



LUND UNIVERSITY

Methodological Notes

for "Income Taxes and Redistribution in the Early Twentieth Century" and "Income Tax Progressivity and Inflation during the World Wars"

Torregrosa Hetland, Sara; Sabaté, Oriol

2021

Document Version:
Other version

[Link to publication](#)

Citation for published version (APA):

Torregrosa Hetland, S., & Sabaté, O. (2021). *Methodological Notes: for "Income Taxes and Redistribution in the Early Twentieth Century" and "Income Tax Progressivity and Inflation during the World Wars"*. (Lund papers in Economic history; No. 2021:223).

Total number of authors:
2

General rights

Unless other specific re-use rights are stated the following general rights apply:

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Read more about Creative commons licenses: <https://creativecommons.org/licenses/>

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

LUND UNIVERSITY

PO Box 117
221 00 Lund
+46 46-222 00 00

Lund Papers in Economic History



No. 223, 2021

Methodological Notes

for “Income taxes and redistribution in the early twentieth century” and “Income tax progressivity and inflation during the World Wars”

Sara Torregrosa-Hetland & Oriol Sabaté

REVISED: September 2022

Methodological notes

for “Income taxes and redistribution in the early twentieth century” and “Income tax progressivity and inflation during the World Wars”¹

Sara Torregrosa-Hetland² and Oriol Sabaté³

Abstract

This document presents the methodological approach used in two papers about historical income taxes: “Income taxes and redistribution in the early twentieth century” (Torregrosa-Hetland and Sabaté, 2021) and “Income tax progressivity and inflation during the World Wars” (Torregrosa-Hetland and Sabaté, 2022). We first describe the general method and sources used to obtain synthetic distributions of income and calculate the effective income tax rates and the corresponding indices of progressivity and redistribution. Secondly, we discuss the most important country-specific issues that have been taken into account in our calculations. Finally, the third section looks at the accuracy of our synthetic income distributions and tax simulations by comparing them with the original series from the tax statistics. The two aforementioned papers summarize this same information in their methodological sections, but this note goes more in depth into some details that might be of interest to some readers.

Keywords: Taxation, Redistribution, Progressivity, Income tax, World Wars.

JEL codes: H23, H24, N42, N44

¹ This work has benefitted from funding from Jan Wallanders och Tom Hedelius stiftelse with the project ref. P18-0034, as well as from Vetenskapsrådet (Swedish Research Council), ref. 2018-01269. Oriol Sabaté also has received funding from the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 752163.

² Lund University, Department of Economic History, Lund, Sweden. Corresponding author: sara.torregrosa_hetland@ekh.lu.se

³ University of Barcelona, Department of Economic History, Institutions, Politics and World Economy, Barcelona, Spain (oriol.sabate@ub.edu).

1. Introduction

This document presents the methodological approach used in the papers “Income taxes and redistribution in the early twentieth century” and “Income tax progressivity and inflation during the World Wars”. We first describe the general method and sources used to obtain synthetic distributions of income and calculate the effective income tax rates and the corresponding indices of progressivity and redistribution. Secondly, we discuss the most important country-specific issues that have been taken into account in our calculations. Finally, the third section looks at the accuracy of our synthetic income distributions and tax simulations by comparing them with the original series from the tax statistics. The two aforementioned papers summarize this same information in their methodological sections, but this note goes more in depth into some details that might be of interest to some readers.

2. General methodological approach

In the two papers we aim at exploring the evolution of progressivity and redistribution in the income taxes of Sweden, the United Kingdom, and the United States. To do so we need to look at incomes and effective tax rates across the entire distribution of incomes (which do not only provide valuable information on their own, but also constitute the basis for the calculation of indices of progressivity and redistribution). Unfortunately, the historical sources only provide information in aggregate form, and for tax units over a certain level of income (the exemption threshold in place). This section explains the method used to approximate incomes below these thresholds, and to estimate effective tax rates and the corresponding indices based on the information available in the tax statistics.

The original data used in the papers comes from the historical tax statistics of each country and provides us with these (incomplete) distributions of income, tax returns and (for some years) tax due. The most comprehensive series (yearly since 1914) exist for the United States, while in the other two countries the information is scattered over time. When it exists, the data suffers from serious shortcomings that make the primary sources unusable without some elaboration. To begin with, and as mentioned earlier, the data is generally limited to tax units that filled the income tax form for the tax authorities, so it lacks information on the number of people exempted

from paying the income tax, as well as on their income. However, for our purposes we need complete information across the entire income distribution (not just on taxpayers) for several years.

Secondly, the aforementioned distributions of income and tax returns are grouped in the sources by income levels that generally do not coincide with those of the brackets in the tax schedules, and also change across countries and over the years. For instance, the distribution of income and tax returns in the United Kingdom in 1938/39 is based on 12 income levels ranging from a minimum of £125 to over £20,000, whereas in 1949/50 the same information is divided in 12 levels which extend from £135 to over £20,000. On the other hand, the tax schedule is divided in 13 brackets ranging from £125 to over £50,000 in 1938/39, and in 12 brackets from £135 to over £20,000 in 1949/50 (that do not coincide, however, with the 12 income levels in the distribution of income and tax returns mentioned above). Thus, in order to make calculations comparable across countries and over time, and to illustrate the distribution of tax rates over the population, we need to adapt these numbers to a synthetic micro-distribution.

To address these issues, we have first gathered information on the total number of tax units (individuals or families that are considered one unit for the purpose of income tax, including those with incomes below the exemption limits) and their total income for each country during the first half of the twentieth century. These data come from various works in the top incomes literature (exact references are given for each country in the next section). The residual between total income of all tax units in the economy and income assessed by the tax authorities (and therefore available in our primary sources) corresponds to the income of tax units below the exemption threshold, whereas the analogous residual between total number of tax units and tax returns is the number of tax units exempted.⁴ This information allows us to have at our disposal the distribution of income, tax returns and tax due, for the entire population of tax units, albeit grouped in non-homogeneous brackets.

We follow the recent method and software developed by Blanchet et al. (2017)⁵ to disaggregate data from grouped statistics, such as cumulative income shares, which

⁴ Note that the residual might also include incomes not reported by those who filed returns (fraud), and it is therefore only an (over)approximation of the revenues of the exempted population. This might be a problem of considerable magnitude in some cases, and it will be tackled in future work.

⁵ Previous versions of this work used the method by Shorrocks and Wan (2008), implemented through the Stata package DASP, provided by Abdelkrim and Duclos (2007).

has been devised precisely for tax data. The procedure generates a synthetic sample consistent in mean and distribution with the original information inputted, using the properties of the Pareto coefficients. The calculus is based on the cumulative share of tax returns and total income for each income bracket, as well as on the total average income for a given year. The resulting samples contain 1 million equally weighted observations for each year; a number high enough to capture the higher variability present in the upper part of the income distribution. These synthetic samples are mostly consistent with the original data in terms of number of units and average incomes in each bracket (see Section 4).

Once the synthetic sample has been generated, we proceed in a way similar to Piketty (2001, annex B.3) in applying the regulations in force in each country to simulate the operation of the tax, estimating tax payments for the average income of each quantile (in our case, one million).⁶ The tax base corresponds to the gross income (total amount originally received by the tax unit) excluding deductions for costs (for instance, for housing repairs) and exempted incomes (such as those below the exemption limits). As a first step we deduct family allowances (deductions for the taxpayer based on his marital status and the number of children, when present) from the tax base to obtain taxable income.⁷ Since these allowances depend on family circumstances, we generate eight synthetic taxpayer types within each observation: singles and couples with zero, one, two, or three children. After deducting the corresponding family allowances from the original gross income (thus obtaining taxable income), we apply the marginal tax rates in the schedule to each synthetic taxpayer type. When we have obtained the tax due corresponding to each of them, we calculate the value for each of the million original observations as a weighted average of the different family types.⁸

⁶ The rationale that follows applies to all countries of our sample, with specificities described in the following sub-sections.

⁷ Other deductions and allowances are present in some cases, but we mostly cannot include them in our calculations due to the lack of statistical information on their distribution. We do include several of them in the case of the United Kingdom and the United States, given the higher quantities involved (see the following sections). Nevertheless, in general, family allowances are the most important ones.

⁸ The information on the distribution by family types comes from the tax statistics themselves, except for the United States, for which we used the *Statistical Abstract of the United States* (1950) that provides information from the Census of 1940. The same family distributions (weights) have been used over the

Next, we calculate the effective income tax rate by dividing each observation's tax due by its gross income. This effective tax rate represents a simple average for all the tax units represented by each of the 1-million synthetic taxpayers. We also estimate average effective tax rates for different percentiles and permilles of the income distribution by averaging the effective tax rates of all observations that fall within the corresponding income levels.

Effective tax rates can also be calculated in an aggregate form: total tax due of the group / total gross income of the group. This is different from our approach described above, and the aggregate results will normally be higher (in a progressive tax).⁹ We have also calculated effective tax rates in this way (for a few income brackets) in order to compare our simulations with the statistics provided by (or directly obtainable from) the original tax sources. As can be seen in Section 4, they are largely consistent. The main differences are generally found at the upper part of the income distribution, where our estimators are sometimes higher. This discrepancy seems to be driven by the low number of observations in these top groups, which renders the estimates imprecise, as well as by the fact that additional tax allowances and deductions which we do not consider benefited mostly the well-off. The comparison, overall, suggests that our calculations are a reasonable depiction of the original tax data.

At this stage, we are ready to calculate the progressivity and redistribution of the income tax system in our three countries of interest. We follow the general framework in public economics to estimate progressivity and redistribution indices (Kakwani, 1977; Lambert, 2001), using the 'progres' stata module developed by Peichl and van Kerm (2007). For progressivity we use the Kakwani index, which is obtained as the difference between the concentration of tax payments C_T and the Gini of gross incomes G_Y :

$$K = C_T - G_Y \quad (1)$$

The index would be 0 for a proportional tax (i.e., where tax payments were concentrated to the same extent as incomes), and gets positive values when the tax is progressive.

years, since there is no yearly information and this feature does not experience abrupt changes; Piketty (2001) followed the same approach for 1915-44.

⁹ The tax rates calculated in this aggregate way are equivalent to the average effective tax rates defined above, weighting tax units by their income.

Redistribution is measured with the Reynolds-Smolensky index, which corresponds to the difference between the Gini indices of gross and net incomes (i.e., before and after tax):

$$RS = G_Y - G_{Y-T} \quad (2)$$

A tax is redistributive if $RS > 0$.

The relationship between these indices is given by the expression:

$$RS = \left[\frac{aetr}{(1-aetr)} K \right] - RR \quad (3)$$

where RR is the effect of re-ranking between tax units. Redistribution by the income tax is thus positively affected by progressivity (K) and the average effective tax rate ($aetr$), defined in this case in aggregated terms.

3. Country-specific methodological approaches

In this section, we describe specific adjustments to the general methodology that had to be made in some cases to adapt the general methodology to the tax regulations in place.

3.1. Sweden

- Total income: the total income series we use comes from Roine and Waldenström (2010),¹⁰ but has been adjusted for consistency with the tax base definition in the Swedish income tax in the period under study. This entails the following:
 - o Social benefits have been removed, since they were not taxable before 1974. The series (starting in 1937) is also provided by Roine and Waldenström (2010), table A1 in the Appendix.
 - o Total municipal income tax revenue has been removed before proceeding to generate the synthetic distribution, for the period 1920-42, since it was deductible from the state tax base and they are therefore not included in the grouped data from which we depart (see Roine and Waldenström, 2010, table 7.1). We have used the series of “*Allmän kommunalskatt*” from the Swedish Statistical Yearbooks (see Taxing for the Welfare State database).

¹⁰ These total incomes were constructed with a “top-down” methodology for the period before 1943 (departing from National Accounts, but downward adjusting by a factor of 0.89), and “bottom-up” since 1943 (departing from the tax data).

The municipal income tax revenue is added again, for each observation, before calculating the effective tax rates and the corresponding indices of progressivity and redistribution (see the section “Local income tax” below). A previous version of the paper estimated the proportional local income tax using the reported tax base in the state tax and the average statutory municipal tax rates in Du Rietz et al (2015), Appendix D, table 3.

- o Wealth imputation: a fraction of the value of wealth was included in the tax base of the Swedish income tax between 1911 and 1948. The fraction was first one sixtieth, and then reduced to one percent in 1939 (Henrekson and Stenkula, 2015, p. 31). Because of this, the total value of this wealth estimate for all taxpayers had to be added to “total income” before proceeding with the disaggregation (prior to 1945, see next point). We have estimated total imputed wealth with the 1921 datum coming from the census (wealth as a share of income for the whole population; *Statistical Yearbook* 1929), and extended the series using the variations in the capital-income ratios from Waldenström (2017).
- o Capital gains: we use the series of total income including capital gains. Capital gains are a part of income according to the classical Haig-Simons definition, but they are often excluded from studies of inequality because of their irregular nature (a capital gain realized in one year may have been generated through a long period). See a discussion of their importance for Swedish income inequality in Roine and Waldenström (2010).
 - The series of total income including capital gains is estimated before 1945 using the percentage from this year, and thus is 1.24% superior to the series without capital gains. According to Roine and Waldenström (2010), this is possibly an underestimate.
 - Capital gains were taxed under the Swedish income tax, together with the rest of incomes, but to a varying extent across our period. Between 1911 and 1951, capital gains generated through more than five years were exempt, while the other were included in the tax base at their whole value. After 1951, more variation was introduced. See Du Rietz et al. (2014), table 1, page 12.
- o Downward adjustment of the incomes of those below the threshold (which is estimated as a residual) has been necessary in the tax year 1921. The

problem is likely due to the high inflation and the extension of the obligation to file a return, which was probably not fully complied with. Additionally, the total income we use as a reference might have been an overestimate for this year. We have imputed to the lowest group the same mean income of 1920 (453 kr, versus a threshold of 600 kr).

- o Incomes of the lowest group have also been adjusted in 1945 and 1946, to a level of 400 kr, which is slightly more conservative than the 300 kr used by Bentzel (1952). The reason for this was that the simulated revenue and number of taxpayers from the local income tax were considerably higher than in the sources (we compared our estimates with the tax revenue estimated in the Taxing for the Welfare State database, obtained from Swedish Statistical Yearbooks and *Skattetaxeringarna*, and with the preliminary estimate of the number of taxpayers in the local income tax from Bengtsson et al., 2022).
- Income distribution (tax base):
 - o The distributions for 1945 and 1946, taken from *Skattetaxeringarna*, correspond to *sammanräknad nettoinkomst* and not *taxerat belopp* (tax base). The concept *sammanräknad nettoinkomst* does not include the wealth imputation, since it just refers to income net of costs of obtainment (but should not be confused with net incomes, i.e. after tax; see explanation in *Skattetaxeringarna* from tax year 1944, p. 5*). We have estimated a simple regression of the wealth share over (logged) income, using tabulated data from 1942 (Quensel, 1944, table VII), and used these results to impute wealth shares in the 1945 and 1946 microdata, adjusting to the total reported wealth supplements given in the sources. The same procedure has been used for the years 1936 and 1941, where the income distributions used come from Bentzel (1952) and do not include the wealth imputation either.
 - o When comparing across countries, it needs to be taken into account that the original tax base distributions for Sweden between 1911 and 1948 include these wealth imputations. We have, however, adjusted them using the distribution of wealth for 1942 from Quensel (1944), so that the reference for the average effective tax rates, the progressivity indices, and the quantiles is always income (without wealth imputations). The adjustment procedure is based on a linear regression estimating the share of wealth in

income with respect to the logarithm of income. Bentzel's (1952) distributions are given in percentage terms; they have been complemented with the total income and total number of tax units from the World Inequality Database. Total income has been used, without adjustment for wealth, social benefits, or local taxes, because the concept used in Bentzel (1952) is *förvärvsinkomst*: market incomes, a broader definition than the fiscal one.

- o The income distribution for 1938 comes from Quensel (1944, table IV and table A4), and includes his own estimate of wealth imputations. The same source also provides a distribution for 1942, but this one has not been used, since it leads to an extremely high average income in the bottom group.
- o Information on incomes subject to *bevillning* (old tax on incomes below 1000 krs) was available in some years, for example regarding 1907 in *Taxeringen 1907*, p. XVI. But the definition of the tax base for this tax was different than that for the modern income tax and thus has not been incorporated in our calculations.
- o Incomes taxed were (an estimation of) those of the current year between 1903 and 1907; in 1908 a transition was made to start paying tax on the income of the preceding year. *Taxeringen 1907*, p. IX: "*Inkomst taxerades ännu år 1907 för det löpande året (icke för det föregående, såsom fallet varit fr. o. m. år 1908*". The transition implied that in taxyear 1908 only half of income tax was paid (on the same incomes already taxed in 1907), and in taxyear 1909 income tax was paid on the incomes earned in 1908. See p. 207 in *Förslag till förordning angående bevillning af fast egendom samt af inkomst och förordning om inkomstskatt..., utarbetadt år 1906 inom Kungl. Finansdepartementet*, P. Palmquists AB, Stockholm, 1907 – SOU 1907:4. The income tax revenue in 1908 was indeed approximately half that of both 1907 and 1909.
- Types of taxpayers:
 - o The original statistics include, together with Swedish residents, other three types of taxpayers: Swedish living abroad, non-Swedish residents, and some types of juridical persons (not corporations). These are included in the

personal income tax tabulations, together with the bulk of taxpayers.¹¹

Ideally, we would like to exclude juridical persons from our calculations, but this is not possible due to lack of data. *Taxeringen till inkomst...* for the year 1917 allows calculating that the three groups together account for around 1% of total taxpayers in each income level (0.3% in the case of juridical persons). Juridical persons represented 2.0% of the taxable base and 3.8% of the tax paid in that year, a percentage remarkably similar to that in 1939 (1.5 and 4.4% respectively, according to the data in *Skattetaxeringarna*). We are thus quite confident that this is not causing any significant bias in our estimates.

- o The Swedish regulation obliged married couples to present a joint return up to the income year 1966. Between 1966 and 1970, there was an option for them to make separate returns, and after 1971 the individual return has been the standard (compulsory). See Söderberg (1996). This issue does not affect our estimations in the paper, since the change to individual taxation of married couples took place after our period of study.
- Family allowances and estimation of taxable income:
 - o Family structure data (share of households corresponding to each of the family types) are taken for 1913 and 1917 from the tax sources, *Taxeringen till inkomst...* These correspond to tax units under 6,000 krs (which were the ones entitled to family allowances). The original data includes three different family types (single without children, couple with or without children), and we have calculated the average number of children for the couples who had them from the statistics of deductions claimed for children (2.59 and 2.34 respectively). For 1945, we have the distribution of tax units with taxable income found in the census (*Folkräkning*). Since this one was very coincident to the 1920 data in the distribution of singles and couples, we have used it as well for that previous year.
 - o Taxable income distribution data exist for 1912, 1913, 1917 and 1920 (by levels of tax base). We have therefore been able to check the accuracy of our estimates, in terms of taxable income as a share of tax base for different

¹¹ But, until 1917, they are not distributed under 6,000 kr (since the rules for allowances did not apply to them).

income brackets. Our estimated taxable incomes correspond quite well with those in the source, even if at lower and middle-income levels they are sometimes slightly higher than in the sources. This could arise from them having more children and therefore more children allowances than the average.¹² The effect might be that our calculations underestimate progressivity.

- o For the period before 1911, deductions for kids and spouse did not exist. A small fraction of income was exempted from taxation (and is thus not present in the *taxerad inkomst* statistics). Between 1.000 and 2.000 krs of income, 800 krs were exempted; between 2.000 and 3.000 krs of income, 600 krs were exempted; between 3.000 and 4.000 krs of income, 400 krs were exempted. See *Taxeringen 1907*, p. VIII. This has not been adjusted for because of insufficient information. This entails that, for 1907 and 1909, the reference income to calculate the effective tax rates and the corresponding indices of progressivity and redistribution is the tax base and not a broader concept of income, which would under-estimate progressivity.
- o In the 1910 regulation, deductions for kids were introduced (100 per child under 15 years), as well as a basic exemption specific to cities with high price levels. The exemption was of up to 300 krs if income was below 1.800 krs, and of up to 200 if income was between 1.800 and 3.000 krs. We have applied to all observations in these income levels a weighted average using the percentage of population in the cities of groups 4 and 5 in 1920 (see following paragraph), which stands at 28,4%.
- o Starting in 1920, family allowances varied according to the price level in the municipality of residence of the taxpayer, distinguishing five groups. We have used the distribution of taxpayers by city group to calculate weighted allowances. Data come from *Taxeringen... 1920*, pp. 6*-7*, 36*, 40*, from *Sociala Meddelanden 1942* (nr. 10, p. 807; corresponding to 1935), and from *Statistisk Årsbok 1950* (tab. 173, p. 193; corresponding to 1947).

¹² Data on number of children by family income level are not available in Sweden. According to Dribe and Scalone (2014), families of lower social classes (by profession) in 1900 had more children, but by 1960 there was no clear social gradient (data from censuses).

- o Since 1938, the allowances for singles followed a special table, which meant they were reduced as taxable income grew. We have taken these values from the original tabulations in the law (one table for each city group).
- o In 1920, the source provides the distribution of taxable income, by levels of taxable income. This allows us to perform an alternative estimation, where we directly disaggregate taxable income, and proceed from there with the estimates of tax dues (without having to simulate family allowances). The results of this alternative estimation coincide very closely with those of our baseline. Even if this has to do with the 1920 tax being widely proportional for most of the distribution, estimated effective tax rates are not far from each other at the top either.
- Tax schedules:
 - o The tax schedules (marginal tax rates applied to each level of taxable income) are taken from Du Rietz et al (2013), Appendix D.
 - o The system of 1911 is more complex than the general explanation, since under 6.000 krs the schedule was given in average rates (i.e., not marginal: the tax rate given in the tables was applied to all of taxable income).
 - o We have included the *extraskatt* in 1919, the *extra statlig inkomstskatt* in 1932-38, and the *värnskatt* in 1941, 1945-46. These were additional taxes on income, which in combination with the general tax resulted in a more progressive schedule. The corresponding increases in revenue were ultimately incorporated to the general income tax.
 - o A defence tax on high incomes was also enacted in 1914 (referring to the incomes of 1913, and including one-tenth of wealth in the tax base). It was paid over the years 1915, 1916 and 1917 (Genberg, 1942; Du Rietz et al., 2013). This tax is not included in our estimates for 1917, so they might be considered a lower bound.
- Local income tax:
 - o The proportional local income tax has been estimated using the average local income tax rate given by Söderberg (1996). Local tax rates varied by municipality. We start including the local income tax after the reform of 1920, which made it a modern one, based on similar principles as the central state tax. This reform was first applied to the incomes of 1921, in the taxyear 1922.

- o The tax base for the proportional local income tax did not include the wealth imputations; these have been removed for years before 1945 based on the wealth distribution of 1942 (Quensel, 1944, table VII), and adapting to the totals for each year given in *Skattetaxeringarna*.
- o Information on basic deductions (*skattefria grundavdrag*), which were introduced in local income taxation after World War I, comes from Söderberg (1996). The rules from city group 3, the intermediate one, are as follows:

Taxyear	Deduction for taxpayer	Deduction for spouse and for each kid
1922	600	150
1923-27	450	150
1928	400	100
1929-52	420	180

Up to 1952, these basic deductions were subject to the same “*bankning*” rules applied in the state tax (adaptation of the amount deducted to the size of the tax base).

Taxyear	Deduction for single taxpayer	Deduction for married taxpayer
1953-58	1290	1840
1959-62	1840	3680
1963-71	2250	4500

We obtain average weighted deductions using the same distributions by city groups as for the state tax. In the last period, 1963-71, city groups were no longer used.

- o Between 1920 and 1970, the proportional local tax paid was deductible when calculating the state taxable income (Du Rietz et al., 2015). Therefore, the proportional local income tax paid was added to the state taxable income to calculate gross incomes, which are used as a reference for the progressivity calculations.
- o Our simulations of proportional local income tax paid are less precise than those of the state income tax, because the income distribution data come from the state tax statistics, which had a higher threshold than local taxation (often, double as high). The simulations result in a higher aggregate revenue for this proportional local tax than the one given in the sources. This does not affect the basic conclusion of our international comparison, since it reinforces the fact that, even if we included local taxation, the Swedish

system did not attain the levels of income tax redistribution of the United Kingdom and the United States.

- o We have also considered in our simulations a local progressive income tax which was in place between 1922 and 1938 (*den kommunala progressivskatten*), and the “state equalization tax” which was in place between 1929 and 1938 (*den statliga utjämningskatten*). These taxes were levied on state taxable income (Söderberg, 1996; Du Rietz et al., 2015), and therefore (unlike the proportional local income tax) were not deductible from the state tax base. The equalization tax can be calculated as a given percent of the local progressive income tax.

3.2. United Kingdom

- Income distribution:
 - o The British income data comes mainly from various volumes of the *Reports of the Commissioners of His Majesty's Inland Revenue* (from now on, the Reports). The 64th Report (corresponding to the fiscal year 1920-21) provides these estimates for the fiscal year 1919-20, and thus constitutes our source of information to study the post-WWI period. According to the commissioners, the Financial Act passed in 1920 made the aforementioned estimations too difficult and costly to calculate, so analogous estimates were not presented again until new Income Censuses were conducted just before and after WWII. We therefore rely on the 92nd Report for the estimates related to the fiscal year 1938-39, and the 94th Report for the fiscal year 1949-50.¹³ The estimates for the post-WWII benchmark have been complemented with the information provided by Scott and Walker (2020).

¹³ Similar information exists for 1948-49 based on the results of the 1949-50 census, but we prefer to use the census year to improve the accuracy of our estimates. The Commissioners of Inland Revenue also conducted a special investigation for the year 1937-38 (that has been further complemented by Scott and Walker, 2020). However, we use the 1938-39 benchmark (which builds on the 1937-38 investigation) because the original distribution of income covers a larger share of tax units (income above £200 in 1937-38 and above £125 in 1938-39).

Data on taxpayers, income, and tax due, for the fiscal year 1911/12 (our pre-WWI benchmark) also comes from Scott and Walker (2020).¹⁴

- o The 64th Report provides information on income assessed *in* the year 1919-20, which includes income generated in this same fiscal year but also income carried forward from preceding years (mainly related to profits from trade assessed under Schedule D). Unfortunately, the Report does not specify how much income corresponds to 1919-20 and to previous fiscal years. According to the Report this amount is “*relatively small and fairly constant*” (p. 108), so we have not made any adjustment to address this issue. Secondly, the income tax in our period of study included not only the regular income tax but also the so-called “super-tax” (imposed upon incomes that surpassed a certain income level – see below). The income assessed in a given year (e.g., 1919-20) was assessable to the regular income tax in the same current fiscal year (in our example, 1919-20), whereas it was assessable to the super-tax in the following year (1920-21). We, nevertheless, have estimated the final tax due for 1919-20 as if the entire income was assessable to the tax in 1919-20, since we are more interested in the tax due generated by the income of a given year (despite of the limitations described above) rather than on the specific year when the income becomes assessable.
- o Information on the distribution of income, tax units, and tax due, for our pre-WWII benchmark (1938-39) comes from the 92nd Report (published in 1950). The Report (Table 85) provides data on the number of incomes, the total income before tax, and tax due, for 12 income brackets (ranging from £125-£150 to above £20,000).
- o Information on the distribution of income and tax units for our post-WWII benchmark (1949-50) comes from Scott and Walker (2020, 53), who update the figures found in the 94th Report (published in 1952). Information on the distribution of tax due comes from the 94th Report (Table 110, p. 117). The

¹⁴ No similar information has been found for 1912-13 or 1913-14. Scott and Walker (2020, 39) rely on unpublished estimates made by the Commissioners of Inland Revenue. According to the authors, these estimates had been kept confidential due to the “extreme political sensitivity of Britain’s high concentration of income and wealth”.

information in the report is disaggregated in 12 income brackets (ranging from £135-£150 to above £20,000), whereas Scott and Walker (2020) disaggregate 6 additional brackets (from £20,000 to above £100,000, although they do not provide tax due data).

- o Along the lines of the post-WWI benchmark, the estimates for 1938-39 and 1949-50 reflect the assessments made in a particular year, not the assessments corresponding to the income generated that particular year. However, the 94th Report states, once again, that *“Most of these assessments will of course relate to the year in which they are made; there will also be a small proportion of assessments in respect of past periods, but broadly speaking these may be regarded as balanced by assessments for the year in question which are made in subsequent years.”* (p. 39) Thus, no adjustment has been made.
- Total income and total tax units:
 - o Both the total tax units in the population and the total income come from Atkinson (2007, Table T4B.1, p. 126), except for total income in 1911 which comes from Scott and Walker (2020). We follow Atkinson and relate each fiscal year (e.g., 1919-20) to the population in the calendar year (1919 in this specific example) in order to *“make some allowance for the lags”* in the income tax statistics already mentioned in the previous paragraph.¹⁵
 - o As for the total income for the 1919-20 benchmark, Atkinson starts from the “actual income” assessed by the Inland Revenue and adds the income of non-filers (exempted from paying the tax) in order to find the total income corresponding to the total number of tax units (what the author calls “adjusted total income”). We also rely on the same category of income (“actual income”) for the income of taxpayers, so we can subtract this assessed income from the Atkinson’s “adjusted total income” to find the income of the exempted tax units.¹⁶ The distribution of “actual income” in which we base our calculations excludes the so-called “non-personal

¹⁵ Fiscal years run from April to March.

¹⁶ Our total income for tax filers is slightly different from Atkinson’s (around 3% lower or higher) because we use the sum of disaggregated income by income groups, whereas Atkinson uses the aggregated figures provided by the same source.

income” (i.e., income of companies that falls under the income tax but is not distributed to shareholders in dividends) and the income accruing to non-residents (which account for about 10% of the total income charged to income tax). Atkinson also excludes the former item to calculate his “adjusted total income”, but not the latter. This inconsistency between the two datasets would overestimate our estimated total income for the exempted taxpayers. Fortunately, both the Report and Atkinson estimate the total amount of non-residents’ income, so we can subtract it from Atkinson’s “adjusted total income”. Even if this adjustment entails that we do not take this source of income into account in our calculations, the amount is small (less than 2% of total income) and it corrects the overestimation of exempted taxpayers.

- o Atkinson (2007) follows the same procedure described above for the year 1938-39, but it varies a little when it comes to the total income for 1949-50. The original total income is taken directly from the Income Census mentioned above (called “total net income”), to which he subtracts undistributed profits and adjusts for non-filers. This “net income” differs slightly from the income category used in the previous benchmarks (namely, “actual income”). As a result, the “total net income” used by Atkinson is 0.7% higher than the total income used in our calculations (based on the disaggregated actual income provided by Inland Revenue). Since the discrepancy is small, we do not make any adjustment to it.
- o Most capital gains are not included in Atkinson’s (2007, 88-89) total income nor in the disaggregated tax figures. Indeed, the Report on the Income Tax published by the Inland Revenue in 1920 states that “*Casual non-recurring or occasional profits arising from transactions that do not form part of the ordinary business of the person who makes them are accordingly held not to be within the scope of the Income Tax, and consequently escape taxation.*” (Royal Commission on the Income Tax, 1920, 19). Hence, income that is not likely to recur annually remains outside of the scope of our estimations.
- Allowances and estimation of taxable income:
 - o The system of allowances and exemption limits varied significantly between the two World Wars. In the two periods, incomes that did not reach the exemption limit were exempted from making a tax return and from paying

the tax. When incomes exceeded such threshold, most part of the income below the threshold continued to be exempted from paying the tax through the operation of a system of personal allowances (although a small portion of such income would become liable to the income tax). Prior to the Financial Act of 1920, the corresponding tax due was graduated by deducting from the tax base a type of personal allowance called “abatements”. These were fixed deductions (i.e., pounds free from paying the income tax) on incomes that fell within certain income brackets. Apart from these abatements, in our calculations we take into account family allowances, which varied according to the number of children and the existence of a spouse (as long as the total income did not exceed certain thresholds).¹⁷ A similar system of allowances was in place after 1920, but family allowances were not subject to thresholds anymore. Moreover, the system of abatements was abolished, and instead single and married allowances were implemented (even if the name for the “single allowance” changed over time). These allowances operated once the total income exceeded the exemption limit (once again, these allowances exempted the bulk of the income below the threshold from paying the tax, although a small portion of it was not covered by them).

- o After 1920, when the income of a married couple included earned income of the wife, the married allowance was increased by a percentage of the wife’s income (with a maximum allowance that varied throughout the period; for instance, in 1920 the maximum allowance was £45, which amounted to 20% of the married allowance). Unfortunately, we do not have information on the distribution of earned income of wives. In order to take this allowance into account, we assume that all married couples claimed an exemption that amounted to half of the maximum allowance (e.g., in 1920 it would be £22.5).
- o Regarding married allowances, it is worth noting that since 1914 married women were allowed to submit their own income tax declaration (separately from that of their husband). However, this provision did not affect the total amount of tax paid by the couple, since marginal tax rates and allowances

¹⁷ The latter were introduced in 1918-19 for incomes below £800.

were calculated taking the joint income into account (and were distributed between them in proportion to their income). As a result, the wife's income was commonly submitted to the income tax as part of that of their husband (who was legally responsible for the joint income for the purpose of the tax, see 65th Report, p. 90, and HMSO 1980, p. 6-7). The option of having their incomes assessed separately as if they were two single persons was not introduced until 1971 (HMSO, 1980, 7). For this reason, we treat all married couples as single tax units for the purpose of the income tax (liable to the provisions of the tax mentioned above).

- o The aforementioned abatements in place prior to the 1920 Financial Act were slightly different for soldiers, sailors and other professionals: since 1915-16, those individuals with incomes below £160 were not required to fill in tax returns, and from £160 to £300 they were granted a special abatement of £160. By contrast, the general public was required to fill in tax returns when their income was above £130, and the abatement for incomes below £300 amounted to £120. Soldiers and sailors with incomes above £300 were subject to the same abatements than the general population. Since our fiscal sources do not differentiate the distribution of income for these two groups of taxpayers, we cannot separate the two systems of abatements. In order to avoid the risk of overestimating the total number of taxpayers and their tax due, we implement an adjustment to tax scheme: we simulate an abatement of £140 for incomes between £130 to £300 (half-way through the general abatement and that of the soldiers). In this way, we aim to reach a compromise between the two systems.
- o Soldiers and sailors were also granted lower marginal tax rates for incomes below £2,500 (in place since 1915-16), which probably generates an upward bias in our effective tax rates estimates (see the 61st Report published in 1918, p. 9). In this case we have preferred not to implement any adjustment to the series to avoid modifying the tax regulation too much. Furthermore, the estimated aetrs for incomes below £2,500 correspond very closely to the aetrs found in the original sources (see next section), which indicates that our calculations are precise enough.
- o Another important allowance in place after the Financial Act of 1920 and taken into account in our estimations was the so-called “earned income

allowance”. Income was divided between “earned income” and “unearned income” (or “investment income”). A percentage of the first £ of earned income was excluded from paying the income tax (see below more information and sources about earned and unearned incomes).

- o As for our post-WWI benchmark (1919-20), an additional allowance was granted to those total incomes that exceeded certain limits, such as the limit above which the individual ceases to be entitled to an exemption or abatement or when the individual becomes liable to a higher tax rate. The allowance ensured that *“the total tax payable does not exceed the sum of the following amounts:—(1) the amount of tax that would have been payable if his total income had reached but not exceeded that limit; (2) the amount by which his total income exceeds that limit.”* (63rd Report, p. 59). For the sake of simplicity, we have only included the operation of such allowances when they are related to the income tax brackets. Hence, when a person jumped to the next tax bracket, she paid the new tax marginal rate only if the condition cited above was not breached. We calculate this allowance for both earned and unearned incomes (this provision was implemented by the 1916 Financial Act, and therefore it does not affect our 1911-12 benchmark; see 60th Report, p. 11).
- Earned and unearned income:
 - o Apart from the aforementioned allowance, the distinction between earned and unearned income had associated two key provisions for the fiscal years prior to the Finance Act of 1920. First, different marginal tax rates schemes applied to each type of income, the rates imposed upon earned income being lower than those imposed upon unearned income. For instance, the marginal tax rate in 1919-20 for gross incomes between £130 and £400 (the lowest bracket) was 11.25% for earned income and 15% for unearned income. Secondly, in the case of mixed income, abatements and allowances were *“to be given, as far as possible, out of the income chargeable at the lower rate”* (63rd Report, p. 59). Hence, family allowances would only reduce the amount of unearned income assessable to income tax if these allowances were higher than the taxable earned income. These provisions were abolished by the Finance Act of 1920 and substituted for the aforementioned earned income allowance.

- o This implies that in our estimates we need to differentiate between these two types of income. The 83rd Report and the 94th Report provide this disaggregation based on several income brackets for the fiscal years 1937-38 and 1949-50. In absence of such data for 1938-39, we impute the 1937-38 estimation for our 1938-39 benchmark. Scott and Walker (2020) also provide the disaggregation between earned and unearned income across income brackets for 1911-12. Unfortunately, such information is not available for 1919-20. In our calculations, we assume that the distribution of earned and unearned income in 1919-20 followed the same pattern than in 1949-50. Even if 1911-12 and 1937-38 are closer in time, we use the 1949-50 distribution since it reflects the impact of WWII on capital incomes.
- o In order to differentiate between earned and earned income, we have estimated a simple regression of the earned income share over (logged) income, and used these results to impute earned income shares in the simulated microdata. Since the relationship between earned income share and (logged) income is not always linear, we allowed for up to four interactions of (logged) income in the right-hand side of the equation.
- Marginal tax rates:
 - o The operation of marginal tax rates before and after 1920 differed in fundamental aspects. Prior to the Financial Act of 1920, there were two types of marginal tax rates: the standard tax rate and the reduced tax rates. The standard tax rate was a unique rate imposed upon taxable income above a certain income level. Reduced tax rates applied to incomes above the exemption limit and below the previously mentioned given income level (as mentioned above, they were different for earned and unearned income). By contrast, after the 1920 reform, reduced marginal tax rates were chargeable on certain portions of the taxable income. In practice, this means that two taxpayers with the same tax base would face different marginal rates depending on the allowances that they were entitled to. For instance, in 1949-50 individuals with gross income above the exemption limit (£135) were supposed to pay a unique standard tax rate (9 shillings per pound, tantamount to 45%) on their taxable income. However, the first £50 of taxable income paid a reduced rate (3 shillings per pound), whereas the next £200 of taxable income paid the second reduced rate (6 shilling per pound).

- o Additionally, tax rates prior to the Financial Act of 1920 applied to the entirety of the individual's taxable income, and not just to the income falling within the corresponding tax bracket. For instance, taxpayers with gross income falling in the second tax bracket would pay the second tax rate for the entirety of their taxable income (and not just for the income above the second tax bracket).
- o We additionally added the super-tax scheme (which taxed incomes above a certain threshold) following the specific rules that applied in each year. Unlike standard tax rates prior to 1920, surtax rates applied only to the income falling within their corresponding tax brackets.

3.3. *United States*

- Capital gains: the total income series we use (from Piketty and Saez, 2003) includes realized and taxed capital gains, to the extent possible. The regulation concerning these has been quite variable through US history:
 - o Between 1913 and 1933, realized capital gains are included in the tax base. They were initially subject to the same taxes as other incomes, but in 1922 a preferential tax rate was introduced (nominally as a different tax). Since we do not simulate this tax privilege, our calculations will be a slight overestimate of the tax burden on high incomes (those who concentrate the capital gains). The quantities raised by this special tax can be seen in the *Statistics of Income for 1945*, table 19 (page 219).
 - o Since 1933, capital gains are only included at some percent of their value, which varies according to the period of generation (similar regulation as in Sweden after 1951). See *Statistics of Income for 1945*, page 55 and ff.; this is also explained in Piketty and Saez (2003), appendix. In practice, according to Piketty and Saez (2003), “the vast majority of capital gains always falls under the most favorable tax regime”, so capital gains would mostly be present at only 60-40% of their value. Similarly to above, we do not compute the tax privilege arising from an “alternative tax” that was in place between the taxyears 1939 and 1944. (See *Statistics of Income for 1945*, p. 251).

- The series including capital gains is, on average, 2.27% higher than the one without capital gains through the period 1913-45, but it attains a maximum of 9% in 1928.
- Total income in 1944 seems to be overestimated, since the average for non-filers results above the threshold (1,000 \$). We have adjusted it to an average of 810, as in the previous year. This would be a problem arising from the transition to the use of “Adjusted Gross Income” in the US regulation, and the change of method of calculation of reference total incomes in Piketty and Saez (2003). Taxyear 1944 (income year 1943) is the last one where Piketty and Saez use a top-bottom approach. Of course, the problem could also arise because of non-filing (i.e. evasion) in the context of a rapid extension of the obligation.
- Income distribution data is taken from the *Statistics of Income for 1945* for the years up to 1946 (table 20, page 221 and ff.), and the *Statistics of Income...* of each year for 1947-50.
 - We always use distributions of all returns with income (including non-taxable, i.e. those who ended up paying no tax because of the effect of family allowances and other deductions).
 - Data for 1942-44 is completed with each year’s publication, in what respects to tax returns filed under an optional form below 3000 \$ (form 1040A). Sources: SOI 1941, table 13 (page 200), SOI 1942, table 13 (page 218), SOI 1943, table 9 (page 208).
 - For the years 1926-32, the filing obligation for individuals was set at 1,500 dollars, while the statistics present a first bracket of 1,000-2,000. Some individuals with gross incomes (net incomes in the US contemporaneous fiscal terminology) under 1,500 dollars had to file returns if their incomes *before deductions for costs of obtainment* were above a certain level. But most of those with incomes between 1,000 and 1,500 would not file returns. We have therefore assumed that the given distribution starts at 1,500 instead of 1,000, in order to be able to use this bracket for the disaggregation procedure and avoid losing information.
- Types of taxpayers:
 - Married couples could file separate returns in the US, which up to 1948 was of interest for those with two income earners, because there was one single

schedule for all returns (Piketty and Saez, 2001, appendix). This issue was tackled in Piketty and Saez's estimation of top income shares (footnote 59 in working paper version), but not in their series of total tax units in the economy, which is calculated as total adults minus married women.

- The *Statistics of Income* provide information about the number of separate returns of married women in each year. They were between 1 and 3% of all returns during most of the period, attained an isolated maximum of 7.35% in 1932, and then lied between 4 and 5% in 1944-48. This entails a missadjustment between the “total tax units” number and the “actual tax returns”, which would imply that we would have less simulated tax units below the threshold, potentially overestimating their average incomes. Therefore, we correct this by adding the number of separate returns of married women to the “total tax units” series of Piketty and Saez (2003).
- However, the fact that separate returns are made by relatively high-income couples means that the distribution of “tax returns” with which we work will be less unequal than the distribution of “households”, or “tax units” in the Piketty and Saez (2003) definition. This is a bias we are forced to acknowledge, but it is limited by the extent to which separate returns were actually made (see point above).
- Family allowances and estimation of taxable income:
 - Family structure data (share of households corresponding to each of the eight family types) is taken from the data on the surtax exemptions in *Statistics of Income for 1945*.
 - Family allowances come from Internal Revenue Service, Historical Tables.
 - Regarding the personal exemptions in 1945-46, the sources say: “For 1944-1945 [our tax years 1945-46], the personal exemption amounts (columns 1-3) were for “surtax” purposes only. The exemption for basic “normal tax” purposes was \$500 per tax return, augmented by the “earned income” of the spouse, up to \$500, on joint returns (SOI, Historical table 23, note 11). We do not have information on the structure of incomes within the couples, so we have simulated the exemption for couples in the normal tax as 700\$ (i.e., assumed an income of 200\$ from the spouse). This was the alternative that provided a better fit with the original data. In these years, deductions for dependents were only applicable to the surtax (see *Statistics of Income for*

1944, p. 7; also *Statistics of Income for 1945*, p. 32), which we incorporate into the calculations by simulating the normal tax and the surtax separately.

- Earned income tax credit: because of its limited character, we have not included the effects of an “earned income tax credit” which was in place, under different forms, between the tax years 1925 and 1944 (see the description in *Statistics of Income for 1945*, page 370).
 - In 1925-32 it was a tax credit (i.e. a reduction of the tax due, with some limitations), and amounted to an average of 4% of final income tax revenue (*Statistics of Income for 1945*, table 19, page 220).
 - In 1935-44 it operated as a tax allowance (i.e. a reduction of the tax base), but a much smaller one than basic and family allowances. In 1939, for example, it represented 11% of the amount deducted by the latter (calculated from data in *Statistics of Income for 1939*, pp. 8-9).
- Tax schedules
 - We have always included the surtax. Combined marginal tax rates for each bracket have been calculated with data from *Statistics of Income for 1945*, table A (p. 351-52) for the normal tax and table B (pp. 358 and ff.) for the surtax. Both taxes are simulated together, except for the period after tax year 1945 (because of difference in allowances).
 - The “defense tax” of 1940 is also included (it was 10 percent of the total “regular” tax, but limited to 10 percent of statutory “net income” in excess of the total regular tax).
 - Maximum effective tax rate limitations, which were in place between the tax years 1945 and 1964, have been taken into account (Internal Revenue Service, *Statistics of Income Tax Stats*, Historical Table 23).
- Deductions: in some years the sources present the distribution of gross income among income groups before subtracting relevant deductions (adjusted gross income), while in others it presents the distribution of gross income after having subtracted such deductions (what the sources for earlier years call net income). In order to operate with a consistent definition of gross income over time, and to include the impact of some quantitatively important deductions, we have performed an adjustment throughout. The type of deductions available in the tax code varies across the years, and so does their distribution in terms of income levels. We use data from Geloso et al. (2018), which provide the distribution for

most of the years, based on the original data in the *Statistics of Income* and some imputations (ref. their appendix). Their distribution is based on information for charitable contribution, interest, and taxes paid. The deduction for charitable contributions was introduced in the War Revenue Act of 1917, but already applied to the incomes of 1916 (tax year 1917), as discussed in Blakey (1917).¹⁸

- 1945-50: we regress the share of deductions in adjusted gross income, make an estimate of deductions for each tax unit, and subtract them from adjusted gross income before simulating taxes paid. For this, we used simple log-linear regressions (which have an R2 over 0.95). We complement the simulation with the standard deduction for tax units where deductions according to the equation were below the corresponding standard deduction. In 1945-56, the standard deduction was of \$500 when adjusted gross income was over \$5,000, and 10% of adjusted gross income if below this threshold.
- 1919-44: we regress the share of deductions in net income on income levels. The shares are calculated from the data in Geloso et al. (2018) (who provide the quantities deducted). We selected the equations with four interactions, which presented the best adjustment in terms of R2 and total estimated deductions. Deductions were estimated for each tax unit, imputing to those below the threshold the share estimated at the threshold. Estimated deductions are added to net incomes after calculating tax paid, to obtain a measure of gross income which is equivalent to AGI used by the official statistics after 1944.
 - Geloso et al. (2018) do not provide a profile for the taxyear 1926. We used in this case the profile of 1927, since in both years the same regulation was in place (Revenue act of 1926), and total deducted amounts are similar.
- 1918: we proceed similarly as in the surrounding years, with the difference that in this year one type of deductions (namely, contributions) were not subtracted from net incomes in the statistics (SOI 1945, p. 252, note 27 - to table 21). So, we have to simulate these contributions separately from other

¹⁸ “Another amendment, applicable to both 1916 and 1917 income taxes, provides for the exemption of gifts for charitable, religious, educational, and scientific purposes, to the extent of 15 per cent of the payer's taxable net income” (Blakey, 1917, p. 804).

deductions. After the imputation of taxes paid based on taxable income, we add all deductions to gross incomes, in order to obtain gross income equivalent to AGI (which will allow us to calculate the effective tax rates and indices of progressivity and redistribution in a consistent manner with other years). Geloso et al's data for this year seems unfit for this purpose, since it is based on the sum of these three components (contributions, taxes, interest), and the total adjustment leads to an overestimation of deductions.

- The distribution of contributions is taken from 1923 (the nearest year with these data) (SOI 1922, table 7, page 95). We regress the share of contributions over net income, impute these shares to the 1918 data, calculate the corresponding contributions, and then adjust proportionally to the total amount deducted for contributions in 1918 (245,080 thousands of dollars).
- The distribution of interest and taxes is taken from 1937 (in this case, 1933 was the nearest available year, but the resulting quantities were too high) (SOI 1938, page 32). We regress the share of these deductions over net income, impute these shares to the 1918 data, calculate the corresponding deductions, and then adjust proportionally to the total amount deducted for deductions in 1918 (885, 763 thousands of dollars).
- 1917: we use the distribution obtained for 1918, adjusted to an estimated total for 1917 of 521,719 thousands of dollars (based on the percent that deductions represented in net income in 1918: 8.3%). The total of “general deductions” given in the source is too high for our purposes, because it includes deductions for costs incurred in the generation of incomes, which in the next year were deducted from each source in the statistics (see SOI 1917, p. 13). Geloso et al do not provide data for this year.

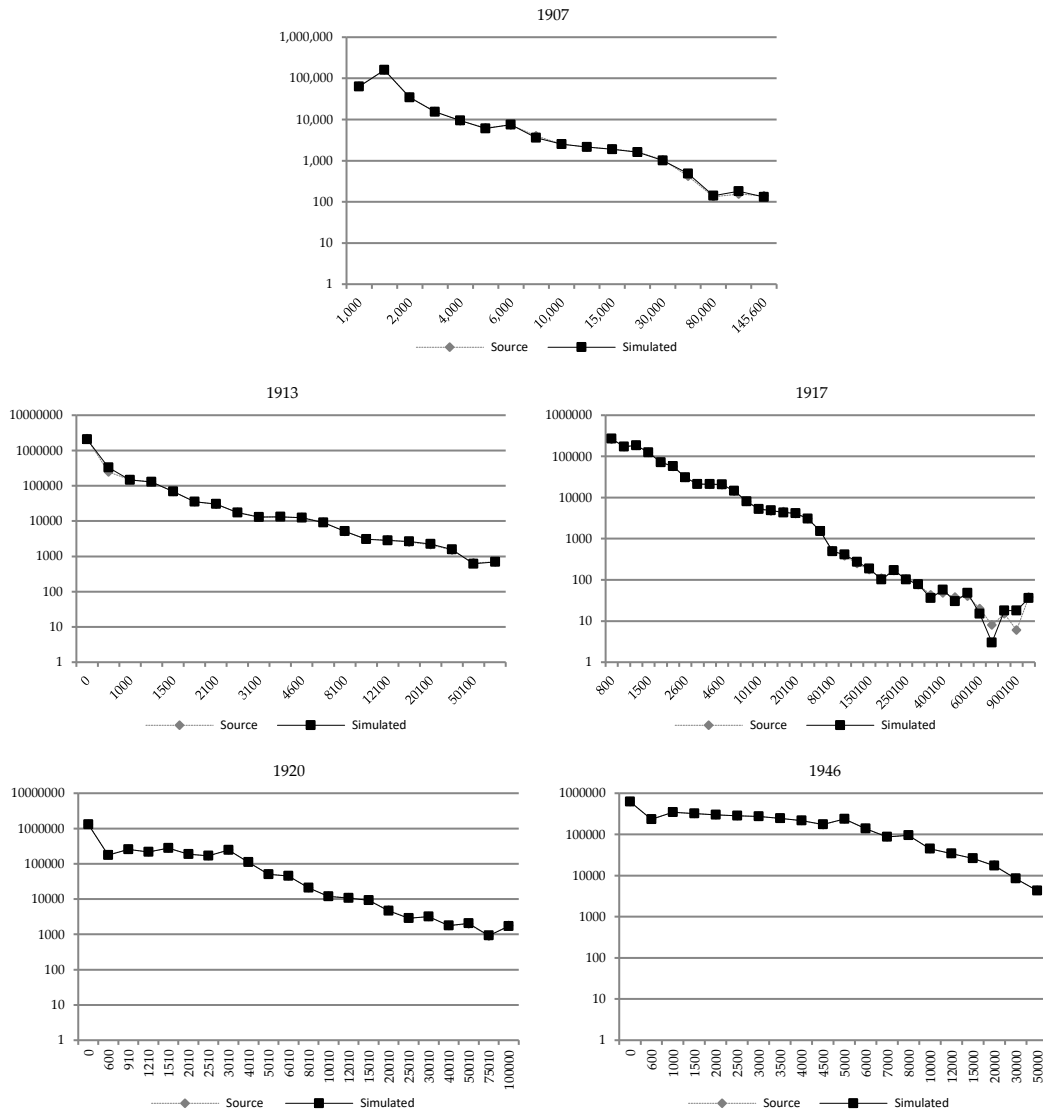
4. Accuracy of our synthetic samples and simulations

In this final section, we compare our synthetic samples with the original series provided by the historical sources to test their reliability. We first compare the number of units and average incomes in each income bracket (i.e., the accuracy of the results of the disaggregation procedure, using the brackets from the original sources). Secondly, we

look at whether our estimated average effective tax rates in each of these income brackets (calculated in the aggregate form) are consistent with the ones provided by (or directly obtainable from) the tax statistics.

Figures 1, 2 and 3 display the distribution of synthetic and original tax units by income brackets in the three countries (Sweden, United Kingdom and United States) for selected years. The grey lines (diamonds) reflect the distribution of tax units in the primary sources whereas the black lines (squares) depict our synthetic units. As mentioned above, the income brackets used in these figures are those employed by the tax statistics to illustrate the distribution of incomes and the operation of the tax.

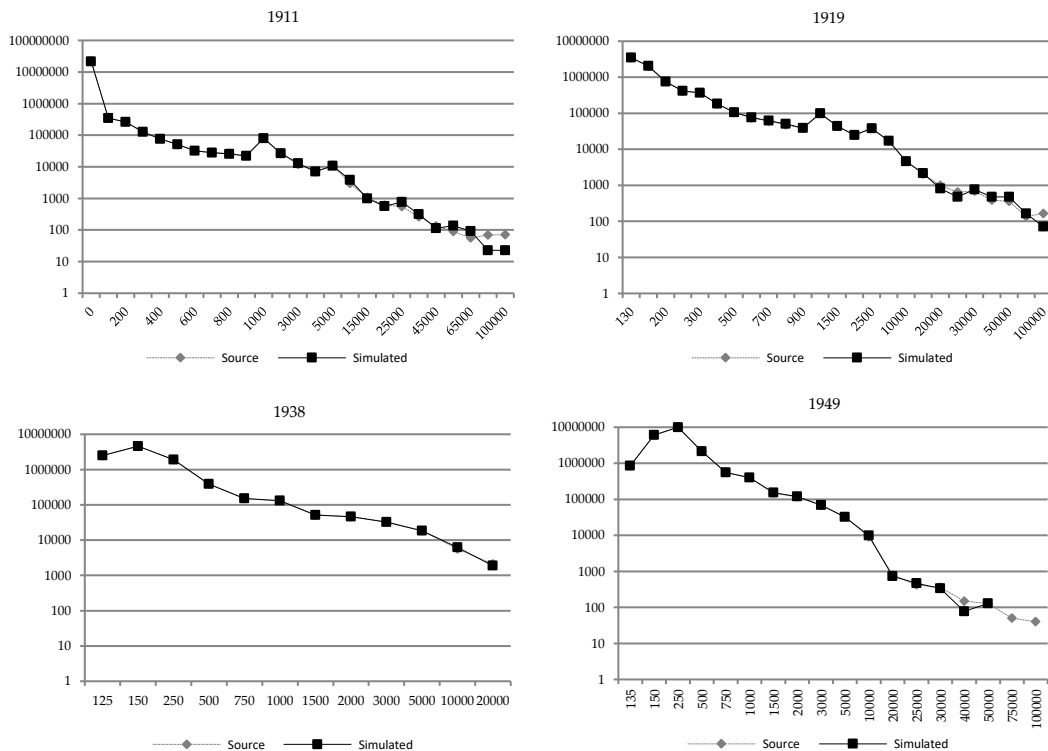
Figure 1. Number of tax units in Sweden, primary sources and synthetic sample



Sources: see text.

Overall, the figures suggest that the synthetic samples are highly consistent with the original data. The two lines overlap to a large extent in most income groups, which suggests that our samples are reasonable depictions of the true distributions of tax units in these countries and time periods. Most of the differences can be found in the upper end of the income brackets, but this is most likely driven by the very small number of people included in these top groups. For instance, the largest differences in the United Kingdom in 1911 can be found in some of the seven higher income groups (above £25,000 per year). These seven groups, however, account for a total of 1,230 tax units according to the primary sources (1,482 in our sample), merely 0.005% of total tax units. In some other cases our synthetic samples do not contain any tax unit in the very upper income groups. For instance, none of our units in the United Kingdom in 1949 earned more than £75,000, while according to the historical tax statistics 50 people earned between £75,000 and £99,999 (the second highest income group in the graph) and 40 more than £100,000 (the top income group). This, again, affects the trifle 0.0003% of the population of tax units.

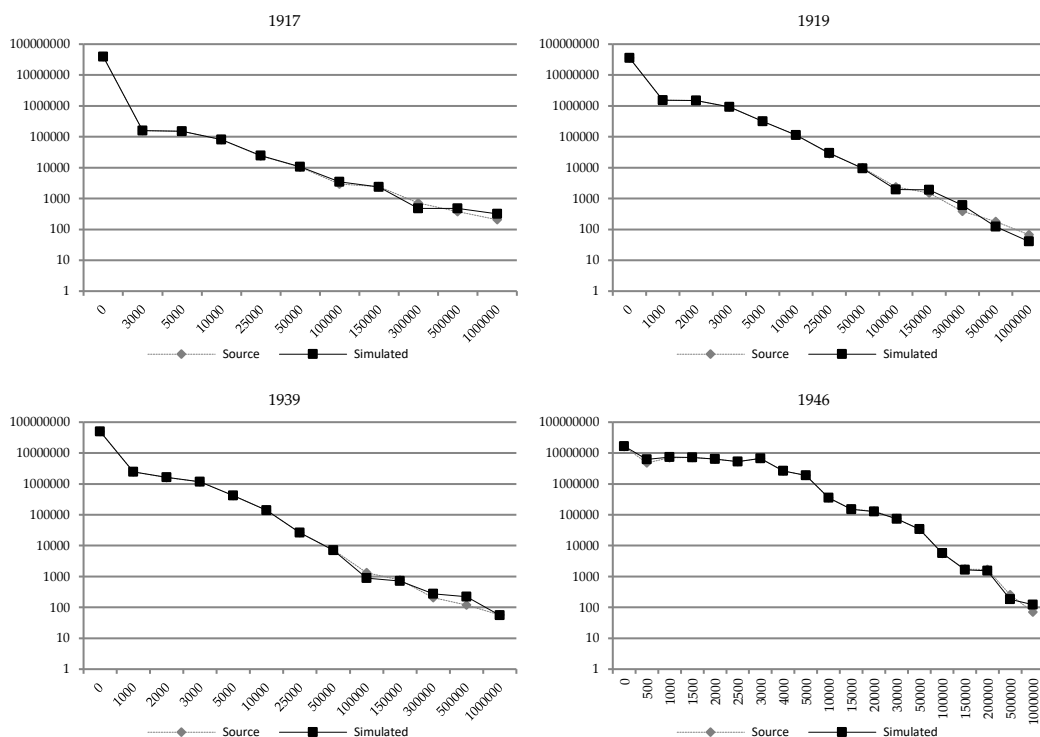
Figure 2. Number of tax units in the United Kingdom, primary sources and synthetic sample



Sources: see text.

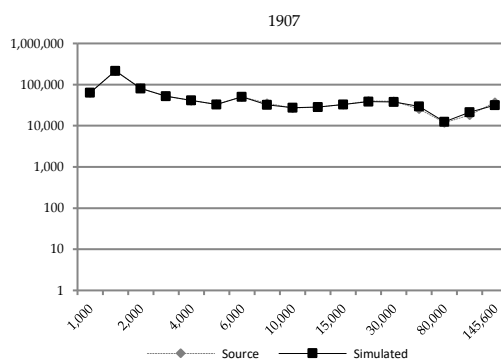
A similar picture can be found when it comes to incomes. Figures 4 to 6 depict the distribution of total income by the same income groups in the three countries and in the same benchmarks. Once again, the dashed and continuous lines (original and synthetic data) overlap to a large degree, except for the highest income groups.

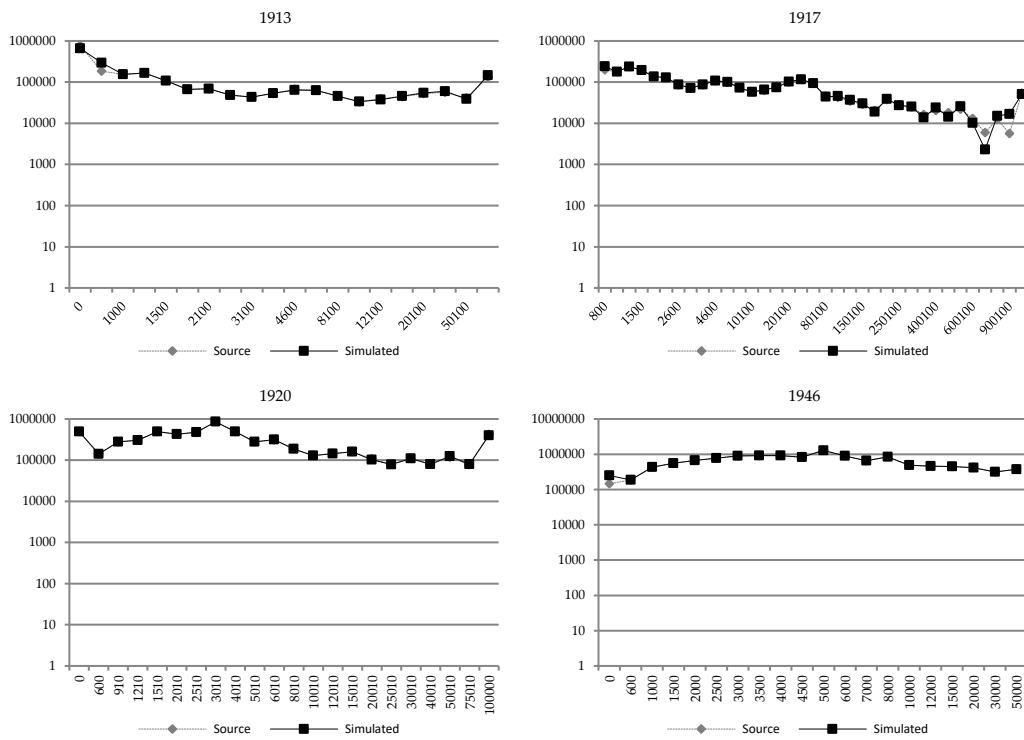
Figure 3. Number of tax units in the United States, primary sources and synthetic sample



Sources: see text.

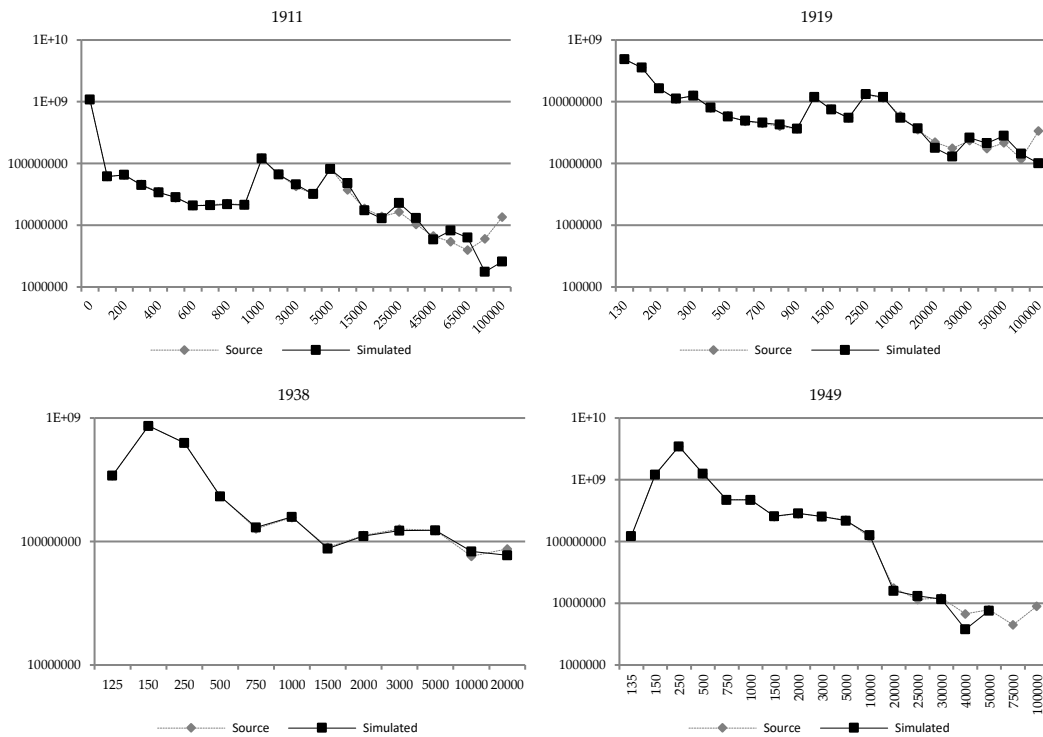
Figure 4. Incomes in Sweden, primary sources and synthetic sample





Sources: see text.

Figure 5. Incomes in the United Kingdom, primary sources and synthetic sample

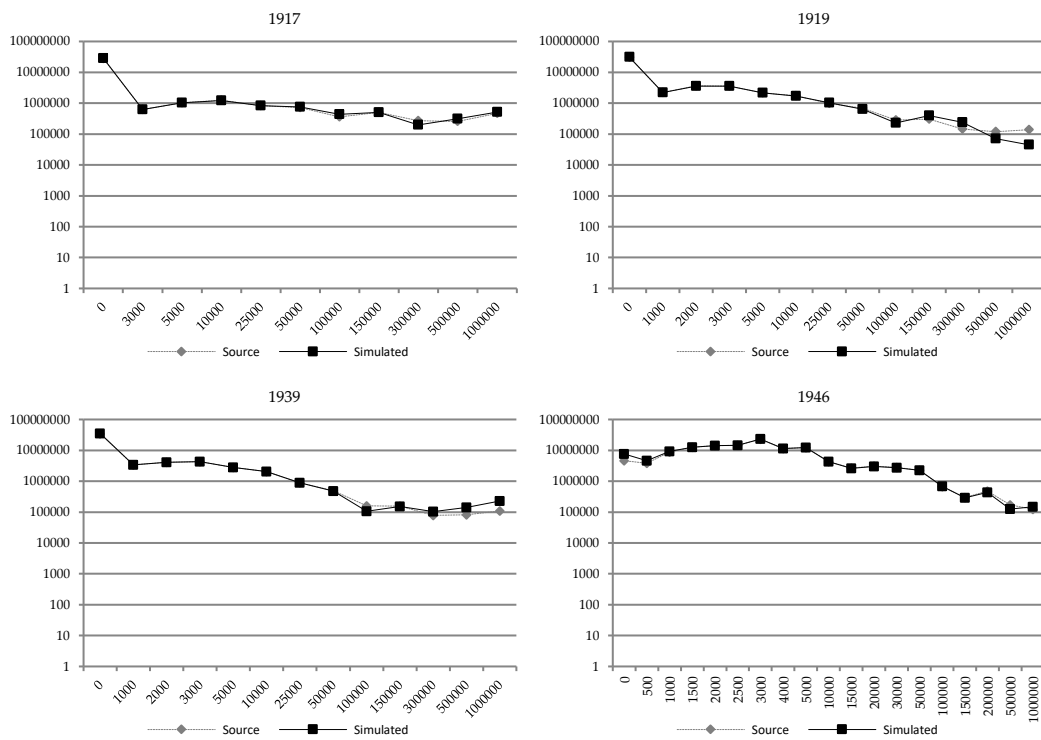


Sources: see text.

Finally, Figures 7, 8 and 9 look at the consistency of our estimated effective income tax rates with the effective tax rates reported by the tax statistics. To do so we first

disaggregate our synthetic distribution into the same income brackets, and then calculate the total tax due per each group and their corresponding effective tax rates. These rates can be compared with the actual effective tax rates provided in the sources. The degree of fit between the original and the simulated tax rates depends both on the accuracy of our distribution of tax units and on how well our calculations emulate the actual operation of allowances and tax rates. In general terms, the figures suggest that we can be confident about the consistency of our estimated rates. The two lines are very close to each other in most cases, particularly in Sweden and in the United Kingdom.

Figure 6. Incomes in the United States, primary sources and synthetic sample

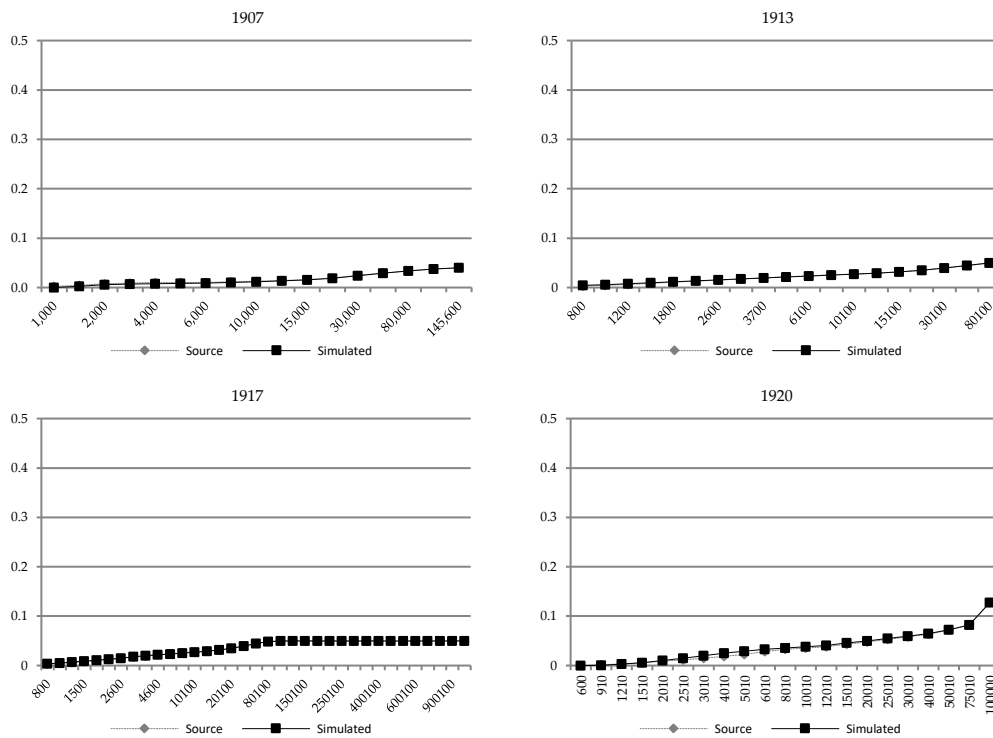


Sources: see text.

Nevertheless, we can observe relevant differences in some country-years. To begin with, synthetic tax rates for low- and middle- income groups are higher than in the original series in several cases, most notably in the United Kingdom before and after World War II and in the United States during World War II. This is most likely the result of the operation of deductions that benefited these income groups but that we cannot take into account in our calculations (due to lack of information). For instance, in the United Kingdom deductions were granted to taxpayers depending on their consumption habits (such as the hiring of housekeepers and for life insurance premiums) or their age. These

differences are generally small and do not affect our main results and conclusions. Only in the United Kingdom in 1938 they lead to a significant overestimation in the number of taxpayers: while in the original source the percentage of taxpayers over total tax units reaches 15.2%, our estimation brings this figure up to 22.9%. This is most likely due to the large concentration of tax units around the exemption limit (£125) that becomes highly sensitive to small changes in effective tax rates.

Figure 7. Effective tax rates in Sweden, primary sources and synthetic sample

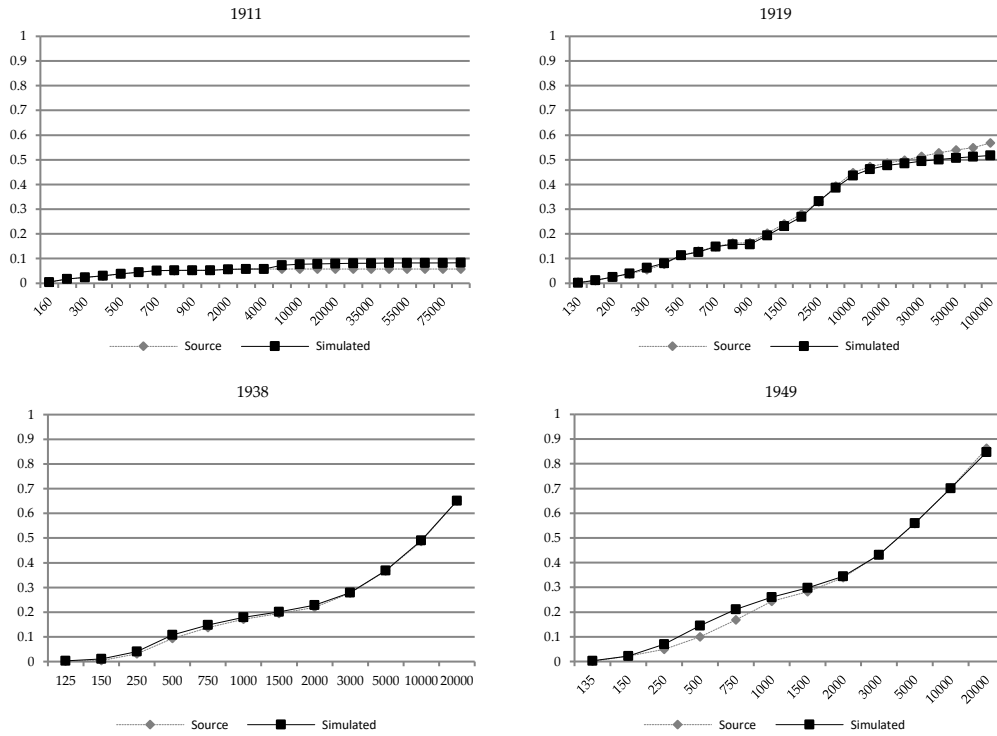


Sources: see text.

Note: the Swedish sources do not provide information on tax paid by income level in the 1940s, so we have not been able to perform this check.

On the other hand, the consistency of our effective tax rates for the highest income groups varies considerably across countries and periods. In the United Kingdom the estimated tax rates at the end of World War I are somewhat lower than the original ones. In the United States, by contrast, our estimates are systematically higher at the upper end of the income distribution (particularly in the post-war periods), which indicates that some additional deductions and tax preferences that we do not take into account benefited disproportionately the well-off (for example, this is expected of the reduced rate on capital gains).

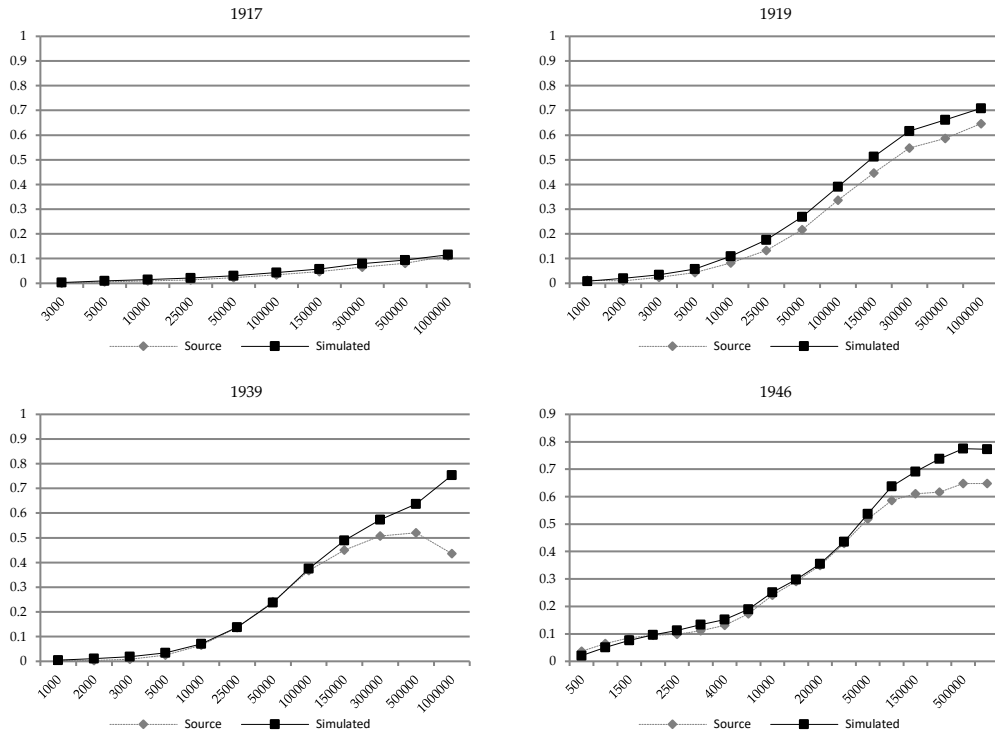
Figure 8. Effective tax rates in the United Kingdom, primary sources and synthetic sample



Sources: see text.

The original statistics we use for our three countries include non-taxable returns. These account for filers that ended up paying no income tax because personal exemptions and allowances reduced taxable income to zero. Including them in the denominator therefore biases the average effective tax rate downwards for the lower groups, and affects the comparison with our estimates (where the denominator only includes effective taxpayers). The quantitative impact of this, however, is very limited and cannot be observed in the graphs (with the possible exception of the United Kingdom after World War II, in which this downward bias effect could – together with the aforementioned allowances that have not been included in our calculations – explain the difference between the original and the estimated effective tax rates).

Figure 9. Effective tax rates in the United States, primary sources and synthetic sample



Sources: see text.

5. Range of error in Gini

In this section, we provide lower and upper bounds for the Gini indices of income. Our estimations of the Gini index are subject to a considerable margin of error when the taxpaying population represents a low percent of the total tax units, which was the case in the early 20th century.

We use the Stata package *giniinc*, provided and described by Hong et al. (2018), which estimates minimum and maximum Ginis consistent with the available data above the exemption threshold. Table 1 shows the results. The estimated ranges of error in the Gini are indeed very broad, with maximums near 90 and minimums near 20 in the early years, when the filing population was restricted to around the top 10% of the income distribution. When income taxes were extended downwards, more information allows reducing the range of error in the Gini: at near 40% of the population filing, the margin of error is close to 20 points (Sweden, early 1920s; US, early 1940s). Ginis are estimated quite precisely in the more recent years, when filing was around 70%.

The Kakwani index would be sensible to the high uncertainty in the Gini, but not so much the Reynolds-Smolensky index (where the same uncertainty is present in two Gini indices which are subtracted from each other). Effective tax rates are not affected by this problem. We can be quite confident about our progressivity and redistribution estimates for the World War II period, while somewhat less for the World War I ones (particularly in the case of progressivity). For that reason, in Torregrosa-Hetland and Sabaté (2022) we complement the Kakwani and Reynolds-Smolensky indices with the reduction in ratios between different levels of income.

Table 1. Ginis in the synthetic data, together with the minimum and maximum Gini using giniinc

<i>Sweden</i>			
Year	Gini (synthetic data)	Gini min	Gini max
1907	64.88	15.28	88.62
1909	59.70	16.15	87.37
1912	56.07	23.02	77.81
1913	56.83	23.62	78.06
1917	66.89	36.92	79.87
1920	57.45	53.39	74.25
1921	56.24	53.07	64.65
1936	56.18	47.77	60.92
1938	51.93	49.42	61.66
1941	55.25	49.60	57.85
1945	50.51	48.85	52.94
1946	49.30	48.04	51.45
<i>United Kingdom</i>			
Year	Gini (synthetic data)	Gini min	Gini max
1911	67.75	14.79	92.86
1919	42.58	27.37	77.45
1938	44.01	31.72	73.73
1949	42.72	39.04	48.58
<i>United States</i>			
Year	Gini (synthetic data)	Gini min	Gini max
1917	66.59	3.88	91.99
1918	33.07	20.33	91.03
1919	29.31	21.52	89.37
1920	27.04	25.06	88.75
1921	28.10	25.99	85.69
1922	38.82	22.77	86.42
1923	37.38	24.85	87.12
1924	31.69	25.85	85.64
1925	34.78	26.56	87.42
1926	41.06	19.76	90.95
1927	40.67	20.62	91.19

1928	40.35	18.85	91.48
1929	42.92	21.37	92.14
1930	39.06	20.75	91.97
1931	38.96	15.07	91.14
1932	43.00	10.71	90.52
1933	43.43	15.35	90.97
1934	45.12	12.92	91.14
1935	41.18	15.01	90.91
1936	37.45	17.07	90.66
1937	37.72	20.32	90.78
1938	35.43	21.97	89.59
1939	37.01	18.75	89.45
1940	39.43	21.68	87.80
1941	48.43	25.66	81.55
1942	50.04	31.56	70.43
1943	39.88	34.08	63.04
1944	41.49	36.90	59.26
1945	47.64	46.41	54.37
1946	47.97	47.09	54.67

Source: author's calculations using the Stata package *giniinc* by Hong et al. (2018), using a non-parametric approach.

References

- ABDELKRIM, A., and DUCLOS, J.Y. (2007): "DASP: Distributive Analysis Stata Package", PEP, World Bank, UNDP and Université Laval.
- ATKINSON, A. (2007): "The Distribution of Top Incomes in the United Kingdom 1908-2000", in A.B Atkinson and T. Piketty (eds.) *Top Incomes over the Twentieth Century: a contrast between continental European and English-speaking countries*. Oxford: Oxford University Press, pp. 82-140.
- BENGTSSON, E., MOLINDER, J., and PRADO, S. (2022): "Incomes and income inequality in Sweden 1870-1970", ongoing work, Lund/Uppsala/Gothenburg.
- BLAKEY, R. G. (1917): "The War Revenue Act of 1917," *The American Economic Review*, vol. 7, no. 4, pp. 791-815.
- DRIBE, M., and SCALONE, F. (2014): "Social class and net fertility before, during, and after the demographic transition. A micro-level analysis of Sweden 1880–1970", *Demographic Research*, Vol. 30, pp. 429-464.

- DU RIETZ, G., JOHANSSON, D., and STENKULA, M. (2015): “Swedish Labor Income Taxation (1862-2013)”, in Mangus Henrekson and Mikael Stenkula (ed.) *Swedish Taxation. Developments since 1862*, Palgrave Macmillan US, pp. 35-122.
- GELOSO, V., MAGNESS, P., MOORE, J. and SCHLOSSER, P. (2018): “How Pronounced Is the U-Curve? Revisiting Income Inequality in the United States, 1917-1945” (February 4, 2018). Available at SSRN: <https://ssrn.com/abstract=2985234> or <http://dx.doi.org/10.2139/ssrn.2985234>
- HENREKSON, M., and STENKULA, M. (2015): “Swedish Taxation since 1862: An Introduction and Overview”, in Mangus Henrekson and Mikael Stenkula (ed.) *Swedish Taxation. Developments since 1862*, Palgrave Macmillan US, pp. 1-33.
- HONG, L., ALFANI, G., GIGLIARANO, C., and BONETTI, M. (2018): “giniinc: A Stata package for measuring inequality from incomplete income and survival data”, *The Stata Journal* 18 (3), pp. 692–715.
- INLAND REVENUE (1917): 60th *Report of the Commissioners of His Majesty’s Inland Revenue for the year ended 31st March 1917*, London: HMSO.
- INLAND REVENUE (1918): 61st *Report of the Commissioners of His Majesty’s Inland Revenue for the year ended 31st March 1918*, London: HMSO.
- INLAND REVENUE (1920): 63rd *Report of the Commissioners of His Majesty’s Inland Revenue for the year ended 31st March 1920*, London: HMSO.
- INLAND REVENUE (1922): 64th *Report of the Commissioners of His Majesty’s Inland Revenue for the year ended 31st March 1921*, London: HMSO.
- INLAND REVENUE (1946): 83rd *Report of the Commissioners of His Majesty’s Inland Revenue for the year ended 31st March 1940*, London: HMSO.
- INLAND REVENUE (1952): 94th *Report of the Commissioners of His Majesty’s Inland Revenue for the year ended 31st March 1951*, London: HMSO.
- INTERNAL REVENUE SERVICE: Historical tables, table 23, “U.S. Individual Income Tax: Personal Exemptions and Lowest and Highest Bracket Tax Rates, and Tax Base for Regular Tax”, <https://www.irs.gov/statistics/soi-tax-stats-historical-data-tables>.

- KAKWANI, N. (1977). "Measurement of tax progressivity: An international comparison", *Economic Journal*, 87: 71-80.
- LAMBERT, P. J. (2001): *The distribution and redistribution of income*, Manchester: Manchester University Press.
- PEICHL, A. and VAN KERM, P. (2007): *Progres: Module to Measure Distributive Effects of an Income Tax*. Statistical Software Components S456867. Boston, MA: Boston College Department of Economics.
- PIKETTY, T. (2001): *Les hauts revenus en France au XXe siècle. Inégalités et redistributions, 1901-1998*. Ed. Grasset.
- PIKETTY, T. and SAEZ, E. (2003): "Income Inequality in the United States, 1913-1998", *Quarterly Journal of Economics*, 118(1), 1-39.
- QUENSEL, C.-E. (1944): *Inkomstfördelning och skattetryck*, Lund: Sveriges industriförbund.
- ROINE, J., and WALDENSTRÖM, D. (2010), "Top Incomes in Sweden over the Twentieth Century", in Anthony B. Atkinson and Thomas Piketty (Eds.), *Top Incomes: A Global Perspective*, Oxford: Oxford University Press.
- ROYAL COMMISSION ON THE INCOME TAX (1920): *Report of the Royal Commission on the Income Tax*, London: His Majesty's Stationery Office.
- SCOTT, P., and WALKER, J. (2020): "The Comfortable, the Rich, and the Super-rich. What Really Happened to Top British Incomes during the First Half of the Twentieth Century?", *Journal of Economic History*, 80(1): 38-68.
- SHORROKS, A., and WAN, G. (2008): "Ungrouping income distributions: Synthesising samples for inequality and poverty analysis". Research Paper UNU-WIDER, No. 2008/16.
- SOCIALSTYRELSEN: *Sociala Meddelanden 1942*, nr. 10, Stockholm.
- STATISTICS SWEDEN (various years): *Taxeringen till inkomst- och förmögenhetsskatt år ...*, Stockholm.
- (1929): *Statistisk årsbok för Sverige 1929*, Stockholm.
- (various years): *Skattetaxeringarna samt inkomstfördelningen inom yrkesgrupper taxeringsåret ...*, Stockholm.

- TORREGROSA-HETLAND, S. and SABATÉ, O. (2021): “Income taxes and redistribution in the early 20th century”. *Lund Papers in Economic History*; No. 2021:224.
- TORREGROSA-HETLAND, S. and SABATÉ, O. (2022). “Income tax progressivity and inflation during the world wars”. *European Review of Economic History*, 26(3), 311-339.
- US CENSUS BUREAU (1950): *Statistical Abstract of the United States: 1950*, Washington DC.
- US TREASURY DEPARTMENT, BUREAU OF INTERNAL REVENUE (1951): *Statistics of Income for 1945. Part I*, Washington, DC.
- WALDENSTRÖM, D. (2017). Wealth-income ratios in a small, developing economy: Sweden, 1810–2014. *The Journal of Economic History*, 77(1), 285-313.

Lund Papers in Economic History
ISRN LUSADG-SAEH-P--21/223-SE+31

© The author(s), 2021

Orders of printed single back issues (no. 1-65)
Department of Economic History, Lund University
Postal address: P.O. Box 7083, S-220 07 Lund, Sweden
Telephone: +46 46 2227475
Telefax: +46 46 131585

Full-text electronic issues (no. 58, 60, 61, 66--)
www.ekh.lu.se

Lund Papers in Economic History are published by the Department of Economic History, Lund University, Sweden. This series replaces the former series under the title *Meddelande från ekonomisk-historiska institutionen, Lunds universitet*. The change of name reflects the orientation of the series towards an international readership. The series is multilingual, but the majority of the working papers appear in English.

Lund Papers in Economic History are published as occasion arises, not at fixed intervals. Printed issues are distributed to libraries. From 1999 and onwards, full-text electronic issues are also available on www.ekh.lu.se. Those who would be interested in receiving information by email on new issues of **Lund Papers in Economic History** are requested to send an email message to Lund.Papers@ekh.lu.se.



LUND
UNIVERSITY