Historical Inequality Series on WID.world - Update

Lucas Chancel
Rowaida Moshrif
Thomas Piketty
Silas Xuereb

February 2023

Historical Inequality Series in WID.world: 2022 updates

Lucas Chancel   Rowaida Moshrif   Thomas Piketty   Silas Xuereb

June 10, 2022

Abstract

The purpose of this technical note is to describe the construction of WID.world's long-run per-adult income inequality series and adjustments made to ensure consistency with our long-run per-capita series and existing per-adult series. These estimates are mainly based on the assumptions and methodology outlined in Chancel and Piketty 2021. For further details, please refer to this paper and its associated appendices. In this document, we detail how we built on these estimates to construct similar income inequality series when income is allocated among adults (20 years of age and older) rather than among all citizens. This entailed a three-step process: 1) estimating adult population totals for each country and region, 2) estimating per-adult income shares, and 3) ensuring maximum consistency between these new estimates and previous estimates, and potentially adjusting earlier estimates. We also describe adjustments made to historical pretax income inequality series that are based on historical fiscal income series.

1   Introduction

This technical note explains the latest update of WID.world to include historical per-adult pretax income distribution series, average national income series, total population and adult population series. Since December 2021, WID.world included per-capita income distribution series for 24 countries, 8 sub-regions and 9 regions with data points for each decade from 1900-2020 and every 30 years from 1820-1880, as described in Chancel and Piketty (2021). We have now added per-adult income distribution series for each of these area-years that were not previously covered by a per-adult income distribution series in WID.world (all area-years were previously covered from 1980 onwards while older estimates were available for a subset of countries). This work required the estimation of adult population totals for each area-year that adult population estimates were not previously available, described in section 2. The estimation of the per-adult pretax income distribution series is described in section 3.

We also systematically checked for consistencies between the new per-adult series, the per-capita series, and the existing per-adult series on WID.world. In the course of this work, we identified several updates and adjustments that needed to be made to each series which are described in section 4. We also checked for discrepancies between our fiscal income series and pretax income series which resulted in revisions to some historical pretax income share estimates described in section 5. This note describes the estimation of the missing adult population totals, the construction of the per-adult income distribution series and the developments that were made to both series to ensure maximum consistency between them.

2   Adult population estimates

We used linear regression models to predict adult population shares for 213 countries/sub-region-years for which we had no available estimates of adult population shares. The model was estimated based on 13,523 country-years for which we have data on adult population, total population, and average national income in WID.world.

We tested a wide range of models that included a subset of the following variables: total population, log of average annual income, log of average annual income squared, year, year squared, period (fixed effects for groups of years), country fixed effects (countries within sub-regions were given a single fixed effect), region fixed effects, interactions between time and country or region.

Model selection was conducted using a two-step process. First, models that resulted in low R^2 values on the estimation sample were excluded. On this basis, we selected four models that included period fixed effects over models with simple linear or quadratic year trends on both metrics. This step also limited potential models to those that included interactions between country or region and periods. Models that were excluded in this step also resulted in unrealistically steep increases in the adult population rate during the 19th century, providing further confirmation of our model selection strategy.

In the second step, we used cross-validation to select the model that minimized average prediction error at the regional level. We removed data for one region at a time for the

period 1960-1989 and estimated adult population shares using each model on the remaining subsample. This process selected a model that included country fixed effects, period fixed effects and their interactions, total population, log average annual income, and log average annual income squared (average prediction error for regional adult shares: 2.6 percentage points). A similar model that included region (instead of country) fixed effects had very similar prediction error but resulted in an unrealistically steep jump in adult population share from 1940 to 1950.

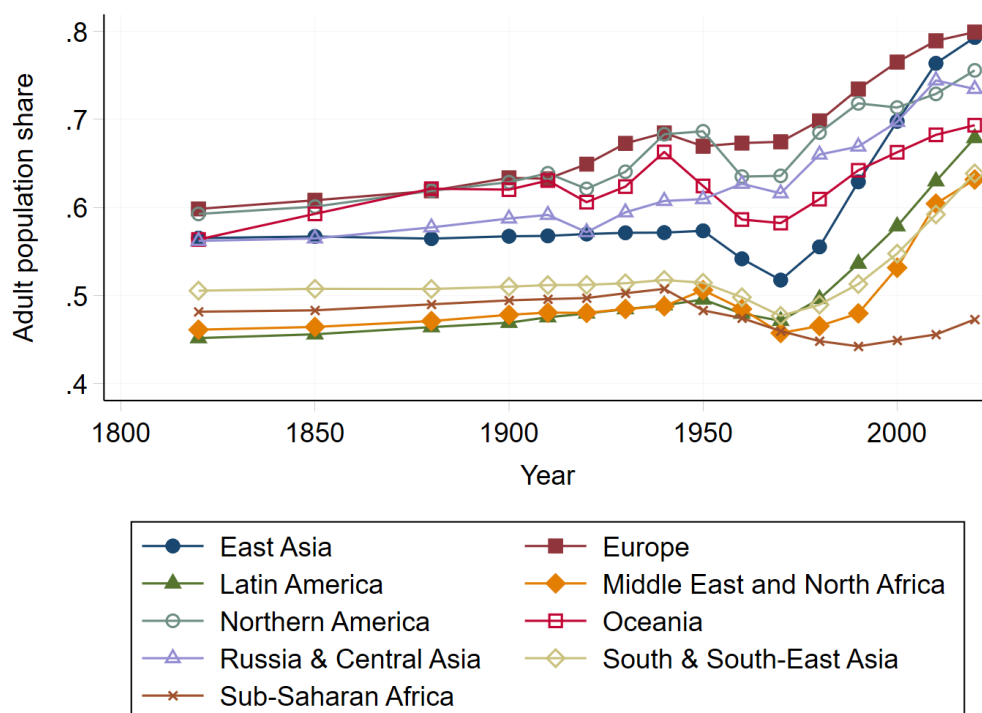Figure 1: Estimated adult population shares by region, 1820-2020.

Figure 1 presents our estimated adult population shares by region over time, combining our predictions with the previously available adult population data. Our estimates align with other research which suggests that fertility rates were relatively stable in non-Western countries during 1820-1950 Ajus, 2010. However, although our estimates are plausible at the regional-level, we acknowledge that further research is needed to estimate precise adult population totals at the country-level in the 19th century.

3   Per-adult historical income inequality estimates

Using the same assumptions as described in Chancel and Piketty 2021, we constructed estimates of the income distribution for each year among 1820, 1850, 1880, 1900, 1910, 1920, 1930, 1940, 1950, 1960 and 1970 that is not currently covered by a per-adult income share estimate in WID.world for the 24 countries and 8 sub-regions covered by a per-capita income share estimate. These estimates thus assume that family incomes are unrelated to family size and thus the distribution of income at the country- and sub-region level is the same in both the per-capita and per-adult series. We plan to relax this assumption in future research. However, our new adult population estimates allowed us to calculate the

average income per adult in each area-year. Thus, average incomes vary between the two distributions based on whether income is distributed to each individual or to each adult.

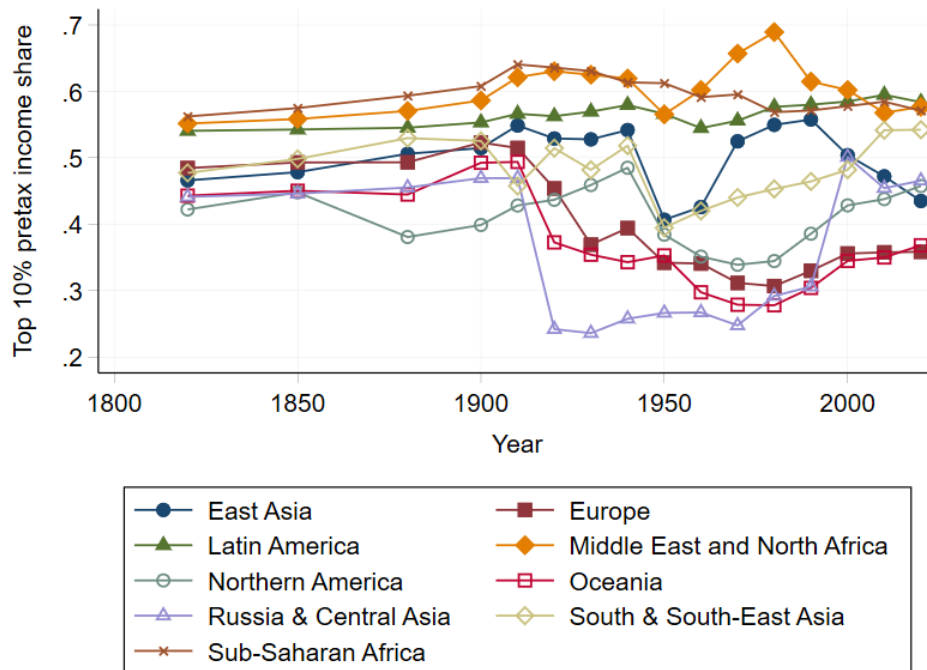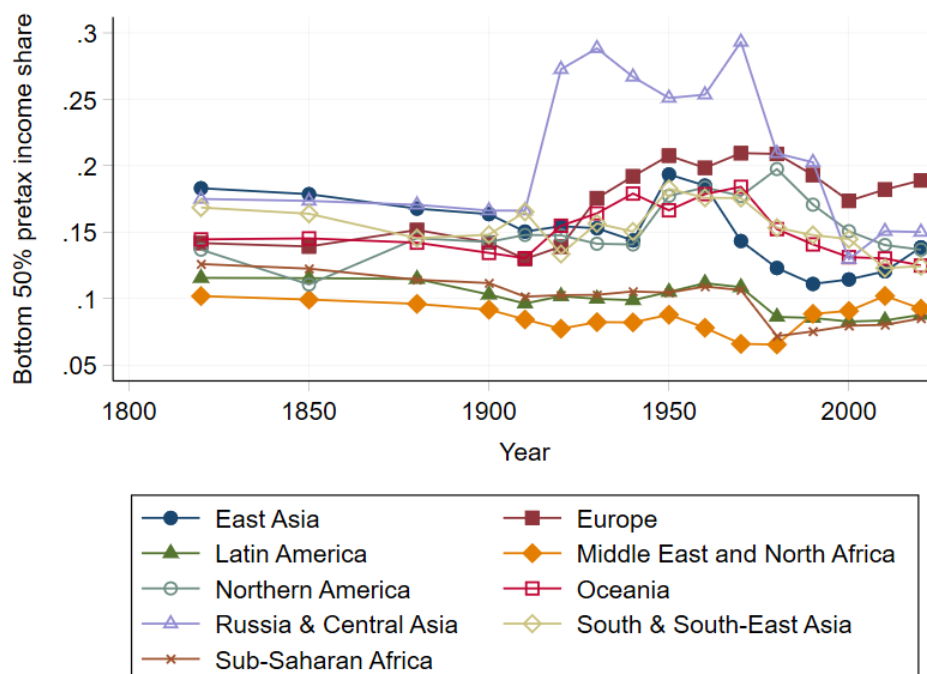Figure 2: Estimated per-adult top 10% income shares by region, 1820-2020.

Figure 3: Estimated per-adult bottom 50% income shares by region, 1820-2020.

To obtain estimates of the full distribution of income in each country- or sub-region-year, we used the generalized Pareto interpolation procedure developed by Blanchet and colleagues 2022. Average incomes for each group were estimated based on the income distribution,

4

average income per capita and adult population totals estimated above. The generalized Pareto interpolation procedure was also used to estimate the distribution of income per adult at the regional level by aggregating the income distributions of each country and sub-region within the region.

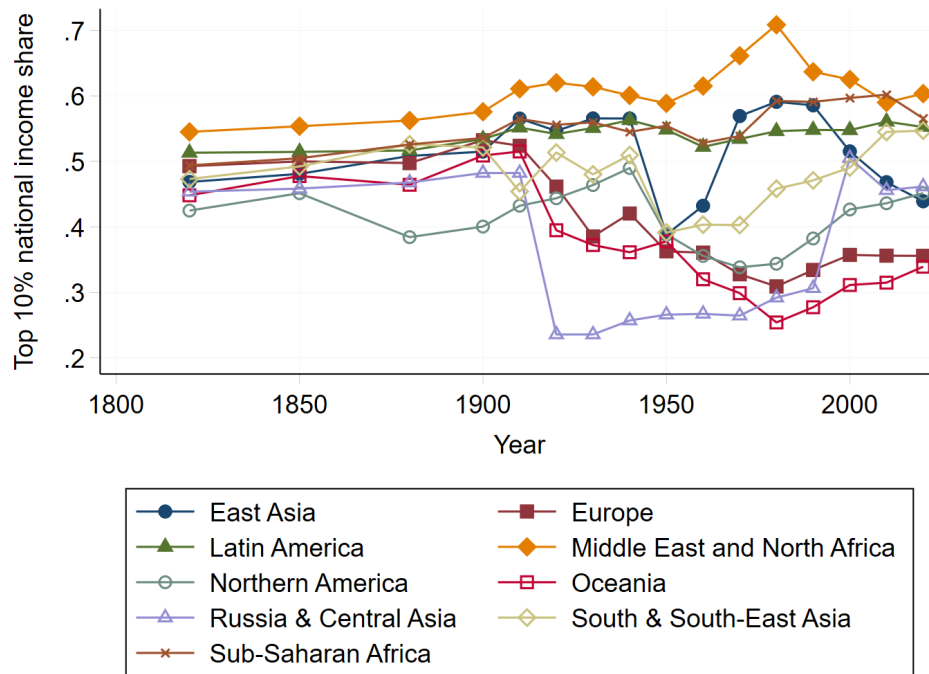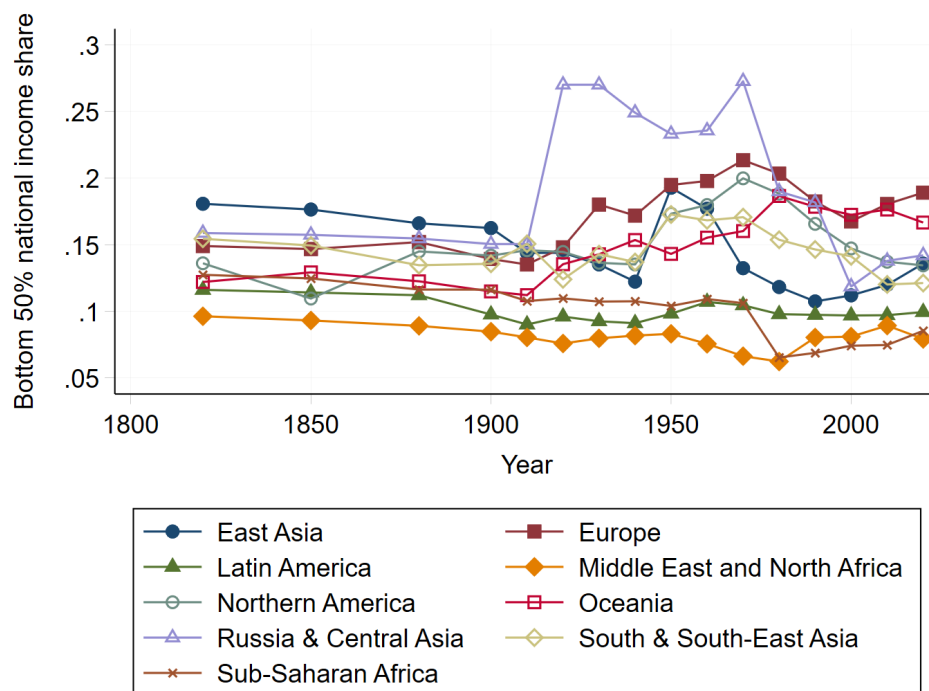Figure 4: Estimated per-capita top 10% income shares by region, 1820-2020.

Figure 5: Estimated per-capita bottom 50% income shares by region, 1820-2020.

New distributions were calculated in this manner for 385 country/sub-region-years. Further-

more, for 32 country-years in which top income shares were already available in WID.world, the lower part of the income distribution was estimated. Figures 2 and 3 display our updated estimates of the top 10% and bottom 50% national per-adult income shares at the regional level over time. Figures 4 and 5 display the same income shares for the per-capita series. Even though country-level per-capita and per-adult pretax income share estimates are identical, estimates may differ at the regional level due to different adult population proportions in countries within a region.

4   Adjustments to available WID.world historical series

Next, we checked for any discrepancy between these new historical per-adult income share estimates, the previously constructed per-capita income share estimates presented in Chancel and Piketty 2021, and the existing per-adult income share estimates available on WID.world as of February 2022.

To find discrepancies between these three series, we flagged an income share (bottom 50%, middle 40%, top 10%, top 1% or top 0.1%) in a given country-year as a discrepancy if it met any of the following criteria:

1. The gap between either historical series and the current WID series was larger than 10 percentage points

2. There was a 75% gap between either historical series and the current WID series for top 1% and top 0.1% shares (since these values are so small they rarely reach the 10pp threshold)

3. There was a 10 percentage point gap between two consecutive values in the combined per-adult series (including all data points from the historical per-adult series and the per-adult series available on WID.world)

4. There was a 75% gap between two consecutive values in the combined per-adult series for the top 1% or top 0.1% share

We manually inspected each discrepancy and identified whether further action was needed.

4.1   Adjustment and updates made to WID.world historical series published in December 2021

Historical per-capita series and the historical per-adult series are based on the same distributional assumptions so we use the term *historical series* to refer to both series in this section. The following developments and adjustments were made:

1. We slightly reviewed Argentina's top historical income shares, which is now more consistent with available historical series from tax data.

2. We revised the United Kingdom's 1940 and 1950 top 10% shares to provide estimates more consistent with our knowledge of distributions in similar countries at the time (see Section 2 and Appendix Table 8 of Chancel and Piketty, 2021 for references of available studies and countries). To do so, we interpolated the top 10% to top 1% ratio between the 1930 value (which is an average of the ratio in France and Sweden where

we do observe top 10% and top 1% shares in this period) and the 1970 value (which is assumed to be equal to the observed 1980 ratio).

3. Top shares were adjusted upwards to align with top 1% and top 0.1% shares available in WID.world.

4. We interpolated the income distribution between 1950 and 1980 for the historical series in Other MENA countries to ensure a smoother connection between pre-1980 and post 1980 series.

5. We identified a jump in the top 10% share in the Other South & South-East Asia region from the historical series through 1970 to the series available on WID starting in 1980, due to different estimation methods. To reduce the effect of this change in methods, we have interpolated the income distribution between 1950 and 1980 for the historical series.

6. We updated the post-1980 historical series, based on the latest version of the WID.world database (and moving forward will be updated synchronously with the rest of the database).

7. We also revised 1970 income share estimates for a subset of countries. This change does not impact historical estimates.

4.2 Adjustments and updates made to WID per-adult series available on WID.world before December 2021

Five adjustments were made to the per-adult income distribution series that was previously available on WID.world.

1. Estimates of Germany's top 1% and top 10% income shares during the 1871-1940 period available on WID.world were found to be too low during research for the historical series. Thus, we have now adjusted the series available on WID.world during this time period to be in line with the estimates in Chancel and Piketty, 2021.

2. Estimates of the United Kingdom's top 10% shares from 1910-1970 available on WID.world were revised based on the recent research by Scott and Walker 2020. We have now adjusted this series to match the historical series, including the correction listed in the previous section.

3. Estimates of Indonesia's top 1% and top 0.1% series from 1921-1939 have been removed and replaced with the estimates from the historical series.

4. Estimates of Other Russia and Central Asia's income distribution depend on the income distributions of the countries that make up this region. Income distribution data was not available until the late 1980s or 1990s for many of these countries and inequality was previously assumed to be flat from 1980 until the first data point in each country. However, given that these countries were all part of the Soviet Union, we have now revised these estimates to interpolate between the income distribution of this sub-region in 1970 and the first available income distribution of the country.

5. We revised the treatment of North Korea and its integration in the Other East Asia region in order to better deal with the lack of available data about national income at purchasing power parity.

5 Consistency between fiscal and pretax income distribution series

Finally, we checked for discrepancies between other WID.world income distribution series. We focused on six series that describe two income concepts (fiscal income and pretax income) for three different population units (individuals, tax units and equal-split adults). Specifically, we checked for discrepancies between all combinations of fiscal income per tax unit, pretax income per tax unit, fiscal income per equal-split adult and pretax income per equal-split adult as well as comparing the two individual series, for a total of seven pairwise comparisons.

We used similar criteria to flag income shares as discrepancies as to compare the historical per adult and per capita series. We flagged an income share (bottom 50%, middle 40%, top 10%, top 1% or top 0.1%) in a given country-year as a discrepancy if it met any of the following criteria:

1. The gap between any pair of series was larger than 10 percentage points

2. There was at least a 75% gap between any pair of series for top 1% and top 0.1% shares

3. There was at least a 5 percentage point gap between the changes in any pair of series

4. There was at least a 3 percentage point gap that was also at least a 100% gap between the changes in any pair of series for top 1% and top 0.1% shares

Most discrepancies identified in this manner were determined to be legitimate discrepancies due to the different definitions of pretax and fiscal income. It should also be noted that fiscal income inequality series available in WID.world are intermediate estimates utilized by researchers in the construction of pretax income inequality series. Fiscal series may not be comparable across countries or across time and may not be updated in concert with pretax income inequality series. However, this process also identified several discrepancies that needed to be addressed.

We noticed a pattern in discrepancies that led to us revise our pretax income estimates for country-years in which pretax income shares were estimated as a function of fiscal income shares. Previous estimates applied a fixed ratio of fiscal income shares to estimate pretax income shares for each country-year-percentile. This fixed ratio was the first observed ratio of the pretax income share to the fiscal income share for that country-year-percentile. We've replaced this fixed ratio with a predicted ratio that varies according to a polynomial in the year and a polynomial in the magnitude of the fiscal income share. Ratios were predicted using a separate model for each percentile which was calibrated using all country-year-percentiles for which we have (non-imputed) estimates of both pretax and fiscal income shares. These ratios were constrained to vary smoothly from the first observed ratio to maintain series continuity. This change resulted in reducing our estimates of top shares in the early 20th century for several countries including Sweden, Norway and Japan. We also now include estimates of the full distribution of pretax income whenever we have some pretax income share estimates available and include estimates of per-capita income shares whenever per-adult income shares are available (see Table 1).

We also identified and rectified inconsistencies in Taiwan's top 1% pretax income share

Table 1: WID.world pretax income shares data coverage

Countries/regions	Years	Per adult estimates	Per capita estimates
All countries and regions [1]	1980-2020	127 g-percentiles	127 g-percentiles
Twenty-four countries, nine regions and eight sub-regions [2]	1820, 1850, 1880, 1900, 1910, 1920, 1930, 1940, 1950, 1960, 1970	127 g-percentiles	127 g-percentiles
Select countries [3]	Select years	127 g-percentiles	127 g-percentiles

[1] For a full list of countries and regions with complete data post-1980, see: https://wid.world/codes-dictionary/#country-code

[2] For a full list of these countries, regions and sub-regions, see https://wid.world/document/longrunpaper/

[3] At least one additional year of data are available before 1940 for Switzerland, Denmark, Finland, Hungary, Ireland, Korea, Netherlands, Norway, Portugal, Vietnam and Zimbabwe. At least one additional year of data are available before 1980 for Cameroon, Ghana, Greece, Israel, Kenya, Mauritius, Nigeria, Seychelles, Singapore, Tunisia, Taiwan, Tanzania, Uganda, Zambia and Zanzibar.

during the 1990s and early 2000s and a significant discrepancy between Malawi's pretax and fiscal series in 1980.

6 Conclusion

In this update, we have added per-adult income distribution series for 24 countries, 8 sub-regions and 9 world regions to the WID.world database from 1820 to 1970. These supplement the previously available per-capita series by estimating the average income available to adults across the income distribution across the world since 1820, a potentially more relevant statistic than average income per individual. Furthermore, the previously available per-capita estimates have now been harmonized with the available estimates of the per-adult income distribution. Historical estimates of the per-adult pretax income distribution that are based on historical fiscal income share estimates have also been updated. We have made a first attempt at estimating adult population totals around the world since 1820. While long-run distributional national income and fiscal income series are now available on WID.world, we stress that we view national income distributional series as reference series, and fiscal income series as intermediate steps computed by researchers and which should only be used by expert users. Overall, the work presented in this document helps us better understand the global income distribution since 1820. For further discussion of the construction of these series and the interpretation the trends observed, see Chancel and Piketty, 2021.

References

Ajus, Ferenc (2010). *Documentation for Children per Woman (Total Fertility Rate) for countries and territories*. en. Gapminder Documentation 8. Stockholm, Sweden, p. 80. URL: `https://www.gapminder.org/documentation/documentation/gapdoc008_v2.pdf` (cit. on p. 3).

Blanchet, Thomas, Juliette Fournier, and Thomas Piketty (2022). "Generalized Pareto Curves: Theory and Applications". en. In: *Review of Income and Wealth* 68.1, pp. 263–288. ISSN: 1475-4991. DOI: `10.1111/roiw.12510`. URL: `https://onlinelibrary.wiley.com/doi/abs/10.1111/roiw.12510` (visited on 03/29/2022) (cit. on p. 4).

Chancel, Lucas and Thomas Piketty (Dec. 2021). "Global Income Inequality, 1820–2020: the Persistence and Mutation of Extreme Inequality". en. In: *Journal of the European Economic Association* 19.6, pp. 3025–3062. ISSN: 1542-4766, 1542-4774. DOI: `10.1093/jeea/jvab047`. URL: `https://academic.oup.com/jeea/article/19/6/3025/6408467` (visited on 03/29/2022) (cit. on pp. 1, 3, 6, 7, 9).

Scott, Peter and James T. Walker (Mar. 2020). "The Comfortable, the Rich, and the Super-Rich. What Really Happened to Top British Incomes during the First Half of the Twentieth Century?" en. In: *The Journal of Economic History* 80.1, pp. 38–68. ISSN: 0022-0507, 1471-6372. DOI: `10.1017/S0022050719000767`. URL: `https://www.cambridge.org/core/product/identifier/S0022050719000767/type/journal_article` (visited on 03/29/2022) (cit. on p. 7).