, the Bahamas, Barbados, Cuba, Cyprus, Eritrea, the GDR, Hong Kong, Korea, Kuwait, Libya, Luxembourg, Malta, Myanmar, New Zealand, Oman, Portugal, Puerto Rico, Qatar, Singapore, Somalia, Taiwan, Tonga, UAE, Yugoslavia, Zimbabwe. WDI thus overlaps EHII in 121 countries.

To make a comparison, we broke the WDI and EHII down into four large regions: the Americas, Eurasia, Asia, and the Middle East and Africa. This permits us to compare coverage and values in manageable portions. Table Two summarizes the coverage for each region.

Table Two
EHII and WDI Coverage by Regions, for Countries in EHII

Regions	WDI Observations 1978-2013	WDI Observations 1978-2008	EHII Observations 1978-2008	EHII Observations 1963-2008	Number of Countries (EHII)	Number of Countries (WDI)	Mean Difference EHII-WDI
Americas	349	293	518	750	29	25	-6.04
Eurasia	382	323	919	1245	44	39	4.75
Asia and Oceania	107	92	496	692	24	16	5.29
Africa and Middle East	165	138	743	1155	50	41	3.98
Total	1003	932	2676	3842	147	121*	1.23

*WDI has in addition 28 countries with 107 total observations that are not covered in EHII; They are excluded above.

A glance at Tables A7 – A14 again shows the dominance of EHII in coverage of every region. However a striking aspect of the comparison is the areas of disagreement over inequality values.[5]

For the Americas, the two data sets are in broad agreement, with just a few anomalous values for the WDI in Jamaica and Peru. Both show only a handful of countries in the moderate-to-low-inequality range typical of the advanced world; in EHII these are Cuba, Canada, and the US, with Uruguay, Costa Rica and Nicaragua during the revolutionary period coming in just above. Uniquely, average values for the Americas in WDI are higher than they are in EHII; otherwise EHII measures tend to exceed those in WDI, especially in Asia where WDI incorporates numerous consumption-inequality measures.

For Eurasia, the main area of disagreement is over the inequality values in the countries of the former USSR. In EHII, outside the Baltics these are all relatively close to each other, and with inequality measures above those for Western Europe, as indicated by their grouping together by rank on the matrix. For the WDI, they are highly diverse, with the Ukraine, Moldova and Kazakhstan showing as much more egalitarian than Russia, almost from the start of the separate existence of those countries. In our view, countries with closely-related economic structures and histories – such as those of the former USSR – likely shared common inequality characteristics at the outset and even now, although they may diverge as time passes.

For Asia, the most important disagreements are over India and Indonesia, as well as Bangladesh, Nepal, Pakistan and Sri Lanka, which are ranked by the WDI as low-inequality countries – with inequality values below those of Australia, in some cases. This is evidently because of the use of consumption inequality measures for those countries, which tend to run some twenty Gini points below the corresponding income measures. All six countries are ranked as high-inequality by EHII, and (as noted above) the EHII measures for India come very close to the singleton income-inequality measure for that country recently published by LIS. It seems obvious that even if the expenditure-inequality

5 The dark background for these charts was necessary to force Tableau to accept the empty columns in countries where EHII had some coverage and the WDI had none.

measures are to be taken at face value, they are not comparable with income-inequality measures and it is essentially meaningless to include them in the same table.

For Africa and the Middle East the two sets of measures are discordant. Notably, the EHII value for South Africa is lower than that reported by WDI and also by almost all studies in the survey literature; this is for reasons we have not been able to explain, though possibly due to South Africa's unique racial history. We are inclined to distrust our estimate. On the other hand, the cluster of oil producers in the high inequality range for Africa and the Middle East, from Qatar and Kuwait to Angola, makes sense to us, as do the (relatively) egalitarian measures for Malta, the Seychelles, Algeria, Mauritius, The Gambia, pre-occupation Iraq and post-revolutionary Iran. In contrast, the WDI measures for Sub-Saharan Africa do not appear to follow any particular logic of regional or economic structure.

Still, whichever source one chooses, for much of Africa there is very little to go on. On one side, surveys are rare and there is no reason for confidence that they were taken in a consistent manner at different times, let alone across different countries; in fact given the mixture of consumption and income inequality measures in this table it is clear they were not. On the other side, the industrial sectors of most African countries are small, and so the foundation of the EHII estimate for this region is comparatively weak. We like our model and the fact that it gives results for many countries that track survey measures very well. For Sub-Saharan Africa, however, it may be best to conclude that while inequality is certainly high, and broadly similar to that found elsewhere in developing countries, all precise measures for the region are open to doubt.

The WDI measures, in addition to being sparse, are sometimes volatile within countries over short periods of time. Notable instances of large jumps in Gini scores in adjacent observations over short intervals – usually less than five years – occur for Angola (59 –> 43), Bolivia (42 –> 54), Central African Republic (44 –> 56), Kenya (57 –> 42), Kyrgyzstan (26 –> 53), Paraguay (41 –> 58), Peru (35 –> 56 over one year!), Senegal (54 –> 41), the Seychelles (43 –> 66!), and Venezuela (53 –> 44). Apart from the collapse of communism in the early 1990s, there is little known basis in the political history of these countries – or almost any country – for such shifts.

For these reasons, in our view the WDI measures of inequality are haphazard. They do not meet the standards set by any of the other comparison data sets, for coverage or comparability.

Divergences and Conclusion

The next task is to compare coefficients in a more systematic way. This we do for LIS, OECD and SILC by adding the mean difference between EHII and each comparison data set to all the values in the latter, so as to erase the influence due to differences in concept. For WDI, since the mean difference overall is small (about 2 Gini points) and since the underlying data set is not built on consistent concepts anyway, we used the raw values. We then compiled a matrix of divergences, for each data set pair, for those observations which are exactly overlapping. The divergences are color coded, from green to yellow to red, according to the absolute value of the difference between the Gini coefficient reported in the comparison data set and in EHII. To present a consistent set of comparisons across tables, the mid-point of the color scale was set in all cases at 7.5 Gini points, which is the mean absolute divergence between EHII and the WDI. Green indicates a divergence of below four Gini points, and dark green, even lower. A table of green-coded values is therefore the desired condition.
Table Three presents summary information on these divergences. Tables A15 – A21 in the appendix provide the coefficients.

Table Three
Summary Measures of Divergence Across Data Sets.

Data Set	Years Covered	Mean Divergence from EHII	Standard Deviation of Divergence from EHII	Volatility* of Gini Coefficient Across Countries	Volatility of EHII Gini Across Matched Countries and Years Covered
LIS	1967-2008	6.34	4.25	1.87	1.55
OECD	1983-2008	6.88	3.36	1.48	1.61
SILC	1995-2008	7.57	3.75	1.56	0.92
WDI: Americas	1978-2008	-4.82	5.97	3.33	1.85
WDI: Eurasia	1978-2008	6.5	6.57	3.04	2.52
WDI: Asia and Oceania	1978-2008	8.51	9.25	3.08	2.86
WDI: Africa and Middle East	1978-2008	4.34	10.34	4.47	2.3
WDI: All EHII Countries	1978-2008	1.65	8.93	3.53	2.37

*Volatility is measured as the mean of country-level standard deviations of Gini coefficients

As a general conclusion, EHII is highly consistent with LIS, OECD and SILC, notwithstanding the difference in concepts measured, or differences in tax systems and welfare states. However its coverage and historical depth is far greater. We take the success in tracking the (fairly reliable) inequality measures for rich country disposable income to be a sign of the general power of the relationship between industrial pay inequality and income inequalities, and therefore an indication that the model underlying EHII is widely applicable around the world.

While the reduction in inequality achieved by passing from gross to net income is important in the wealthy countries that predominate in these data sets, differences in that reduction, both across countries and though time, appear to be second-order. In poorer countries, the reduction in inequality achieved by tax systems is much smaller, and may be effectively nil, and again the differences in this effect are small. Therefore it appears that in both rich and poor countries, taken as separate groups, differences and changes in gross income inequality are the primary source of differences and changes in inequalities generally. However, for the richer countries gross income inequality will overstate the degree of inequality actually experienced in household living standards.

EHII is much less consonant with the WDI, which is the only other data set that covers the entire world, even though the mean divergence is small. Rather, a major source of inconsistency appears to be the mish-mash of different concepts covered in the WDI, an apparent artifact of the Bank's deference to the reporting preferences and survey histories of its member states. The conclusion we draw is that the WDI inequality coefficients are merely erratic, and that the data set should not be used for comparative purposes. Which means, in effect, that it should not be used at all.

Finally, in our view, the case for EHII, as a broadly useful comparative data set with wide geographic and deep historical coverage speaks for itself.

Appendix 1. Gini Coefficients Compared Across Data Sets: EHII, LIS, OECD, SILD and WDI

Figure A1 EHII Inequality Observations for Countries in LIS

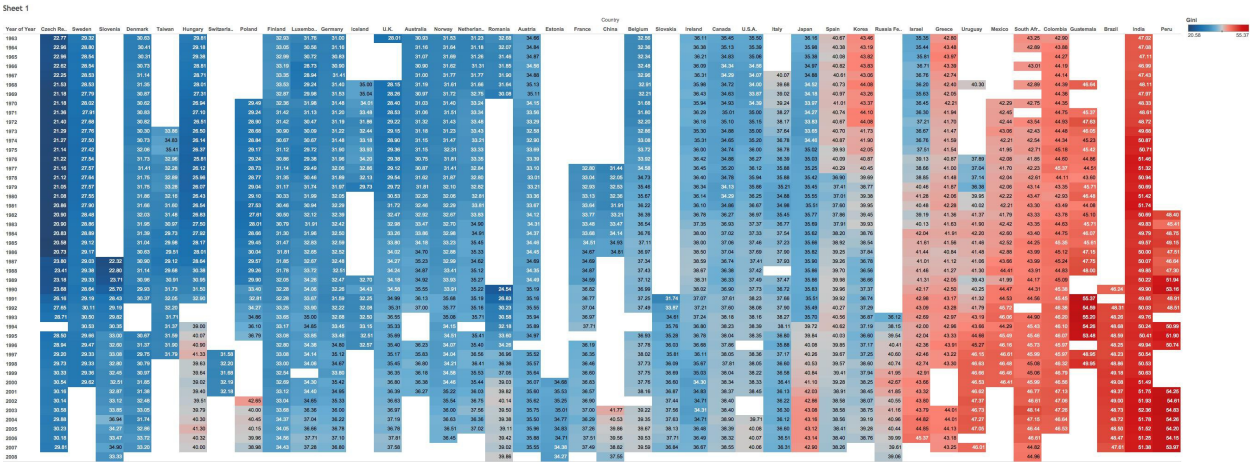


Figure A2 LIS Inequality Observations for countries in EHII

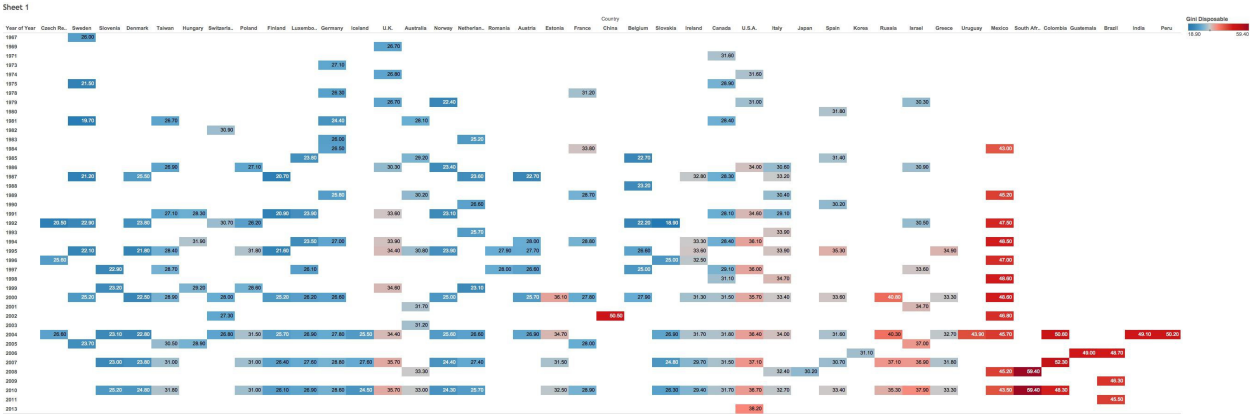


Figure A3 EHII Inequality Observations for OECD Countries

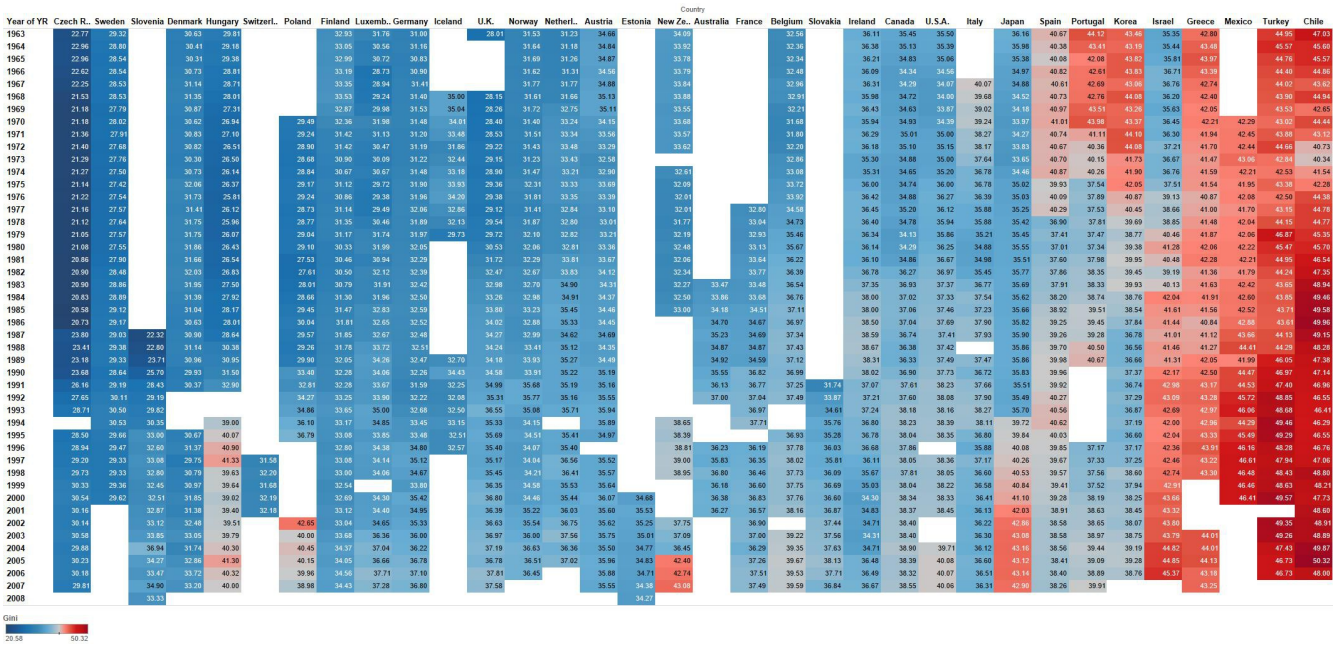


Figure A4 OECD Inequality Observations for OECD Countries

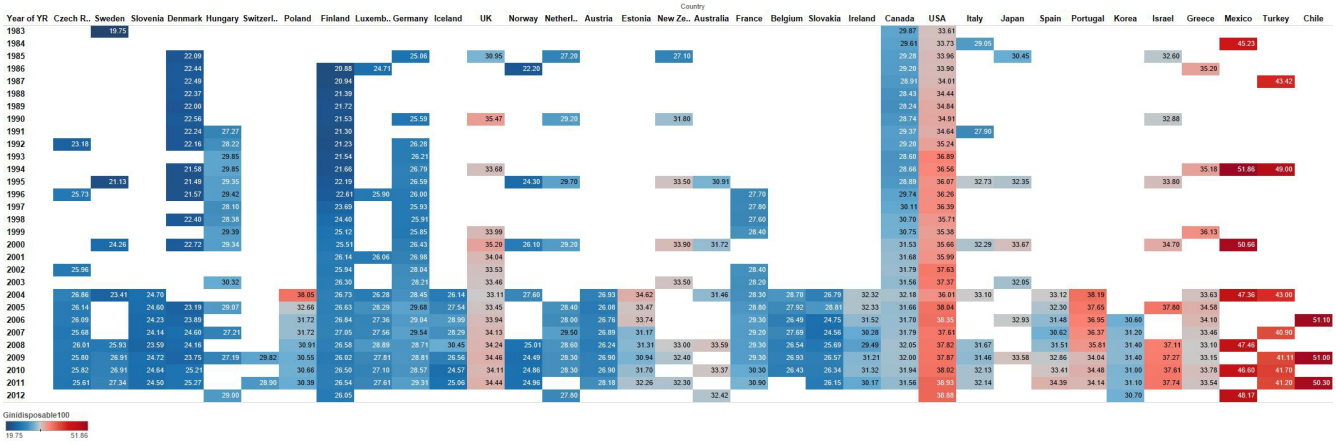


Figure A5 EHII Inequality Observations for EU SILC Countries

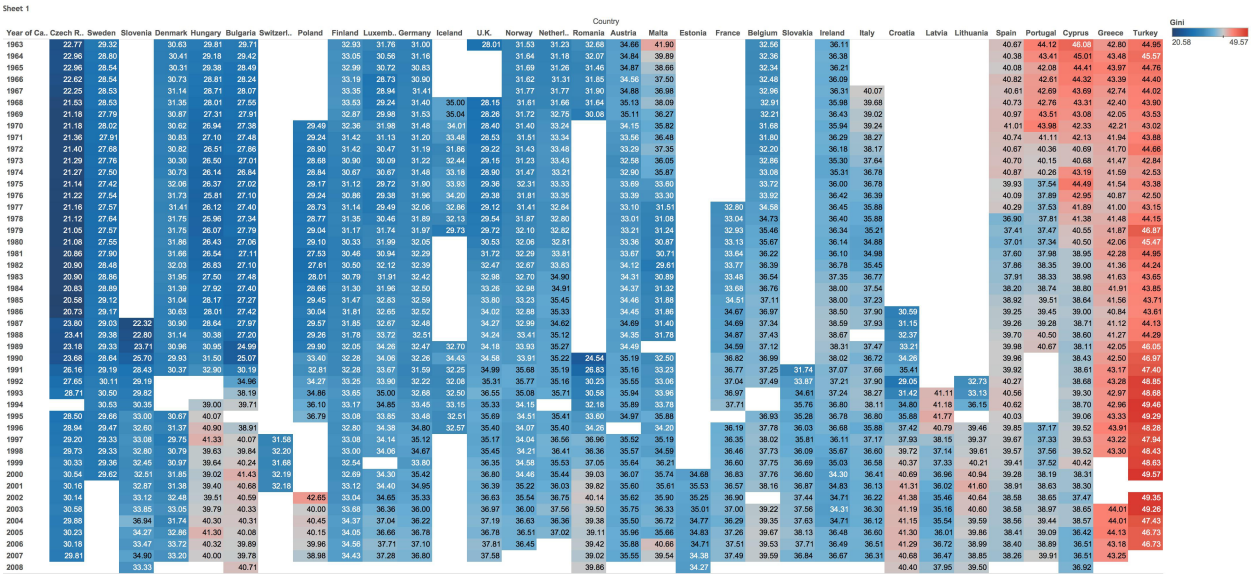


Figure A6 EU SILC Inequality Observations

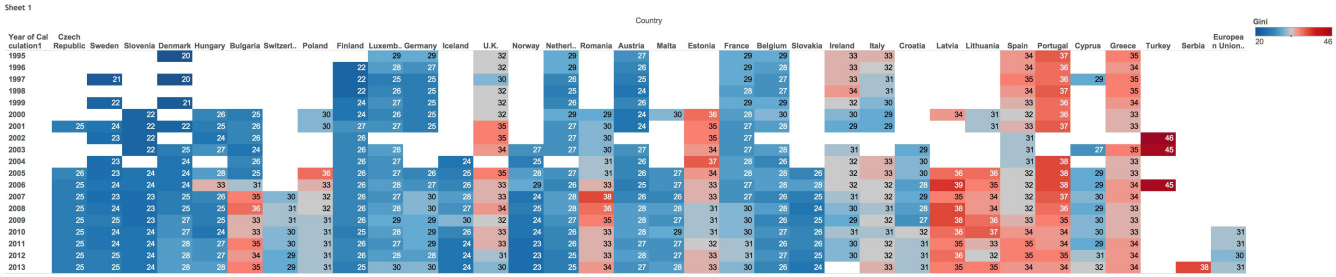
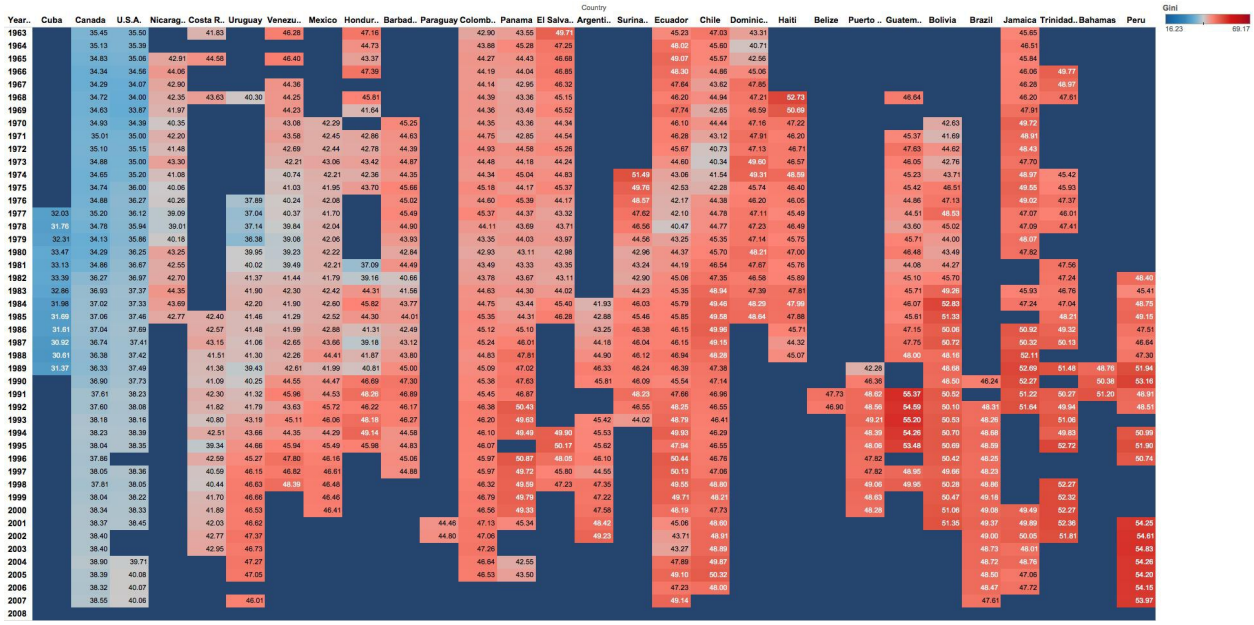


Figure A7 EHII Inequality Observations for Countries in WDI – The Americas



(The dark background in tables A7-A14 was necessary for Tableau to accept the presence of empty columns.)

Figure A8 WDI Inequality Observations for Countries in EHII – The Americas

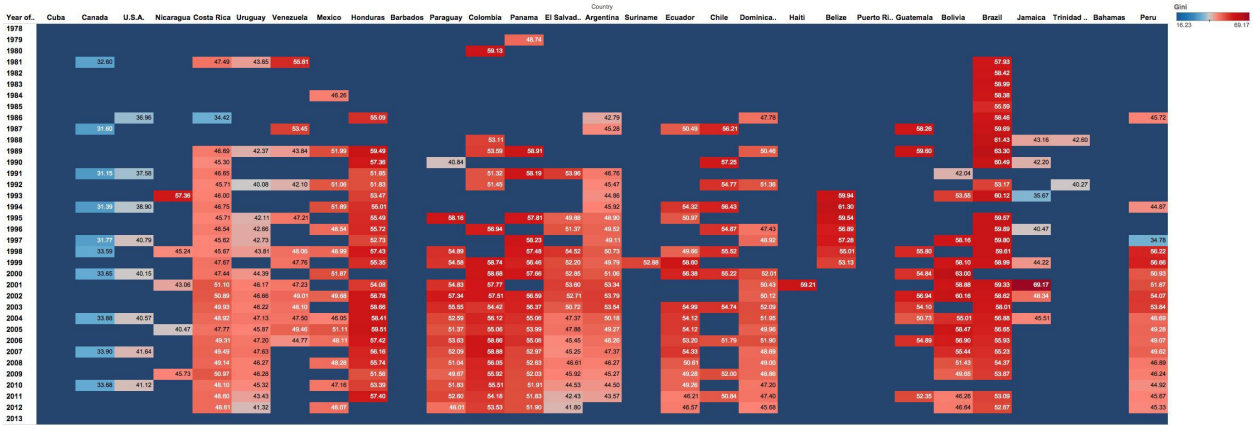


Figure A9 EHII Inequality Observations for Countries in WDI – Europe and Eurasia

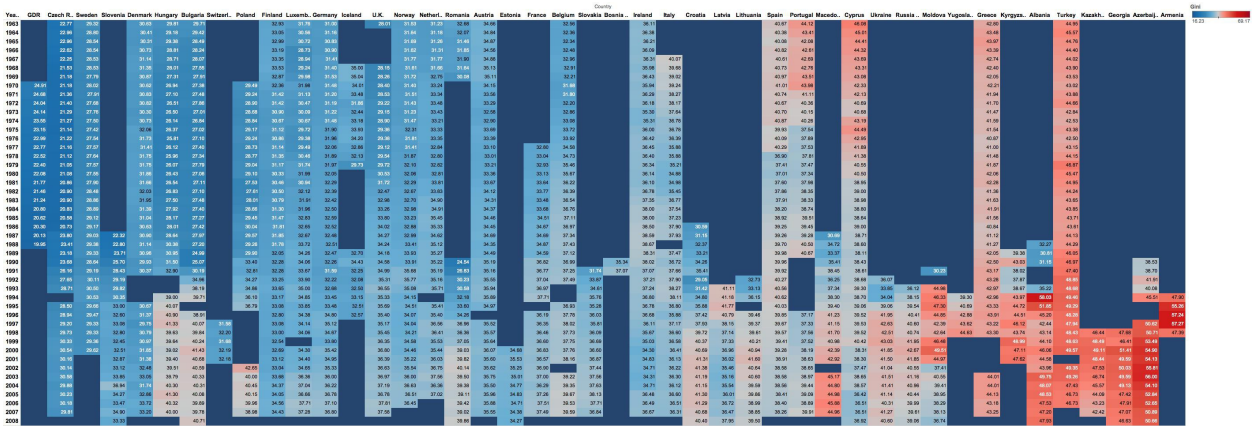


Figure A10 WDI Inequality Observations for Countries in EHII – Europe and Eurasia

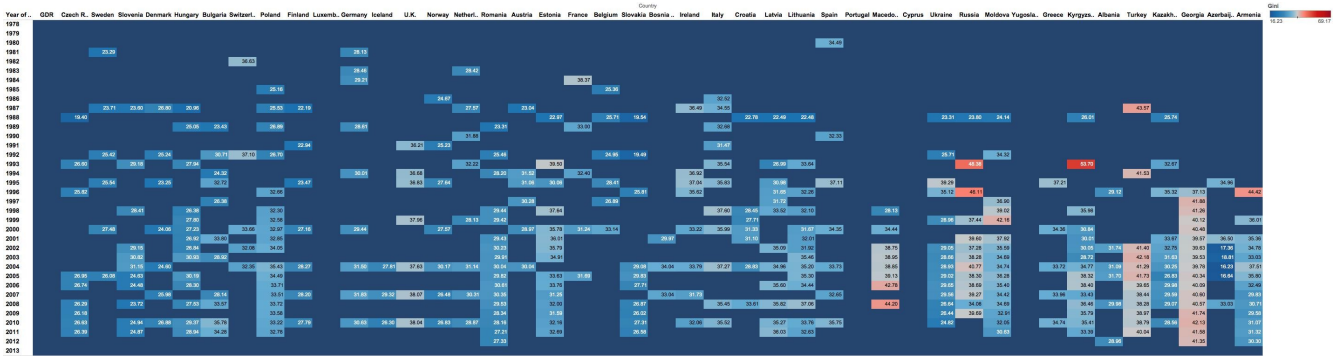


Figure A11 EHII Inequality Observations for Countries in WDI – Asia and Oceania

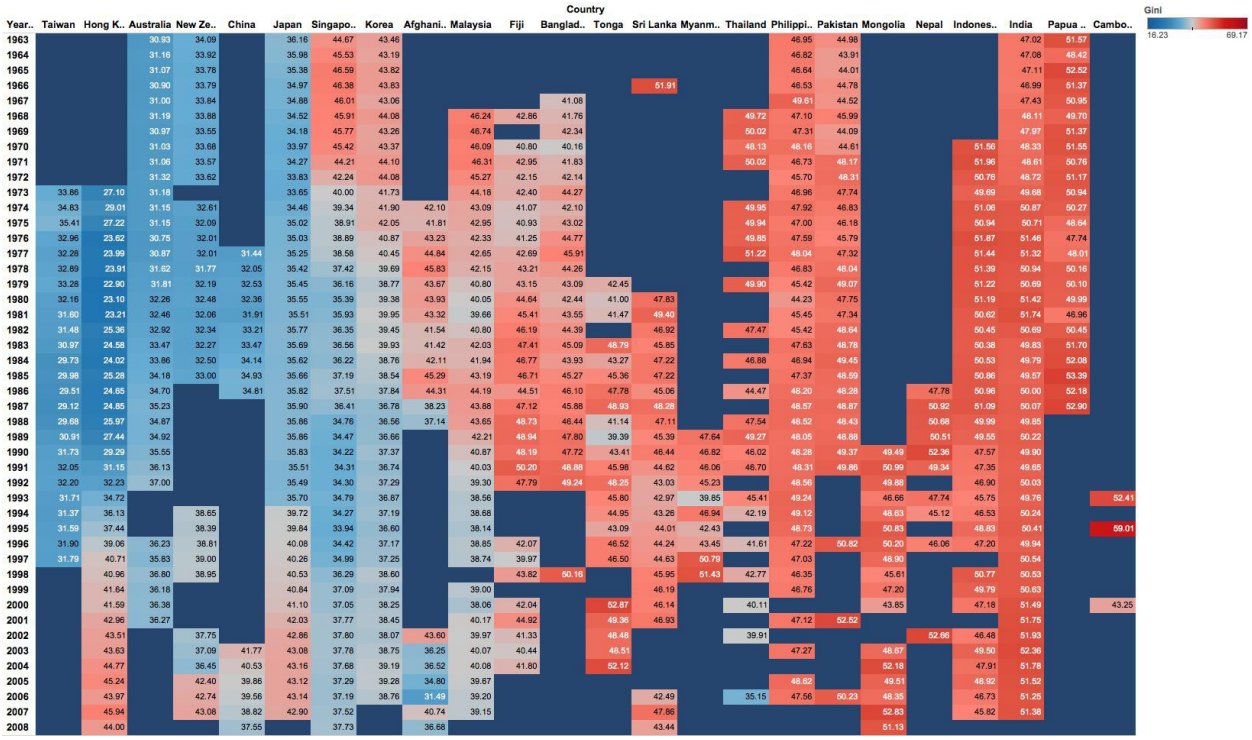


Figure A12 WDI Inequality Observations for Countries in EHII – Asia and Oceania

Figure A13 EHII Inequality Observations for Countries in WDI – Africa and Middle East

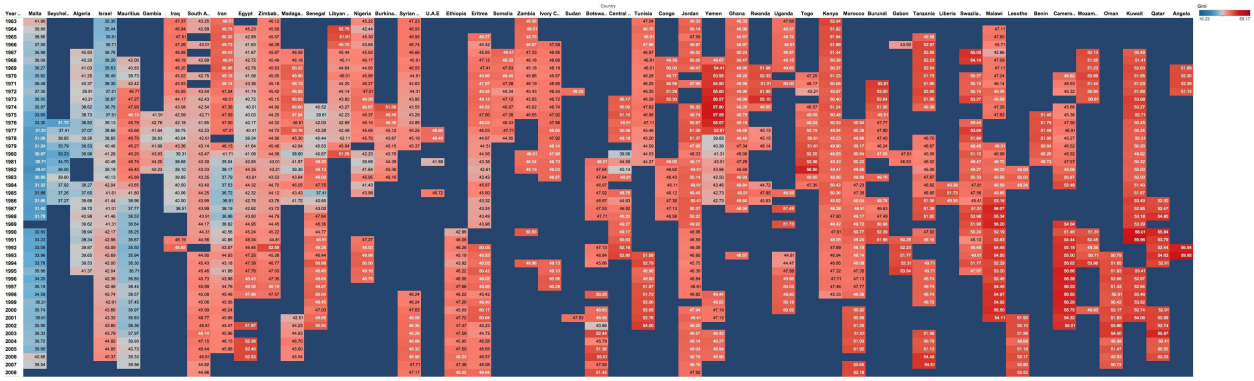


Figure A14 WDI Inequality Observations for Countries in EHII – Africa and Middle East

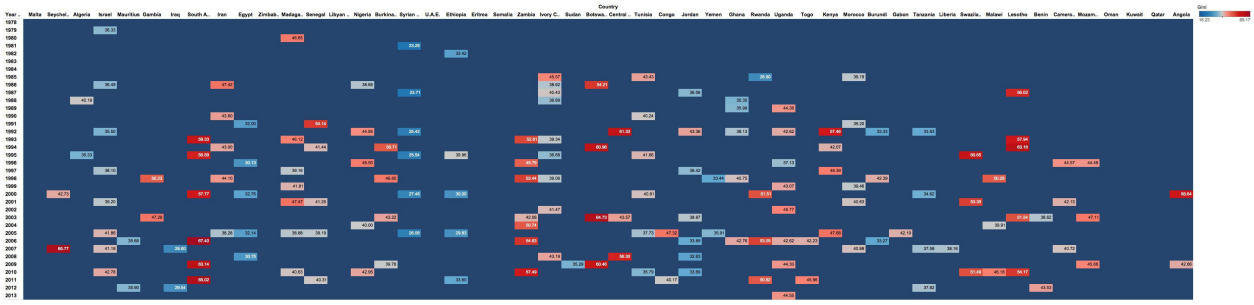


Figure A15 Absolute Differences: EHII – LIS, Mean-adjusted

Figure A16 Absolute Differences: EHII – OECD, mean-adjusted

Figure A17 Absolute Differences: EHII – SILC, mean-adjusted

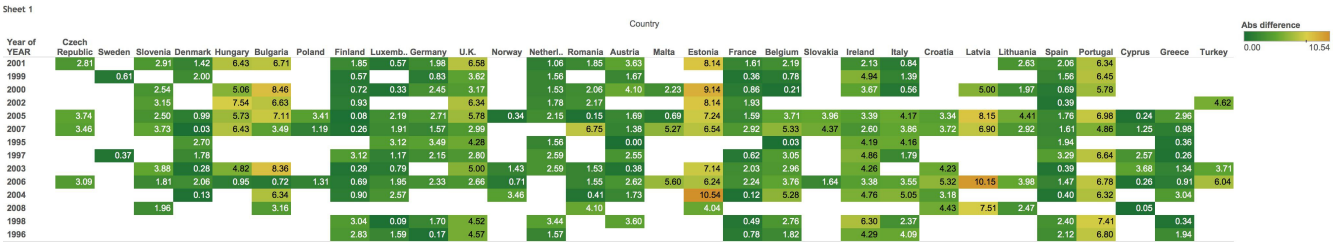


Figure A18 Absolute Differences, EHII – WDI – The Americas

Figure A19 Absolute Differences, EHII – WDI Europe and Eurasia

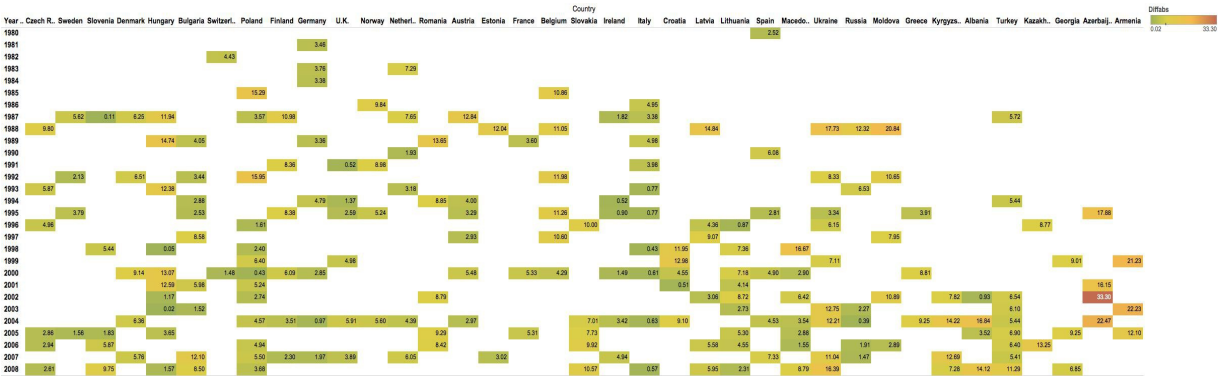


Figure A20 Absolute Differences, EHII – WDI, Asia and Oceania

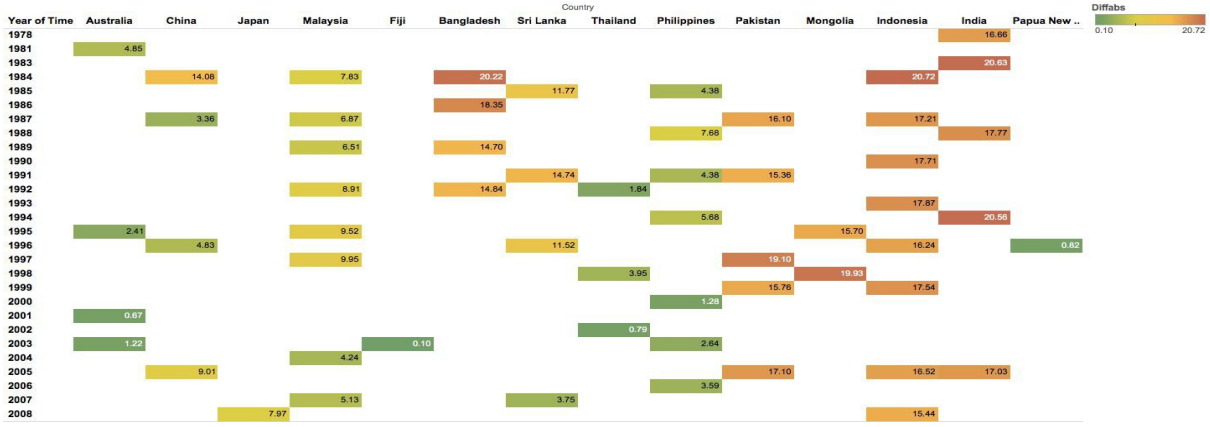
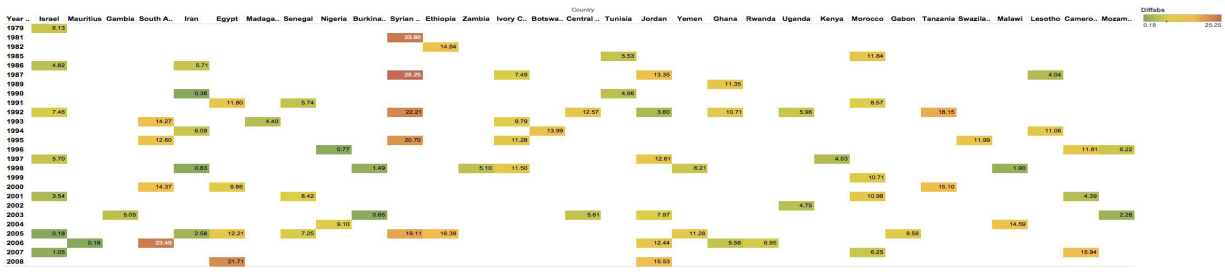


Figure A21 Absolute Differences, EHII – WDI, Africa and Middle East



Appendix 2. Summary of Values by Country: EHII, WDI, LIS, OECD and SILC

		EHII			WDI			LIS			OECD			SILC		
Region	Country	# Obs	Mean	Std dev	# Obs	Mean	Std dev	# Obs	Mean	Std dev	# Obs	Mean	Std dev	# Obs	Mean	Std dev
Asia and Oceania	Afghanistan	22	40.9	3.9												
	American Samoa															
	Australia	36	33.2	2.2	6	33.1	1.1	8	30.9	1.8	6					
	Bangladesh	27	44.6	2.6	8	30.2	3.2									
	Bhutan				3	41.2	4.9									
	Brunei Darussalam															
	Cambodia	3	51.6	7.9	7	35.8	3.1									
	China	16	35.6	3.5	11	36.3	5.9	1	50.5							
	Fiji	32	44.1	2.9	2	44.8	2.8									
	French Polynesia															
	Guam															
	Hong Kong	36	32.9	8.4												
	India	45	50.0	1.5	6	32.7	1.7	1	49.1							
	Indonesia	36	49.5	1.9	10	31.2	2.5									
	Japan	45	37.2	3.1	1	32.1		1	30.2		6	37.0	1.4			
	Kiribati															
	Korea, Rep.	44	39.8	2.5				1	31.1		7	35.3	0.4			
	Korea, Dem. Rep.															
	Lao PDR				5	33.9	2.4									
	Macao															
	Malaysia	39	41.5	2.5	9	46.4	3.4									
	Maldives				2	50.0	17.9									
	Marshall Islands															
	Micronesia, Fed. Sts.				1	61.1										
	Mongolia	17	49.1	2.3	4	33.2	2.6									
	Myanmar (Burma)	10	46.1	3.6												
	Nepal	10	49.3	2.6	4	35.5	6.0									
	New Caledonia	2														
	New Zealand	33	35.1	3.4							8	36.6	2.5			
	Northern Mariana Islands															
	Pakistan	32	47.6	2.2	9	31.4	1.8									
	Palau															
	Papua New Guinea	25	50.6	1.6	1	50.9										
	Philippines	41	47.4	1.1	10	43.5	1.8									
	Samoa															
	Singapore	46	38.5	3.9												
	Solomon Islands															
	Sri Lanka	26	45.9	2.2	6	36.3	3.7									
	Taiwan	25	31.8	1.6				9	28.9	1.9						
	Thailand	23	46.3	4.2	13	43.1	2.2									
	Timor-Leste				1	30.4										
	Tonga	23	45.9	3.6												
	Tuvalu															
	Vanuatu															
	Vietnam				8	36.5	1.3									
Eurasia	Albania	19	44.1	7.2	6	30.4	1.3									
	Andorra															
	Armenia	5	53.0	5.0	14	33.7	4.0									
	Austria	44	34.6	1.0	6	29.2	3.1	6	26.3	1.9	8	30.3	0.7	18	26.3	1.4
	Azerbaijan	17	49.5	6.1	7	24.8	9.5									
	Belarus				17	27.3	2.1									
	Belgium	42	35.8	2.5	6	27.4	3.1	6	24.6	2.3	7	34.5	1.1	18	27.4	1.2
	Bosnia and Herzegovina	2	36.2	1.2	3	32.4	2.1									
	Bulgaria	45	31.9	6.0	11	30.2	4.2							14	30.4	4.7
	Channel Islands															
	Croatia	23	37.1	4.2	7	29.1	3.5							11	29.6	1.5
	Cyprus	46	40.2	2.5										11	29.5	1.4
	Czech Republic	44	24.5	3.9	9	25.7	2.4	3	24.2	3.3	11	29.2	1.1	10	25.1	0.4
	Denmark	42	31.3	0.9	7	25.3	1.4	7	23.6	1.3	23	26.8	1.3	15	24.5	2.7
	Estonia	9	34.8	0.4	15	33.3	3.9	4	33.7	2.1	8	35.6	1.5	14	33.5	1.9
	Faeroe Islands															
	Finland	45	32.3	1.2	7	25.7	2.7	7	23.8	2.6	27	29.5	2.8	18	25.1	1.6
	France	30	35.6	1.6	5	33.3	2.9	7	29.6	2.2	16	30.7	1.0	19	28.6	1.3
	GDR	19	22.3	1.6												
	Georgia	11	48.4	1.6	17	40.5	1.2									
	Germany	45	32.9	1.8	9	29.8	1.3	11	26.8	1.3	22	31.4	1.7	16	27.5	2.1
	Greece	41	42.4	1.0	5	34.8	1.4	5	33.2	1.1	11	35.3	1.0	18	33.9	0.8
	Greenland															
	Hungary	43	31.9	5.9	14	27.5	2.4	4	29.6	1.6	15	30.8	1.1	13	26.5	2.4
	Iceland	20	33.0	1.2	3	27.8	1.5	3	25.9	1.6	8	31.3	2.3	10	25.8	2.0
	Ireland	45	36.4	1.1	8	34.6	2.2	8	31.8	1.6	8	34.8	1.3	17	31.2	1.5
	Isle of Man															
	Italy	40	37.0	1.2	11	34.9	2.0	11	32.6	1.8	9	32.7	1.8	17	31.5	1.2
	Kazakhstan	10	46.2	2.3	12	30.5	2.8									
	Kosovo															
	Kyrgyzstan	13	43.4	3.8	15	34.8	6.4									
	Latvia	16	37.7	2.2	12	32.5	4.1							10	36.1	1.4
	Liechtenstein															
	Lithuania	16	38.8	2.6	14	32.9	3.4							11	34.0	2.1
	Luxembourg	44	32.6	2.4				8	25.6	1.6	11	30.8	1.4	18	27.5	1.3
	Macedonia	20	40.0	4.4	8	38.2	5.0									
	Moldova	17	41.7	4.9	16	34.7	3.9									
	Monaco															
	Montenegro				7	30.0	0.8									
	Netherlands	42	34.3	1.7	8	29.8	1.8	8	25.5	1.5	11	33.0	0.9	18	26.8	1.3
	Norway	44	33.3	1.7	7	26.9	1.8	8	24.0	1.0	8	28.8	1.8	11	24.9	2.3
	Poland	32	32.2	4.7	18	31.8	3.2	7	29.6	2.3	8	32.6	2.6	11	31.7	1.6
	Portugal	39	39.7	2.1							8	37.4	1.7	17	36.1	1.3
	Romania	26	34.6	4.5	17	28.6	1.9	2	28.0	0.1				14	32.6	2.6
	Russia	16	40.2	1.6	13	38.6	5.8	4	38.4	2.6						
	San Marino															
	Serbia				9	30.8	2.1							1	38.0	
	Slovakia	17	36.2	1.6	10	25.8	3.6	5	24.4	3.2	8	29.7	1.5	9	25.4	1.3
	Slovenia	22	31.0	4.0	11	26.8	2.9	5	23.5	1.0	8	25.4	0.4	13	23.1	0.9
	Spain	45	39.5	1.1	7	34.3	1.7	8	32.3	1.7	8	33.9	1.3	19	33.0	1.4
	Sweden	38	28.7	0.9	6	25.3	1.5	8	22.8	2.1	9	28.5	3.5	14	23.6	1.1
	Switzerland	5	32.0	0.3	5	34.4	2.4	5	28.7	1.9	2	29.9	0.7	7	29.8	1.0
	Tajikistan				5	31.8	1.6									
	Turkey	43	45.7	2.3	12	40.5	1.7				7	43.9	2.9	3	45.3	0.6
	Turkmenistan				3	34.2	7.3									
	U.K.	41	33.1	3.3	7	37.3	0.8	11	32.1	3.7	16	36.9	1.2	18	32.6	1.4
	Ukraine	17	40.2	2.8	14	28.9	4.1									
	Uzbekistan				4	34.6	8.4									
	Yugoslavia	5	42.2	2.2												
	Algeria	26	39.3	1.4	2	37.8	3.4									
Africa and Middle East	Angola	6	54.0	3.5	2	50.7	11.3									
	Bahrain															
	Benin	7	50.9	1.1	2	41.1	3.5									
	Botswana	21	48.1	2.7	4	60.1	4.4									
	Burkina Faso	10	46.3	2.6	4	45.1	4.7									
	Burundi	17	49.7	2.4	3	36.3	5.2									
	Cabo Verde				2	47.2	4.7									
	Cameroon	28	51.3	4.3	3	42.5	1.9									
	Central African Republic	19	48.1	3.7	3	53.7	9.2									
	Chad				2	41.5	2.5									
	Comoros				1	64.3										
	Congo	14	48.5	2.1	2	43.7	5.1									
	Congo, Dem. Rep.				1	44.4										
	Djibouti				1	40.0										
	Egypt	39	44.5	3.7	5	31.6	1.1									
	Equatorial Guinea															
	Eritrea	42	46.7	2.2												
	Ethiopia	19	46.6	1.5	5	33.2	4.1									
	Gabon	8	49.8	3.4	1	42.2										
	Gambia	8	42.3	1.4	2	48.8	2.1									
	Ghana	26	48.8	1.1	5	38.6	3.1									
	Guinea				5	41.0	5.1									
	Guinea-Bissau				2	41.7	8.7									
	Iran	42	44.4	4.3	5	43.3	3.3									
	Iraq	27	43.4	3.3	2	29.1	0.7									
	Israel	44	40.3	3.0	8	38.9	2.8	8	34.0	3.1	9	41.8	2.4			
	Ivory Coast	22	47.9	1.3	9	40.1	2.9									
	Jordan	42	48.7	1.5	7	36.4	3.7									
	Kenya	36	49.5	1.7	4	48.4	6.5									
	Kuwait	32	52.4	2.2												
	Lebanon															
	Lesotho	11	50.7	1.7	5	56.6	4.4									
	Liberia	3	50.2	1.4	1	38.2										
	Libyan Arab Jamahiriya	17	45.8	3.5												
	Madagascar	26	45.7	3.3	7	43.0	3.7									
	Malawi	35	50.6	3.6	3	45.5	5.2									
	Mali				4	40.6	7.3									
	Malta	44	34.8	3.0										10	27.7	1.0
	Mauritania				6	42.0	4.5									
	Mauritius	40	40.9	4.3	2	35.8	0.2									
	Morocco	31	49.7	1.6	5	39.9	0.8									
	Mozambique	13	51.4	1.2	3	45.8	1.3									
	Namibia				3	66.5	6.9									
	Niger				5	37.9	6.0									
	Nigeria	28	45.8	2.6	5	42.6	3.3									
	Oman	15	51.5	1.3												
	Qatar	15	53.1	1.6												
	Rwanda	12	48.8	3.5	4	46.1	11.5									
	Sao Tome and Principe				2	42.3	12.0									
	Saudi Arabia															
	Senegal	29	45.7	4.0	5	43.3	6.1									
	Seychelles	11	36.3	2.4	2	54.3	16.3									
	Sierra Leone				2	37.4	2.9									
	Somalia	12	47.2	1.4												
	South Africa	41	44.3	1.5	6	61.5	4.3	2	59.4							
	South Sudan															
	Sudan	2	48.0	0.1	1	35.3										
	Swaziland	26	50.3	2.5	3	55.2	4.8									
	Syrian Arab Republic	28	46.4	1.7	6	25.3	1.5									
	Tanzania	34	50.1	2.5	4	36.0	2.0									
	Togo	14	49.4	3.4	2	44.1	2.6									
	Tunisia	29	48.3	2.8	6	39.9	2.8									
	U.A.E.	4	46.4	3.1												
	Uganda	19	49.0	1.9	8	43.1	2.6									
	West Bank and Gaza				3	35.7	2.6									
	Yemen	27	48.7	5.7	2	34.7	1.7									
	Zambia	18	47.6	1.7	7	51.5	4.9									
	Zimbabwe	36	45.4	1.7												
	Antigua and Barbuda															
Americas	Argentina	17	45.7	2.0	23	48.1	3.2									
	Aruba															
	Bahamas	3	50.1	1.2												
	Barbados	28	44.5	1.5												
	Belize	2	47.3	0.6	7	57.6	2.9									
	Bermuda															
	Bolivia	32	48.0	3.2	15	54.2	5.9									
	Brazil	17	48.5	0.7	28	57.8	2.7	3	46.8	1.7						
	Canada	45	36.5	1.6	10	32.7	1.1	12	30.0	1.6	29	34.2	1.5			
	Cayman Islands															
	Chile	44	46.4	2.5	11	54.5	2.1				3	52.0	0.4			
	Colombia	43	45.1	1.1	20	55.8	2.5	3	50.4	2.0						
	Costa Rica	22	42.0	1.2	26	47.3	3.2									
	Cuba	13	32.1	0.9												
	Curacao															
	Dominica															
	Dominican Republic	23	46.7	2.1	18	49.5	1.9									
	Ecuador	45	46.4	2.4	16	52.1	3.5									
	El Salvador	28	45.6	2.0	18	48.8	4.1									
	Grenada															
	Guatemala	26	47.8	3.7	9	55.3	2.8	1	49.0							
	Guyana				2	48.0	5.0									
	Haiti	21	47.0	1.9	1	59.2										
	Honduras	26	44.0	3.1	23	55.7	2.5									
	Jamaica	34	48.6	2.0	8	46.1	10.0									
	Mexico	31	43.6	1.7	14	49.2	2.1	11	46.3	2.0	7	49.3	2.4			
	Nicaragua	21	41.9	1.6	5	46.4	6.5									
	Panama	40	45.5	2.5	19	54.9	2.9									
	Paraguay	2	44.6	0.2	16	52.4	4.1									
	Peru	21	50.9	3.1	18	48.6	5.2	1	50.2							
	Puerto Rico	12	47.8	1.9												
	Sint Maarten (Dutch part)															
	St. Kitts and Nevis															
	St. Lucia				1	42.6										
	St. Martin (French part)															
	St. Vincent and the Grenadines															
	Suriname	20	46.2	2.2	1	52.9										
	Trinidad and Tobago	26	49.3	2.3	2	41.4	1.6									
	Turke and Caicos Islands															
	U.S.A.	42	36.9	1.7	8	39.7	1.7	11	35.2	2.2	30	40.7	1.8			
	Uruguay	32	42.7	3.4	20	44.6	2.2	1	43.9							
	Venezuela	34	43.1	2.5	13	48.0	3.6									
	Virgin Islands															
Total		3842 Gross Income			1110 Mixed			235 Net Income			382 Net Income			443 Net Income		

References:

Klaus Deininger and Lyn Squire. 1996. "A New Data Set Measuring Income Inequality." *World Bank Economic Review* 10(3):565-591.

—. 1998. "New Ways of Looking at Old Issues: Inequality and Growth." *Journal of Development Economics* 57(2):259-287.

James K. Galbraith, Lu Jiaqing and Willliam A. Darity, jr. "Measuring the Evolution of Inequality in the Global Economy," January 1999. UTIP Working Paper No. 7.

---. Hyunsub Kum. 2003. "Inequality and Economic Growth: A Global View Based on Measures of Pay." *CESifo Economic Studies* 49(4):527-556.

—. 2005. "Estimating the Inequality of Household Incomes: A Statistical Approach to the Creation of a Dense and Consistent Global Data Set." *Review of Income and Wealth* (1):115-143.

--- . Amin Shams, Béatrice Halbach, Aleksandra Malinowska and Wenjie Zhang (2014). "The UTIP Global Inequality Data Sets 1963-2008: Updates, Revisions and Quality Checks" *UTIP Working Paper No. 68 Forthcoming, United Nations University Working Paper series.*

Luxembourg Income Studies: http://www.lisdatacenter.org/ Accessed December 19, 2014.

Thomas Piketty *et al. The World Top Incomes Database*: http://topincomes.parisschoolofeconomics.eu/ Accessed December 19, 2014.

Thomas Piketty, 2014. *Capital in the Twenty-first Century.* Cambridge: Harvard University Press.

Frederick Solt, *The Standardized World Income Inequality Dataset:* http://myweb.uiowa.edu/fsolt/swiid/swiid.html Accessed December 19, 2014.

World Bank. 2007, *World Development Indicators Online*. http://www.worldbank.org/.

Wang, Chen & Caminada, Koen, 2011. "Disentangling income inequality and the redistributive effect of social transfers and taxes in 36 LIS countries," LIS Working Paper No. 567.