Estimating long-run income inequality from mixed tabular data: Empirical evidence from Norway, 1875-2017

Rolf Aaberge, Statistics Norway

Anthony B Atkinson, Nuffield College, Oxford

Jørgen Modalsli, Oslo Business School at Oslo Metropolitan University and Statistics Norway

Abstract. Most evidence on the long-run evolution of income inequality is restricted to top income shares. While this evidence is relevant and important for studying the concentration of economic power, it is incomplete as an informational basis for analysing inequality in the income distribution as a whole. This paper proposes a non-parametric approach for estimating inequality in the overall distribution of income on the basis of tabular data from different sources, some in a highly aggregated form. The proposed approach is applied to Norway, for which rich historical data exist. We find evidence of very high income inequality from the late nineteenth century until the eve of World War II, followed by a rapid equalization until the 1950s. Income inequality remained low during the post-war period but has increased steadily since the 1980s. Estimates of a measure of affluence demonstrate that overall inequality has largely been governed by changes in the top half of the distribution and in in the ratio between the mean incomes of the lower and upper halves of the population.

JEL Codes: D31, D63, N33, N34

Keywords: distribution of income, long-run inequality, the Gini coefficient, Norway

Acknowledgement: We would like to thank Andrea Brandolini, Magne Mogstad, Agnar Sandmo, and Espen Søbye and two referees and the editor for helpful comments. Financial support from the Research Council of Norway, grant numbers 231625 and 261985, is gratefully acknowledged.

1. Introduction: Inequality in the long-run

The extensive country-specific top income studies initiated by Piketty (2001) and provided by Atkinson and Piketty (2007, 2010) gave rise to a broad public debate on the rising income inequality in OECD countries, although the results of these studies dealt exclusively with the evolution of top income shares. Indeed, most of the discussion on long-run income inequality concerns the increasing share of total income received by the top 1 and top 10 per cent of income recipients. This is a legitimate and important concern, as high top income shares reflect the fact that a disproportionate fraction of total economic resources is being controlled by a small minority of the population. However, an exclusive focus on the concentration of the top 1 and 10 per cent ignores the distribution of income among the remaining 99 and 90 per cent of the population and can contribute to misinterpretation of the long-run evolution in overall income inequality. The debate between Autor (2014) and Piketty and Saez (2014) on the driving forces behind the steep rise in income inequality in the US in recent decades underlines the importance of accounting for the rise in income inequality among "the other 99 per cent".

The main objective of this paper is to propose a recipe for how the inequality of the income distribution as a whole can be estimated on the basis of different sources of tabular data from historical statistical income publications, which are available in many countries. This is possible without making assumptions about the distribution of incomes within wage groups (as in the "social tables" approach, e.g. Lindert and Williamson, 2016 for the United States) or relying exclusively on annual tabulations covering the majority of the population as is the case for Denmark (Atkinson and Søgaard, 2016). Complete detailed tabulations, like those for Denmark, appear to be an exception. Most developed countries have, however, collected income taxes on a regular basis and as a minimum published various aggregated quantities in some periods and detailed tabulations in other periods. This paper demonstrates that such combined data provide sufficient information to obtain reliable estimates of the the Gini coefficient across time. The presence of various aggregated quantities of individual incomes for each year, allows estimation of points on the Lorenz curve and proves to provide a useful basis for estimating the Gini coefficient and any other rank-dependent measure of inequality. Thus, in constrast to most previous studies, we do not limit the description of the evolution of long-run income inequality to years where complete tabulations are available, nor do we rely on interpolation of observations between years. Indeed, the paper demonstrates that much can be learned even in cases where detailed annual tabulations do not exist. By combining data from different historical statistical sources, a more complete picture of the distribution can be attained than that obtained from central government tax records alone. Similar procedures can probably be applied to other countries, to examine whether the development found for Norway also applies to other institutional and geographical settings.

The starting point is the information provided by the detailed tabulations of incomes by ranges as used by Aaberge and Atkinson (2010) in their study of top income shares in Norway. Section 2 of this paper describes the tabulated data available from the published income tax records from 1875 (annually from 1892) and in the form of micro-data from 1967 onwards. The tax information is a rich source, but it varies in form from year to year, and is limited in coverage, as it excludes non-taxpayers. The incomplete coverage of the population by the tax data means that there is a challenge involved in seeking to measure overall income inequality, as represented here by the Gini coefficient. We meet this challenge by creating "upper" and "lower" bounds on the Gini coefficient. Section 3 gives an account of the data and assumptions that are used to arrive at these bounds. In particular, we rely on aggregate information from the municipal and central government tax records, which are available annually for a long period. Combining these data with assumptions about the relative positions of different groups allow us to narrow the bounds on the estimated Gini coefficient. To this end, we add a further source of evidence about incomes at the bottom of the scale: administrative data on the number of recipients of public assistance, and the average amounts received.

The mixed tabulated data provide detailed information on the upper part of the Lorenz curve even in the 19th century, but less information on the lower part of the Lorenz curve. As is well known, the Lorenz curve is an increasing *convex* function taking values between 0 and 1. For distribution functions that are skewed to the right (heavy right tail), the Lorenz curve will exhibit weak (approximately linear) curvature in the lower part and strong convex curvature at the very top. Thus, to obtain a reliable estimate of the Lorenz curve for right-skewed distributions, it is necessary to have access to detailed tabular data for the top of the Lorenz curve, whereas it is sufficient to know a few points of the Lorenz curve below the median. As will be demonstrated in Sections 3 and 4, such tabular data constitute an appropriate basis for using a non-parametric approach to estimate the Gini coefficient and any alternative measure of inequality that are explicitly expressed in terms of the Lorenz curve.

By deconstructing overall Gini inequality with respect to measures of affluence and poverty, we show in Section 5 that overall inequality is governed very much by what happens to inequality in the distribution of income in the top half of the distribution and to the ratio between the mean incomes of the lower and upper halves of the population, which means that the estimates of the overall Gini coefficient are less sensitive to assumptions made on how the income attributable to non-taxpayers is distributed. However, as demonstrated in Section 4 this does not mean that the evolution of the income shares of the top 1 or top 10 per cent provides a complete picture of long-run income inequality in Norway.

A second objective of this paper is to provide new insight into long-run income inequality in Norway. The results presented in Sections 4 and 5 show that income inequality was high until the end of the 1930s, with substantial changes during the First World War. The turning point and origin of the low post-war inequality came with the German occupation between 1940 and 1945. The decline in inequality continued until the mid-1950s and remained stable at a low level between 1953 and 1980, but has increased steadily but moderately since 1980. Section 6 elaborates on how our results on the long-run evolution of income inequality contribute to an understanding of economic development in Norway since the late 19th century. It is shown to be a rich story that can be considered in terms of episodes of change.

1.1. Relationship with previous research

Our paper offers evidence of changes in inequality of the overall income distribution over a period of almost 150 years, and shows that changes in the ratio between the mean incomes of the upper and lower halves of the income distribution and in the inequality in the distribution of incomes among the richest 50 per cent explain most of the changes in overall inequality. Apart from Atkinson and Søgaard (2016), who have had access to detailed annual income tabulations for the majority of the population in Denmark, previous research relies on less informative data and has mostly provided limited evidence on income inequality for selected years before 1945. Moreover, many of the scattered estimates of the overall distribution that do exist for earlier years are not comparable with modern series. The estimates for the United States provided by Spahr for 1890 and by King for 1910 (see Merwin, 1939) have been described in a review paper by Williamson and Lindert (1980, p. 91) as "eclectic size distribution guesses", with the conclusion that "it is better to pass over these". Williamson (1985) has produced estimates for the Gini coefficient for England and Wales, and Scotland, for selected years ranging from 1688 to 1915. None of these estimates can readily be linked to the modern series, but are made available in separate tables in a survey by Lindert (2000). Kuznets (1955) provided a comparative study by compiling income distribution estimates for a few scattered years for the United States, United Kingdom and Germany. Milanovic (2016, Chapter 2) collected evidence for several pre-industrial economies (based on social tables, wealth data and some income-based inequality series) and argues that inequality varies cyclically over time.

To our knowledge, there are three bodies of academic work that attempt to produce comparable estimates of overall inequality from the early twentieth century (or earlier) and onwards. First, Atkinson and Søgaard (2016) have estimated the Gini coefficient for Denmark for 1870 and from 1903 to 2010 based exclusively on annual detailed tax-based income tabulations, which emerge as an extraordinarily informative dataset compared to historic data from other countries. The Danish dataset suffers however, from a series break in 1970 when the unit of account changes from family to individual. Moreover, the Danish dataset only contains detailed tabulation for one year in the 19th century. Secondly, Vecchi (2017, p. 331) reports estimates of the income Gini coefficient for Italy between 1861 and 1931. The estimates of the early period are constructed by fitting a generalized beta distribution on household budget data, and these series has been supplemented with tax-based estimates for the later period. Thus, the overall series may suffer from weak comparability, whereas the estimates of the early period may depend heavily on the chosen parametric distribution. Thirdly, Garbinti, Goupille-Lebret and Piketty (2018) extend their previous series of top income shares in France by including estimates for the bottom 10 per cent as well as the 10-50 per cent from 1900 until 1985 (and a more complete income distribution after 1985).[1] Atkinson et al. (2017) provide a review of historic income inequality estimates, including how data points from separate studies can be merged to construct long-run series of income inequality for the United States (from 1918) and the United Kingdom (from 1938).[2] These estimates suggest that income inequality decreased in the early twentieth century and increased from the early 1980s, but reliable results are still not available for sufficiently many countries to justify a general trend. The longest previous series for income inequality in Norway were reported by Soltow (1965), who constructed a series of Gini coefficients based on samples of tax records for selected years between 1850 and 1960 for eight cities in southern Norway. The results of Soltow (1965) show decreasing income inequality among taxpayers living in these cities.

The methodological approach of the present paper extends previous analyses by combining detailed tabulations of income tax data for a limited proportion of the population with income data from other sources. For most years, we have access to detailed tabulations for the top of the income distribution. These data are supplemented by annual aggregate data from two different taxation schemes (municipal and central government taxes) and from poverty statistics. As is demonstrated in this paper, the shape of the Lorenz curves for right-skewed income distributions makes it feasible to use a non-parametric approach to estimate the Lorenz curve and the Gini coefficient when detailed tabulations are available at the top of the income distribution and aggregate data provide estimates of a few points of the Lorenz curve for the lower half of the income distribution.[3] By contrast, Garbinti et al. (2018) rely on the condition of constant income shares for the lower 90 per cent, while Blanchet, Fournier and Piketty (2017) and Vecchi (2017)

[1] Before 1970, Garbinti et al. (2018) assume constant income shares within the bottom 90 per cent, e.g. the following shares are assumed to be constant for all years in the period 1900-1970 in France: the income of the lower 10 per cent is 0.39 per cent of the total for the bottom 90 per cent, the next 40 (10-50) 26.30 per cent and the next 40 again (50-90) 73.31 per cent. See Appendix Table TD3 to Garbinti et al.

[2] Kopczuk et al. (2010) provide evidence of earnings inequality in the United States from 1937 onwards based on social security data and Kuhn et al. (2017) have produced estimates of income inequality for the United States starting in 1949.

[3] An illustration of the shape of Lorenz curves for Pareto distributions is displayed in Appendix D.

rely on parametric distributions for broad intervals as a basis for estimating the overall income distribution.

The limited evidence on the evolution of overall inequality in the literature has been supplemented with useful information on the evolution of top income shares, not least owing to the top income books edited by Atkinson and Piketty (2007, 2010) and the extensive review provided by Roine and Waldenstrom (2015). Several studies combine results for the overall income distribution in modern times with estimates of top income shares for earlier years. For example, Piketty, Saez, and Zucman (2016) report top income shares and estimates of the functional income distributions for the United States back to 1913, but provide no information on overall income distribution before 1964. Some studies suggest that top income shares could be a good proxy for overall inequality. The evidence provided by Leigh (2007), Roine and Waldenström (2015) and Morelli, Smeeding and Thompson (2015) is however mixed. The call for prudence made by Morelli et al. (2015) is supported by the present paper. Actually, we demonstrate that top income shares might give a misleading picture of the evolution of overall income inequality, partly because changes in top income shares are normally accompanied by significant changes in the distribution of incomes in the upper half of the income distribution as well as by changes in the ratio between the mean incomes of the upper and lower halves of the population.

2. Income tax data in Norway

We begin with a brief account of Norwegian income tax data, and the way in which they can be used to produce results for income distribution as a whole. Since similar data are available for most countries, the method introduced below might be used as a recipe for estimating historical Gini series. This section is principally concerned with the years from 1875 up to 1951 when the published data are more fragmentary and vary in coverage. From 1952, the tabulations are more detailed, and from 1967 to the present we have access to micro-data. The income data originate directly from tax records (they are not inferred from taxes paid).

The same income definition, "antatt inntekt" (*assessed income*), is used over the entire period. It refers to income before tax (including capital income, taxable capital gains, taxable transfers and pension income), but after some pre-tax deductions. The pre-tax deductions represented relatively small amounts until the mid-1980s and were related to the expenses that were deemed necessary for the acquisition of income (Historical Statistics 1994, p. 280). For the period where excluding such deductions from our income measure might create a bias - the late 1980s onwards - we show by using an alternative measure of income that the pattern of the historic series is similar to the pattern based on a more comprehensive income measure for this period.[4] Self-employment income was accounted for by assessing the productive capacity of farms (in particular smaller farms) and deriving figures from company accounts.

The tax unit (nuclear family), which is either a married couple or a single individual, defines the unit of analysis in this study. This choice is dictated by the tax statistics, as married couples were taxed jointly until 2018.

2.1. The income tax data from 1875

The income tax sources are municipal (MUN) and central government (CG) tax assessments: *Kommunenes skattelikning* and *Statsskattelikningen*.[5] The key feature here is that, for a number of years, the government has published tabulations of the distribution of income tax payers by income range. The sources are listed in Appendix B. In addition, we have data on the total number of MUN and CG taxpayers for all years, starting in 1892, as well as the total income earned by each group. As the MUN tax data are more extensive (tax thresholds are lower and more people pay MUN than CG tax), we assume that CG taxpayers are a subset of MUN

[4] Liberalization of the credit market in 1984-1985 combined with the right to deduct interest expenses and high marginal tax rates on capital incomes until the tax system was reformed in 1992 encouraged households to borrow, which led to a significant rise in interest deductions. However, although these reforms might have weakened the comparability of the historic income data, it should be noted that the evolution of the Gini coefficient for income after tax since the mid-1980s as displayed in Figure 4 in Section 3 is consistent with the evolution found for the historic Gini series over this period.

[5] This information, and further information below, comes from Gerdrup (1998) and the Introduction to Part XIII of Historisk Statistikk (HS) 1968.

taxpayers. Given the similar tax base and the way these sources are treated in the tax statistics, this is a reasonable assumption.

The coverage of the detailed income tax tabulations varies over the period. CG tax was introduced in 1892, so there is only distributional information on MUN tax for the years prior to that. The published tabulations for 1892 to 1903 only relate to CG tax, and the same applies to 1938 and 1948-1951. To summarize, in decreasing order of completeness, over the period up to 1951:

(i) MUN and CG distributional data: 1906, 1913 and 1929;
(ii) MUN distributional data: 1875 and 1888;
(iii) CG distributional data: 1892-1903, 1938, 1948-1951.

We supplement the distributional data with the data on the total number of taxpayers and their total income, which is available for nearly all years. This means that, in addition to the Lorenz curve from the distributional data, we have in case (iii) a further point corresponding to the total MUN taxpayers (and hence total taxpayers).

The tabulations of taxpayers by income range from 1952 to 1966, which precede the micro-data available from 1967, vary in their coverage (see below). Income is equal to income as assessed by municipal tax assessment for the years 1952-55. In the tabulations for the years 1957 to 1966, income is defined as income as assessed by central government tax assessment if central government tax is levied. If not, income is defined as income as assessed by municipal tax assessment. There are no data for 1956 on account of the introduction of Pay-as-You Earn.

Since 1967, all individual incomes have been available on computer files at Statistics Norway. The income concept used is "antatt inntekt", income after some standard deductions, which is the same definition as that used in the pre-1967 tabulations. Using data from the Central Population Register, we merge married couples into single units, adding together the incomes of husband and wife to form the nuclear family. An adjustment is required for the data from 1960 to 1967 to account for changes in tax unit definitions, as explained in Appendix E.

In the period 1921-1947, corporate incomes (as well as individual incomes) are included in the aggregates in the tax statistics publications. From 1937 onwards, we can obtain figures for individuals from other sources; between 1921 and 1936 we make adjustments to the totals based on observed rates from 1937-1947.

2.2. Control totals

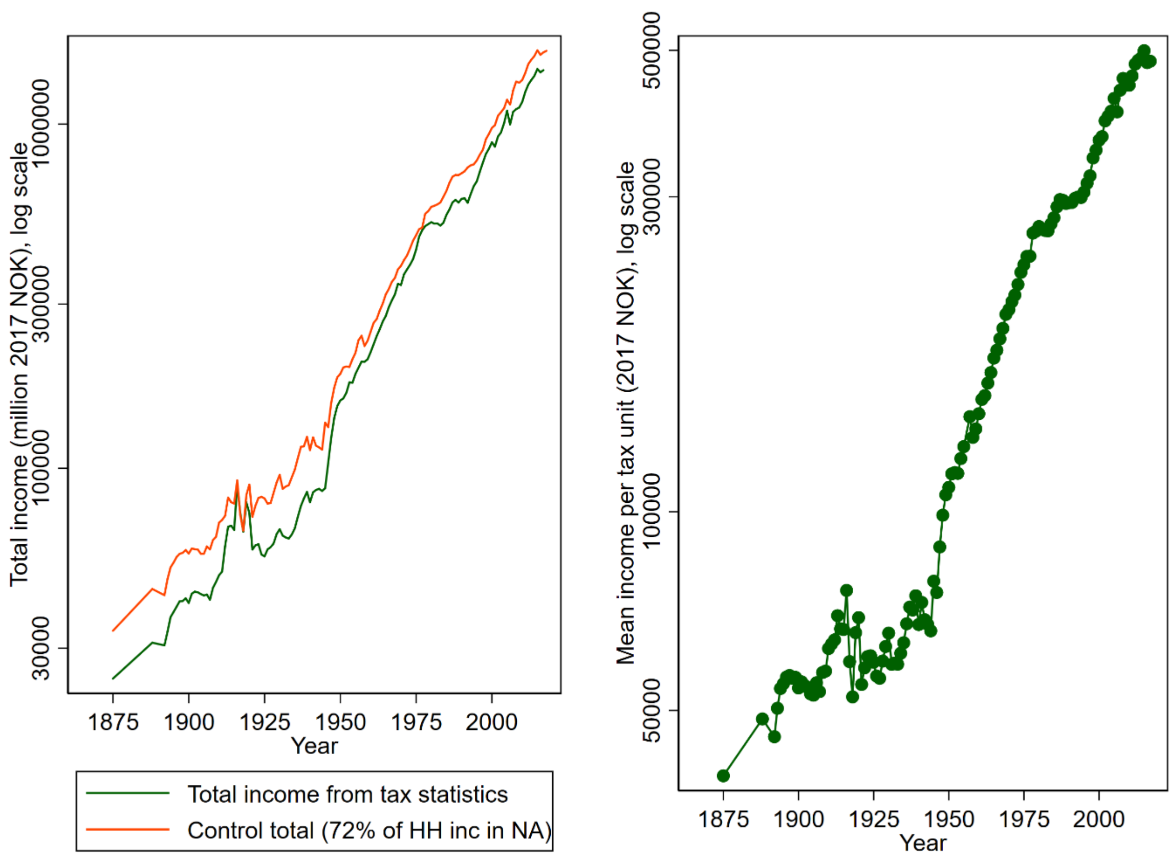
The CG and MUN income tax tabulations for the late 18th and early 19th century do not cover significant proportions of the population as a whole. In order to arrive at an estimate of overall income inequality for the entire population, rather than only for the taxpayers, this study uses estimates of total number of tax units and total household income as starting points. The sources of these "control totals" are described in Appendix C. The first step in calculating total tax units is the adult population, defined here as those aged 16 and over. The second step is to subtract the number of married women. Defined in this way, the tax unit population (nuclear families) as reported in the population statistics increases from 847,000 in 1875 to 1.7 million in 1951 and 3.4 million in 2017. In 1875, 83 per cent of nuclear families were covered by the tax statistics. In 1892, this figure had decreased to 52 per cent. The share subsequently increased gradually to 80 per cent by 1920. During the next period it decreased, to 66 per cent in 1933, and subsequently increased again, reaching 86 per cent in 1951. As explained below, after 1951 we rely on several different tabulations that together cover the entire population.

For total income, we use total household income for 1978 to 2017 from the National Accounts (NA) and extrapolate backwards using comparable historical series (see Appendix C). The resulting series for total household income as measured by the national accounts exceeds the total income recorded in the tax statistics (the internal total) in three main respects. First, the omission of the income of those not covered by the tax statistics. Second, understatement of income in the tax statistics. Third, differences in income definitions. Aaberge and Atkinson (2010) observed that the highest percentage for total NA household income recorded in the tax statistics was 72, and thus chose 72 per cent of NA household income as control total. We use the same approach. [6]

[6] Aaberge and Atkinson (2010, p. 476) provide further details and indicate that a similar approach has been used for Sweden and the United Kingdom.

Total household income as measured by the NA is made up of (i) compensation of employees (not including employers' social security contributions), (ii) operating surplus of self-employed businesses, (iii) property income, (iv) transfers from government and from abroad, and (v) income not classified elsewhere. A comparison of the control total from the National Accounts and the total from the tax statistics is given in the left panel of Figure 1 and in Appendix C (Table A5 and Figures A3-A4).

Figure 1. Total income from tax statistics and control total (left panel) and mean income per tax unit (right panel). In 2017 NOK



Note: "Control total" refers to the total income one would obtain based on official (reconstructed) national accounts data, with the caveats explained in the main text. "Tax unit" refers to married couples and single individuals.

The control totals provide estimates of the mean income per tax unit, which is displayed in real terms (as 2017 NOK) in the right-hand panel of Figure 1. In the period since 1875, real income has risen by a factor of around 13.[7] But the growth has not been steady. Before 1914 there was an irregular pattern of recessions and recoveries. The inter-war period saw little improvement in real incomes. The post- World War II period, in contrast, experienced rapid growth up to the mid-1970s, which later slowed and was interrupted by the recession and banking crisis of the late 1980s and early 1990s.

3. Estimating the Lorenz curve and the Gini coefficient

We now move to the estimation of the Lorenz curve and Gini coefficient based on the data on MUN and CG taxpayers as well as the control total. Given that the data are typically incomplete, we have to make assumptions and work throughout with an upper and lower bound Gini coefficient. By consistently choosing assumptions that lead to higher inequality for the upper bound and lower inequality for the lower bound, we are able to efficiently bracket the true Gini coefficient that we would obtain if we had full information on the exact incomes of all nuclear families, and also to obtain a measure of the precision of our estimates.

The discussion in this section will be based on the available Norwegian historical data sources described in Section 2. However, the existence of several types of income tax as well as data on social assistance is by no means unique to Norway in this period. For this reason, the methods proposed here, utilizing tabular data to assess points on the Lorenz curve, are also applicable to other countries.

3.1. Estimation of Lorenz curves

The Lorenz curve plots cumulative income shares (on the vertical axis) against cumulative proportions of the population (on the horizontal axis), with the population ordered from low-

[7] GDP per capita (in fixed prices) has grown by a factor of 18 over the same period. The discrepancy is largely due to the extensive demographic changes during this period; in 1985 Norway had a much younger population. The total population grew by a factor of 2.8 from 1875 to 2013, while total tax units (as defined here) grew by a factor of 3.6.

income to high-income individuals. This means that the Lorenz curve will always be a convex function below the diagonal, as illustrated in panel (a) of Figure 2. It is well-known that the Gini coefficient is defined by twice the area between the diagonal and the Lorenz curve. Hence, the bounds on the Lorenz curves constructed here correspond directly to bounds on the estimated Gini coefficients. A basic feature of the data used in this paper is that in all years, taxpayers amount to more than 50 per cent of the population, and that the total number of taxpayers and their income are reported annually. The annual aggregates from the municipal and central government tax statistics provide accurate estimates for several points on the Lorenz curve every year.

Different formats of the overall Lorenz curves are shown in panel (b) of Figure 2, which illustrates the case where we have distributional information on MUN taxpayers (with or without information on CG taxpayers) and panel (c) of Figure 2, which illustrates the case where we have only aggregate information on MUN taxpayers. In our estimates, we assume that the total population of tax units is correctly measured by our control total. The difference between this total and the total recorded in the income tax tabulations is referred to as the "missing population". Moreover, we assume that all individuals not represented in the statistics on MUN and CG taxpayers have incomes lower than those who pay tax. This means that the Lorenz curve for taxpayers is scaled down and connected with the final point for the missing population. In the case shown, the individuals in the missing population are all assumed to have identical incomes, so the first section of the Lorenz curve is a straight line. Further assumptions made about the distribution within the missing population are discussed below. Points H1 and H2 are points on the Lorenz curve constructed from MUN and CG taxpayer data. Panel (c) of Figure 2 shows the case where there is no tabulated MUN data, only aggregates. On the assumption that those paying MUN tax but not CG tax all receive the mean income, the Lorenz curve for this group is represented by the dotted line.

Figure 2. Estimation of Lorenz curves from tabular data

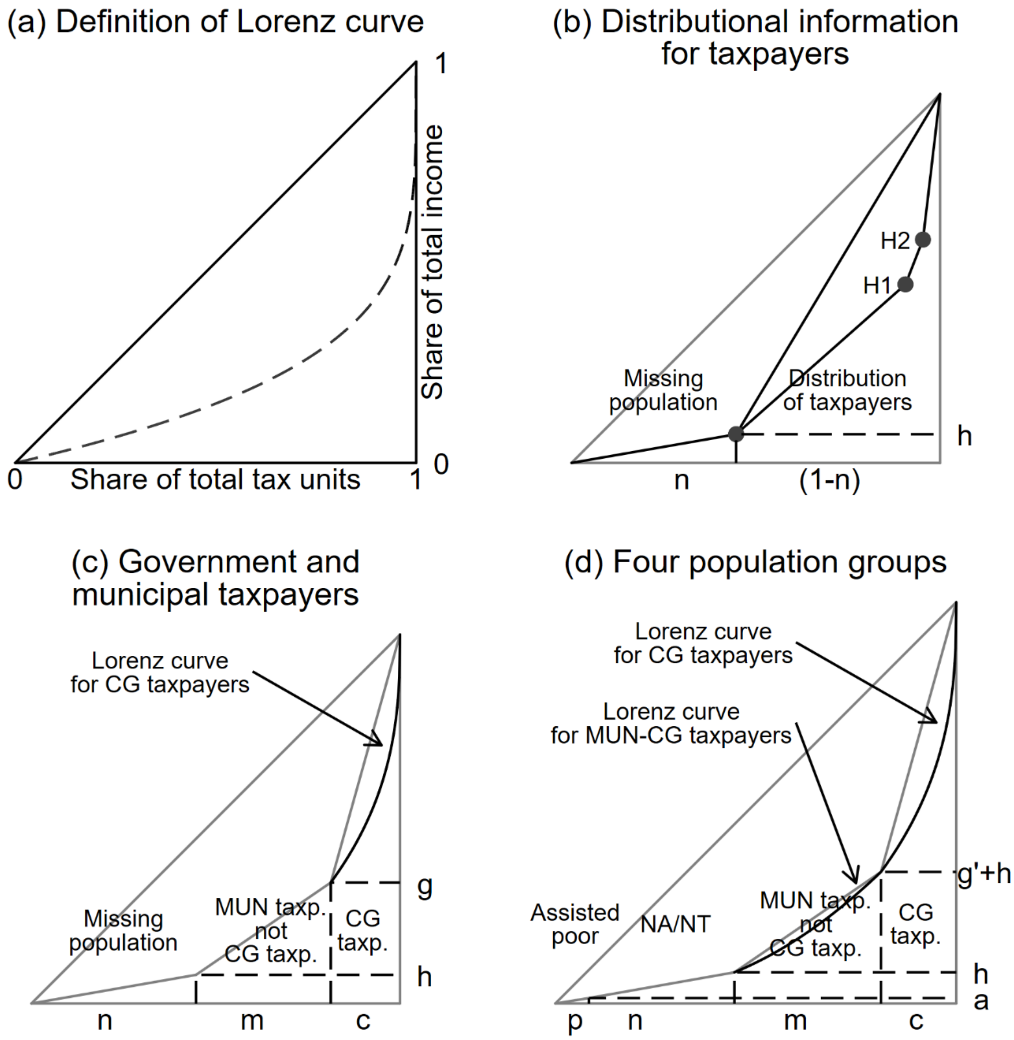


Figure note: Figure 2 shows how Lorenz curves are estimated on the basis of tabular data and assumptions of within-group distributions. Panel (a) shows the definition of the Lorenz curve as a plot of cumulative income shares against cumulative population proportions, where the diagonal line illustrates a hypothetical distribution with complete equality and the line tracing the lower and right corner of the figure illustrating extreme inequality (one tax unit owns the total income). Panel (b) illustrates how a Lorenz curve with distributional information for taxpayers (with two points H1 and H2 known from tabular data) is combined with the proportion of the population who pay tax to arrive at a Lorenz curve for the overall distribution. "Missing population" refers to the share of the population who are not covered by the income tax statistics. Panel (c) illustrates a Lorenz curve with two groups of taxpayers: those paying both central government (CG) and municipal (MUN) tax (denoted "CG taxpayers") and those paying only the municipal (MUN) tax. In this example, the distribution among the CG taxpayers as well as the proportion of the population not paying income tax are known. Panel (d) shows our baseline specification with four population groups, where the missing population is divided into those who receive poverty assistance (the assisted poor) and those who do not receive poverty asistance (the "non-assisted, non-taxed", NA/NT population). In this panel, we have applied within-group distributions for the central government taxpayers as well as for people who only pay municipal tax. For further details of estimation and definitions, see the text below.

The income attributable to the missing population is one element contributing to the difference between the income control total described above and the total income recorded in the tax statistics, where the latter is referred to as the "internal total". In the period 1875 to 1951, there was a difference of around 20 per cent between the internal and control totals (see Figure A3), apart from during World War I. In our estimates, total income is taken as equal to the control total. This means that we can consider bounds on the Gini coefficient in terms of allocating the difference either to under-reporting in the tax data or to the missing population. Suppose that the amount by which the control total exceeds the internal total is equal to a proportion, α, of the internal total, and that a proportion β of the internal total represents under-statement in the tax data. This leaves $(\alpha-\beta)$ times internal total income to be allocated to the missing population, or $(\alpha-\beta)/(1+\alpha)$ times overall control income. If non-taxpayers constitute a fraction n of the total population, then the amount allocated per head to the missing population, expressed relative to the overall mean, is $(\alpha-\beta)/[n(1+\alpha)]$. This would be the overall slope of the first segment of the Lorenz curve in this example.

3.2. Implications for the Gini coefficient

The implications for the Gini coefficient are most easily seen in terms of the area under the Lorenz curve, since the Gini coefficient is equal to 1 minus twice the area under the Lorenz curve. For taxpayers alone, twice the area is equal to

$$B = \Delta F_1 \mathrm{H}_1 + \Delta F_2 \{H_1+H_2\} + \ldots + \Delta F_k \{H_{k-1}+1\} \qquad (1)$$

where ΔF_i is the density in the range and H_i denotes the cumulative share of total income up to and including range i, where there are k ranges. It follows that the Gini coefficient for taxpayers alone is

$$G^* = 1 - B \qquad (2)$$

The introduction of the missing population as in panel (b) of Figure 2 has two effects. It squeezes the Lorenz curve for taxpayers to the right. In equation (1), this does not affect H_i but reduces ΔFi, and hence the area B, by a factor (1-n). The second effect is that it adds additional area under the first segment. If it is assumed that all incomes are non-negative, then the least such addition is zero (i.e. β is set equal to α), in the case where the Lorenz curve in panel (c) of Figure 2 initially follows the horizontal axis. Together, these two effects give an upper bound G_U for the overall Gini coefficient, which can be expressed

$$G_U = n + (1-n)G^* = G^* + n(1-G^*) \qquad (3)$$

It is a weighted average of 1 and G^*. In 1875, for example, values of n = 16.8 per cent and G^* = 47.6 per cent imply that the upper bound is 56.4 per cent.

Conversely, a lower bound might be sought by allocating all the difference to the missing population (β is set equal to 0), but this may violate the assumption that the missing population have incomes below the lowest income of taxpayers. Moreover, for some years there is contemporary evidence on which we can draw. For 1875, the tabulations published by Kiær (1892-3), which we are using, include an estimate of the numbers and income of the missing population.[8] The mean for the range NOK 0 – NOK 400 was NOK 230, which was 40.9 per cent of the overall mean. If, as an illustration, we attribute this amount per unit to the missing tax units, it means that, of the uplift moving from the NOK 345.5 million internal total to the NOK 475.8 million control total, 32.6 million NOK, or 28.3 per cent of the uplift, is allocated to the missing population.

The lower bound adopted here is calculated by considering the area under the Lorenz curve, where the missing population is allocated a fraction h of total income. Twice the area under the Lorenz curve is therefore increased by h times n. At the same time, the Lorenz curve for taxpayers is squeezed vertically by a scale factor (1-h), reducing its area but adding a rectangle, which adds 2h(1-n). The resulting lower bound Gini coefficient is

$$G_L = n + (1-n)G^* - h\left[1 + (1-n)G^*\right] = G_U - h\left[1 + (1-n)G^*\right] \qquad (4)$$

[8] Incomes below NOK 400 were exempt from taxation.

The last term shows that the difference between the upper and lower bounds – a measure of our uncertainty about the extent of income inequality in that year – increases, as we would expect, by the value of *h*, magnified by a factor of (1+ (1-*n*) G^*). The 1875 value of *h* = 8.6 per cent, coupled with *n* = 16.8 and G^* = 47.6 per cent, generates a difference of 9.6 percentage points from the upper bound, or a value of 46.8 per cent for the lower bound.

3.3. Using aggregate taxpayer data

For certain years, we have only the aggregate number and total income of the MUN taxpayers who are not liable for CG tax, and nothing is known about the distribution among this intermediate group. (We do however know the distribution among CG taxpayers.) This is the situation shown in panel (c) of Figure 2.

Let us denote the proportion of the population in the MUN-CG group by *m*, the proportion of CG taxpayers by *c*, and the proportion of those in neither group by n (so *c*+*m*+*n* = 1). The contributions of the three groups to the overall Gini coefficient may be seen in panel (c) of Figure 2. Denote the income share of the bottom group by *h*, and the combined share of the bottom two groups by *g*. Subtracting twice the area under the Lorenz curve from 1 gives the overall Gini coefficient:

$$G = 1 - \left\{ hn + (g+h)m + c\left[1 + g - (1-g)G^*\right]\right\} \quad (5)$$

where G^* is the Gini coefficient for the CG taxpayers. This may be re-written by introducing a new parameter *g′* = *g*-*h* and replacing *g* with (*g′*+*h*) as

$$G = 1 + c(1-g')G^* - \left\{ g'm + c(1+g') \right\} - h\left\{ 1 + m + cG^* \right\} \quad (5a)$$

The upper bound is obtained by setting *h* = 0 and keeping the other parameters constant. The final term in (5a) shows that the difference between *G* and the upper bound is proportional to *h*, with a magnification factor that is less than 3, but which may nonetheless be substantial. In 1892, the first year for which there is CG data, *m* = 36.6 per cent, *c* = 18.8 per cent and G^* = 44.8 per cent, so that the magnification factor is 1.45.

What, if anything, can we say about years for which there are no detailed tabulations for CG taxpayers? The formula (5a) allows us to see the role played by inequality within the group of CG taxpayers when $h = 0$. The term $c(1\text{-}g')G^*$ is an addition to the overall Gini coefficient. Suppose that we do not know G^*, but do know c and g'? Then the difference between the bounds would be widened to an extent that depends on the product of the population share and the income share of the CG taxpayers. Whereas the product may have been small in the nineteenth century, it was substantially higher in World War I and later. On the other hand, in the years for which we have tabulations, the Gini coefficient among taxpayers has rarely exceeded 50 per cent or (apart from two exceptions) fallen below 30 per cent.

3.4. Using data on the assisted poor

In order to provide a more solid basis for our treatment of the lower part of the distribution, we need additional information on the incomes of those below the tax threshold. In search of this, we explore one possible source: administrative data on the number of recipients of public assistance and the average amounts received. It is assumed that the recipient unit can be equated to the tax unit and that the poverty assistance is the same as subsistence market income. This means that all individuals/couples are assumed to have positive market income (where some could have zero); on the other hand, some of the recipients of assistance might also receive small amounts of market income.

In effect, using this additional administrative information means introducing into the three-group model a fourth group, by dividing those not paying tax into those who are assisted (the “assisted poor”) and those who are neither assisted nor taxed (NA/NT). The key assumption underlying our construction of the Lorenz curves and calculation of the Gini coefficient is that the groups can be ranked in order of increasing income, as shown in panel (d) of Figure 2. As liability for taxation depends on both income and wealth, there could be cases where people are liable for MUN taxation on account of wealth (and hence are included in the tax authorities' calculations of the number and total income of MUN taxpayers) but have low incomes that would place them below people in the NA/NT group. But it seems a reasonable approximation.

When the proportion of assisted poor is denoted by *p*, the proportion in the NA/NT group by *n*, and the share of the first group by *a*, the Gini coefficient is now given by

$$G = 1 + c(1-g')G^{*} - a(n+p) - g'(c+m) - c - h\left\{1 + m - p + cG^{*}\right\} \qquad (6)$$

The population proportions, *p*, *n*, *m* and *c* are known. The total income received by the assisted poor and by the two groups of taxpayers is known. The total income of the NA/NT group is not reported in the tax statistics. Here we have to make assumptions regarding the upper and lower bounds, but with the advantage that this group – given our earlier assumption – is "sandwiched" between two groups about which we have information. The upper bound is calculated on the assumption that the NA/NT group has the same average income as the assisted poor, the lower bound on the assumption that the average income of the NA/NT group is equal to one third of the average income of the MUN-CG group.[9] For some years, the MUN-CG mean income turns out to be less than three times the mean poverty support. In these cases, the imputed income for the NA/NT group will be the same for the upper and lower bounds.[10]

Expression (6) for the Gini coefficient does not account for possible dispersion within any of the three groups with lowest incomes. However, the POOR and NA/NT groups are always relatively

---

[9] A number of further adjustments have to be made to the published tabulations in making these 4-group calculations. Assumptions are required when estimating the upper and lower bounds. For G^{*}, if the within-group Gini of the CG taxpayers is not available, the upper bound uses the maximum of the previous and the next observations of G^{*}. Similarly, the lower bound uses the minimum of the previous and the next observation if there are no data available. For the years 1875 to 1891, when there were no CG taxation, the average income of the NA/NT group for the upper bound Gini is taken as NOK 150. NOK 150 was 25 per cent of the mean income of workers and 33 per cent of the mean income of farmers (including cotters) in 1888/89 (Sth. Prp. Nr 48, 1890).). Note that our "upper bound inequality" applies within the framework of assumptions outlined in this chapter. If, for example, we assume that the entire difference between total income from the tax statistics and national accounts was entirely "hidden income" that accrued exclusively to the rich, inequality would be higher.Given the nature of the tax system we do not find this assumption plausible. The lower bounds are assessed within the framework of the control total as described in Section 2.2.

[10] Alternatively, one could attribute zero income to recipients of poverty support on the grounds that one wanted the income definition to respect a strict "pre-tax" definition. A counter-argument is that the poverty support is likely to reflect the subsistence income received by these households. Changing the income level of the poor to zero (while maintaining the income levels for the NANT group) would increase the Gini coefficient by between 0.004 and 0.036. Results from this exercise are available on request.

small and, given our assumption that groups are ranked by income, limited by the income ranges of neighbor groups (or zero, in the case of the poor). This puts a strict upper limit on the contribution to the overall Gini that could result from within-group dispersion in these groups. For example, the maximum consistent inequality in the poorest group would mean that the richest individuals in this group had the same income as the NA/NT mean income and the poorest individuals in this group had zero income. The effect of such a distribution would be largest in 1888, where the lower bound Gini measured in percentage points would increase only from 56.91 to 56.93.[11]

On the other hand, the MUN-CG group constitutes a relatively large proportion of the population, and the data show that the differences between the MUN-CG and CG mean incomes are substantial. For this reason, within-group dispersion is introduced for the MUN-CG group. Specifically, the incomes within this group are assumed to follow a uniform distribution. The details of this imputation are given in Appendix D, where the relationship between the dispersion parameter *z* and the within-group MUN-CG Gini coefficient $G^{**} = z/3$ is explained. As we maintain the assumption that there is no overlap between the income groups, there is a limit to the upper value of *z*. Overall, a value of *z*=0.4 is consistent with the introduction of some dispersion without any MUN-CG taxpayers having either higher incomes than the lowest in the CG-group or lower incomes than the NA/NT group. Note, however, that the overall Gini coefficient proves to be insensitive to changes in *z*.

Finally, in 1875 and 1888 (the years before the introduction of CG tax in 1892) there was no state taxation, but instead detailed tabulations of the incomes of MUN taxpayers. We then assume that the lowest tabulated income group in the MUN tabulations is equivalent to the MUN-CG group in later years, and that the higher-income groups would have been subject to CG tax had that been in effect in these years.

---

[11] Graphically, we obtain the upper bound from 5 by extending the line for the NA/NT group (the slope of this group is the mean income of NA/NT relative to the population mean) down to zero. The resulting triangle (the contribution of the poor group to the overall Gini coefficient) is $\frac{a}{2} \cdot \left(p - \frac{an}{h-a}\right)$. Introducing dispersion to the NA/NT group would decrease the maximum consistent contribution from the poor group.

3.5. Estimation of benchmark series from 1875 to 1951

The comprehensive approach described in Section 3.4 provides the basis for our analysis of the long-run evolution of inequality in Norway in this paper. The Gini coefficient for the years 1875, 1888 and 1892-1951 will then be given by

$$G = 1 - pa - n(a+h) - m(2h+g') - c(1+g'+h) + c(1-g'-h)\ G^{*}) + g'm\,G^{**} \qquad (7)$$

where

a = total income of the poor relative to control total,

h = total income of the poor and non-assisted/non-taxed (NA/NT) relative to control total,

g' = total income of MUN taxpayers who are not CG taxpayers,

p = the poor as proportion of total tax units,

n = NA/NT as proportion of total tax units,

m = MUN-CG taxpayers (those who pay municipal tax but not central government tax) as proportion of total tax units,

c = CG taxpayers as proportion of total tax units,

G^{**} = Gini coefficient among MUN-CG taxpayers,

G^{*} = Gini coefficient among CG taxpayers.

Expression (7) takes as its starting point extreme inequality where the Lorenz curve follows the horizontal axis between 0 and 1. The first four terms then subtract the areas of the triangles and parallelograms below the Lorenz curve as illustrated in Figure 2. The latter two terms add in the within-group Gini coefficients for the two richest groups, scaled by group sizes and income shares. [12]

3.6. Bounds for 1952 to the present

The above discussion has described the bounds applied for the period 1875 to 1951. For the post-1951 period, when coverage was greater, relatively high numbers of tabulated intervals have been published by Statistics Norway (Historical Statistics 1978). From 1967 onwards the incomes of the entire population of taxpayers are available as micro data. For this reason, the set of necessary assumptions for this period is smaller, similar to the situation shown in Figure 2, where the assumptions relate only to the mean income of the missing population. These assumptions are designed to be comparable with those for the earlier period, while taking account of the changing role of assistance to the poor in the 1960s and later. In particular, there is a break in the poverty support series between 1964 and 1967, making mean payout per supported individual a less appropriate value for imputation at the lower end of the income distribution.

The upper bound of the Gini coefficient is based on the assumptions that (i) those not covered by the tax tabulations have a mean income equal to mean assistance (as before) for the years up to 1964 and (ii) from 1967 the group receives 50 per cent of the minimum pension for a single person.[13] The lower bound is based on the assumption that those not covered by the tax tabulations receive mean income equal to 150 per cent of the mean income assumed for the upper bound.

We should emphasize at this point that the final series is based on a consistent population throughout the period. Despite the change from household-based to individual-based taxation, we

[12] While the Gini coefficient is calculated directly from (7), we can also construct Lorenz curves using the assumptions outlined here. These are available as an online appendix. In this case, a Pareto distribution is imposed for the richest (CG) group, with the dispersion and lower bound parameters set to match the mean income and Gini coefficients of this group. As long as these two conditions are satisfied, the choice of within-group dispersion has no impact on the estimated Gini coefficient for the entire population or any part of the population that includes the entire CG group.

[13] For the years 1965 and 1966, the minimum pension was projected back from 1967 (when it was introduced) in line with the growth of seamen's pensions, which were introduced in 1950. The same process applied to 1964 yielded a figure of NOK 2,140, which was close to the poverty support level in that year of NOK 1,975.

can replicate pre-1960 nuclear families in the post-1966 microdata by merging spouses using personal ID numbers in the latter data that link taxpayers and the population recorded as individuals. The first year in which married women could choose to file taxes individually is 1960. We therefore transform individual data for the years 1960-1966 into household-based data using data from the 1960 census as well as the distribution of spouse's incomes, marriage and tax status in 1967. Similarly, adjustments are made to take account of a separate taxation system for sailors (1948-1966) and company taxation (1921-1947). These adjustments are all described in detail in Appendix D.

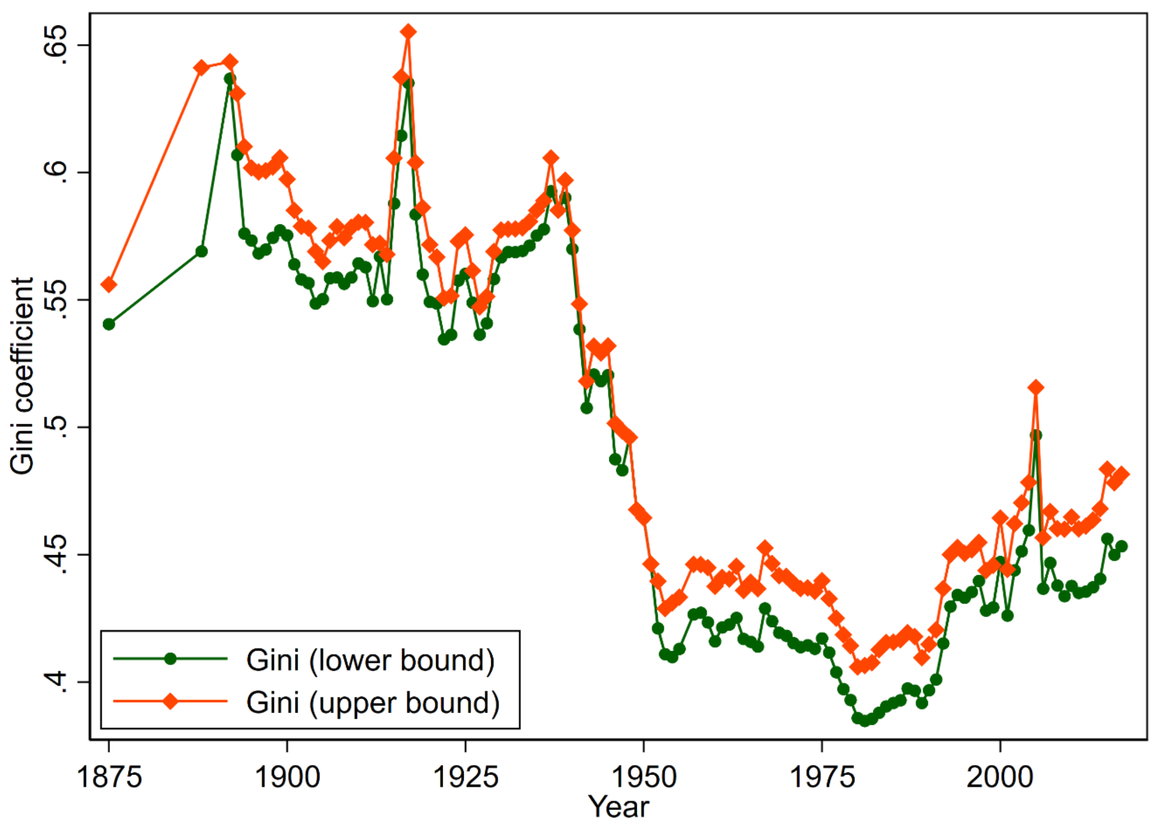
3.7. Long-run inequality in Norway

The results of the calculations discussed in Sections 3.5 and 3.6 are brought together in Figure 3, which shows the upper and lower bounds for the Gini coefficient. The difference between the upper and lower bounds is largest for the pre-1914 period. The average difference over the period from 1892 to 1914 is 9.8 percentage points, whereas the average difference from 1915 to 1951 is 2.2 percentage points. While the difference represents potential error introduced at the stage of data analysis and is not comparable with the sampling error typically considered in distributional analysis, it is nonetheless interesting to compare their magnitudes. From that perspective, the 1892 to 1914 figure appears quite large, but the 1915 to 1951 average difference is not dissimilar to the confidence intervals obtained from the reported standard errors for the Gini coefficients obtained from household surveys: for example, the 95 per cent confidence interval for the Gini coefficient of the distribution of disposable equivalent (household) income in Norway varied between 1.4 and 3.6 for the period 1986 to 1993.

In Appendix A, we perform an evaluation of the sensitivity of estimates of the Gini coefficient to the employment of additional data sources, and a robustness check of the Gini series based on two measures of inequality that complement the information provided by the Gini coefficient. The results displayed in Figure A1 show how the estimated Gini coefficients depend on the choice of data for interpolating the Lorenz curve, starting with the simplest approach in Section 3.1 and increasing the sophistication of the method to arrive at our preferred estimate. It is shown that the Gini coefficients produced by the naive estimator are far too low.

Irrespective of whether we use a measure of inequality that are particularly sensitive to changes that take place in either the lower or the upper part of the income distribution the evolution of the associated inequality estimates shows, as demonstrated by Figure A2 and Table A8, a similar pattern as the estimates of the Gini coefficient. However, the measure that is most sensitive to changes in the upper tail of the income distribution shows significantly larger relative changes than the Gini coefficient during the post-war period. Moreover, the Gini coefficient shows significantly larger relative changes than the measure that are most sensitive to changes that occur in the lower part of the income distribution. As will be demonstrated in Section 4.3, these results prove to be consistent with the information obtained by comparing the evolution of the overall Gini coefficient with the evolution of the ratio of the upper and lower mean income, the upper tail Gini and the measure of affluence, which is discussed in Section 4.

Figure 3. Gini coefficient for the distribution of income in Norway, 1875-2017. Upper and lower bounds



Note: For sources, methods and assumptions, see text. "Lower bound" refers to inequality estimated using assumptions on average income level of the non-taxed and distribution among central government taxpayers that lead to lower inequality, while "upper bound" refers to inequality estimated using assumptions that lead to higher inequality. See Subsections 3.1-3.6.

3.8. Different income definitions

The standard "official" estimates of the Gini coefficient for the distribution of income in Norway accounts for taxation, public cash transfers as well as for the needs of household members and has been published since the mid-1980s. As indicated in the introduction our choice of definitions has been dictated by constraints in available historic data sources. This is why we have adopted a gross income definition, whereas statistical agencies today provide inequality estimates on the basis of disposable equivalent income.

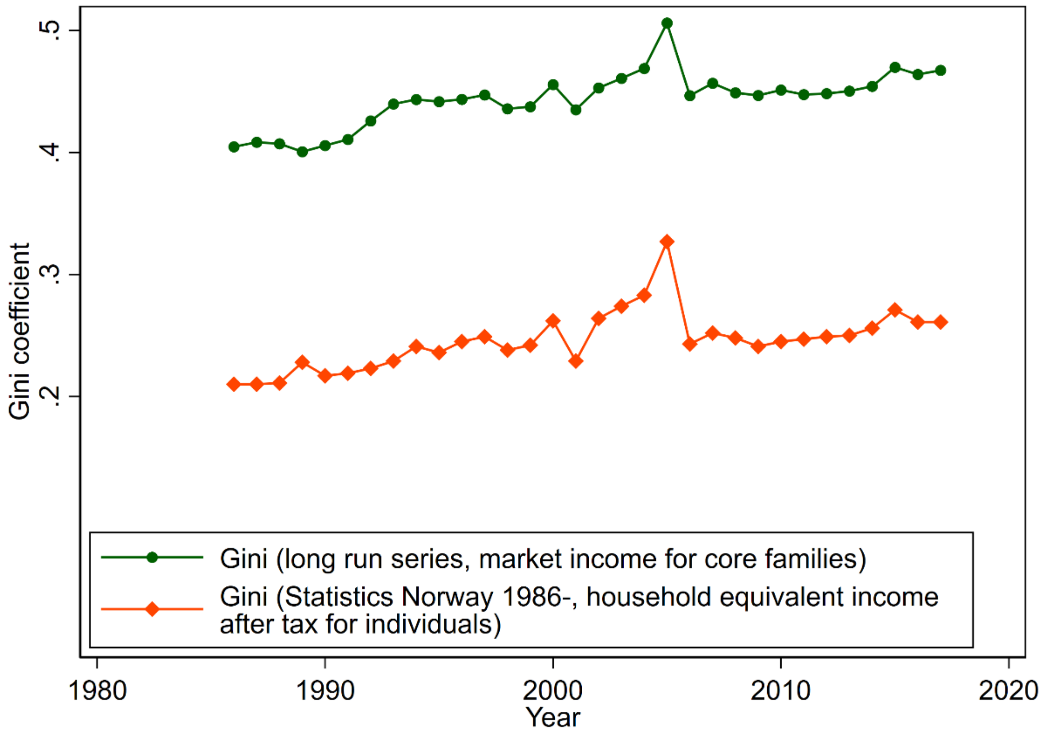
The closest Norway gets to an official definition of income inequality is Statistics Norway's time series from 1986 onwards.[14] The construction of this series diverges from the approach used elsewhere in this paper in three ways. First, the household definition includes everyone living together with joint consumption except students not living at home. To account for scale economics the standard EU equivalence scale is used. Second, a somewhat larger set of income sources (various types of non-taxable transfers) is included than the "gross income" concept used in this paper. Third, the income basis is post-tax rather than pre-tax.

Figure 4 compares the evolution of the Gini coefficient since 1986 for the two alternative definitions of income. As expected, inequality in the "official" series is much lower than the long-term series. This is largely due to the redistributive effects of public transfers and a progressive tax system, but it also reflects the treatment of the income unit. The use of a wider definition tends to reduce recorded inequality, since it assumes a greater degree of income-sharing. Taking account of economies of scale in larger households has also a significant effect on the measured level of inequality. However, since our focus is on the evolution of inequality, we find it reassuring that the pattern of the historic series captures the pattern of the official series from 1986 onwards. Most important here is that the development of inequality over time is similar for

[14] See http://www.ssb.no/inntekt-og-forbruk/statistikker/ifhus

the two definitions. There was a significant increase from 1986 to around 2000, turbulence around the tax reforms of the early 2000s and a slight increase thereafter.

Figure 4. Gini coefficient for the distribution of income in Norway based on two alternative income definitions, 1986 - 2017



Note: The long run series is the average of the upper and lower bound reported in Figure 3. For sources, methods and assumptions, see text.

4. The relationship between overall inequality and inequality at the top

Although Eurostat, OECD and national statistical agencies publish top income shares, ratios of income quantiles and decile means on a regular basis, such quantities cannot be regarded as measures of inequality as they don't satisfy the Pigou-Dalton principle of transfers. Thus, in order to provide information on overall inequality, these institutions regularly publish estimates of the Gini coefficient. Since most of the discussion of the long-run evolution of inequality in OECD countries concerns the increasing top income shares, it is interesting to explore what we learn from the new series of overall inequality (as in Figure 3) compared to the top income series previously published by Aaberge and Atkinson (2010). To this end, we compare the evolution of the income shares of the top 1 per cent and overall inequality, where we take the mean of the upper and lower bounds of Figure 3 to give an "average series".[15]

4.1. The share of the top 1 per cent and the Gini coefficient

When comparing top income shares and the Gini coefficient, it is useful to apply the following approximate decomposition proposed by Atkinson (2007, p. 19-20), and proved by Alvaredo (2011),

$$G \approx (1 - S) \cdot G^{bottom99\%} + S, \qquad (8)$$

where S is the income share of the top 1 per cent and $G^{bottom\,99\,\%}$ is the Gini coefficient of the bottom 99 per cent.[16] The approximate decomposition of the Gini coefficient is shown in Figure 5.

Figure 5 demonstrates that the evolution of the share of the top 1 per cent does not capture the evolution of income inequality in Norway, although overall inequality and top income shares have moved closely together in recent decades. Over the period 1882 to 1939 the Gini coefficient is seen to vary significantly, even though the Gini coefficient measured in percentage points only declined modestly, from 64 per cent in 1882 to 59 per cent in 1939. By contrast, the share of the

[15] The averaging is done because we recognize that what many researchers require is a single series, and that if we do not provide an average ourselves, users will do so. At the same time, there is no evident justification for taking a simple average. It can be argued that the upper bound attributes an unreasonably low income to those recording zero. The appropriate weights may vary over the time period. But the simple average provides a point of reference.

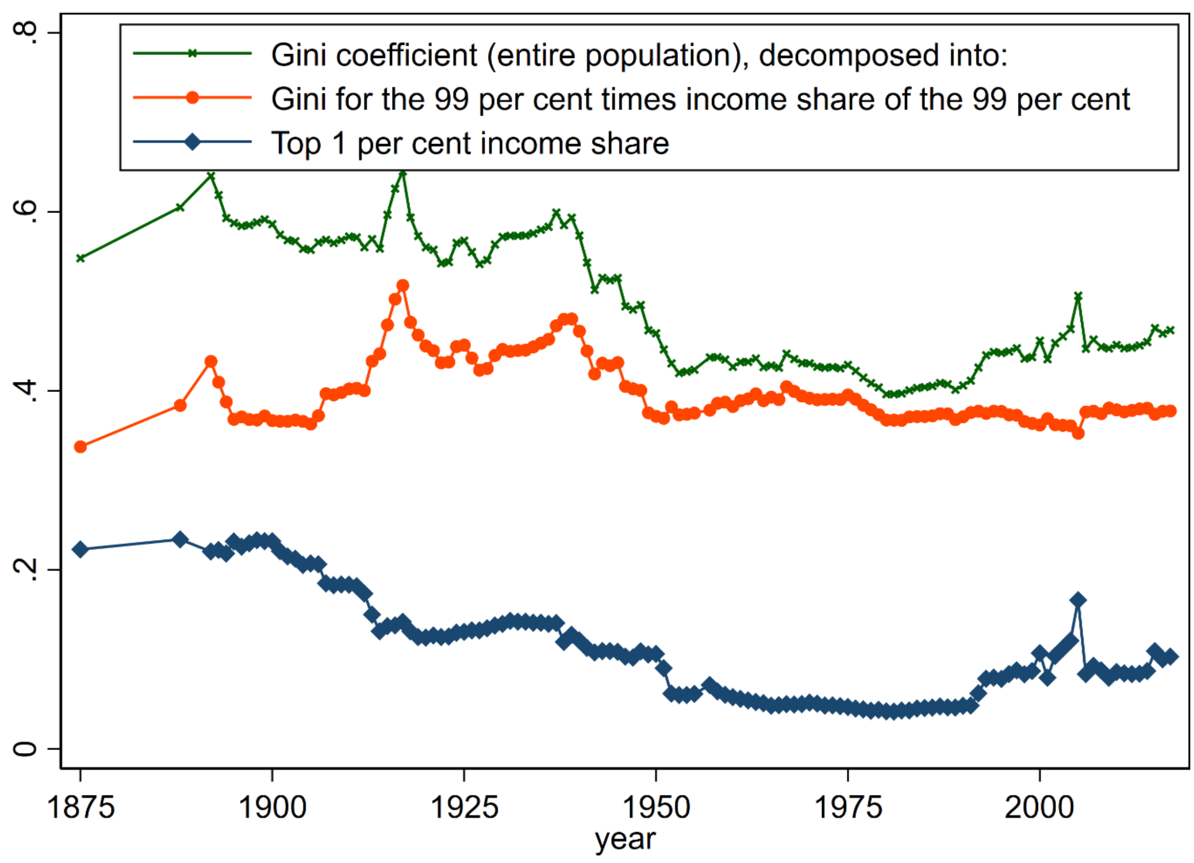
[16] The exact decomposition is given by

$$G = 0.01 \cdot S \cdot G^{top1\%} + (1 - S) \cdot G^{bottom99\%} - 0.01 \cdot (1 - S) \cdot G^{bottom99\%} + S - 0.01,$$

where the first, third and fifth terms are relatively small compared to the second and fourth terms, which justifies the simplified expression in (8). The first term in (8) is always 0.001 or less in our data, while the third term is maximum 0.005. The last term is constant at 0.01 by definition. Note that the top 1 per cent series shown here (and given in the Appendix) differs slightly from those published by Aaberge and Atkinson (2010), simply because the present series utilizes additional data sources, relies on different assumptions on the distribution of unmeasured income and uses interpolation of the top distribution across some years, as explained in Section 3.

top 1 per cent decreased significantly, from 22 per cent to 13 per cent, over the same period, while the Gini of the 99 per cent increased from 43 to 48 per cent.

Figure 5. Decomposition of the Gini coefficient by the income share of the top 1 per cent and the Gini coefficient of the distribution of income among the bottom 99 per cent, Norway 1875 - 2017



Note: The components are estimated using the same estimated Lorenz curves as were used for estimating the Gini coefficients in Figure 3. The graphs display the mean of the upper and lower bound estimates. For sources, methods and assumptions, see text.

During the Second World War and the early post-war period, both overall inequality and top income shares showed a substantial decline. Between 1939 and 1953 the Gini coefficient fell from 59 per cent to 42 per cent, while the share of the top 1 per cent fell from 13 per cent to 6 per cent. The evolution over the next three decades was again rather different. There was a substantial decline in the share of the top 1 per cent, from 6 per cent in 1953 to 4 per cent in 1980, whereas the Gini coefficient was fairly stable. Since 1990, the share of the top 1 per cent has regained the lost ground, and was 10 per cent in 2017 according to our estimates here, and the Gini coefficient too has risen – although only to around 47 per cent. This difference between the time paths of the top shares and the Gini coefficient shows that, while the top share may have driven much of the recent increase in overall inequality, there have been other forces in operation as a result of which not all of the post-war equalization has been lost. Note that the evolution of the share of the top 10 per cent parallels the evolution of the share of the top 1 per cent (see Aaberge and Atkinson, 2010). We refer to Online Appendix G for a decomposition of the Gini coefficient by the income share of the top 10 per cent and the Gini coefficient of the distribution of income among the bottom 90 per cent. Furthermore, Appendix G provides results of the evolution of decile-specific income shares.

Since the evolution of top income shares and overall inequality differs in a number of periods, estimates on upper tail inequality and the ratio between the mean incomes of the lower and upper half of the population might provide essential information on whether changes in overall inequality are due to a widening of the income gap between the upper and lower half of the population and/or changes in the distribution of income among the richest 50 per cent of the population. As will be demonstrated below, these distributional measures contribute to explain the driving forces behind the evolution of overall inequality.

4.2. Affluence

Before World War II, taxpayers comprised between 52 and 81 per cent of the annual populations of tax units, which means that the data base for describing the upper half of the income distribution is richer than that for describing the lower half. This makes it particularly relevant to consider the evolution of the mean and the Gini coefficient for the most affluent 50 per cent of the population and use the associated estimates as a basis for estimating "affluence", a measure introduced by Aaberge and Atkinson (2016). Affluence has been given an axiomatic justification and is defined by

$$A = \frac{1}{3}\left(\frac{\mu_U}{\mu}(G_U + 1) - 1\right) = \frac{2}{3(1+\gamma)}\left(G_U + \frac{1}{2}(1-\gamma)\right), \qquad (9a)$$

where μ is the overall mean income, μ_U and G_U are the mean and the Gini coefficient, respectively, of the richest 50 per cent of the population, $\gamma = \mu_L/\mu_U$ and μ_L is the mean of the poorest 50 per cent of the population. Expression (9a) shows that affluence, *A,* increases with increasing inequality in the income distribution of the richest 50 per cent and decreases with increasing mean income ratio γ [17]. Inserting the well-known expressions for μ_U and G_U into (9a) yields the following alternative expression for *A*,

$$A = \frac{4}{3}\int_{\frac{1}{2}}^{1}(2t-1)\left(\frac{F^{-1}(t)}{\mu} - 1\right)dt \ , \qquad (9b)$$

where $F^{-1}(t)$ is the income of the individual with rank *t* in the income distribution F . Expression (9b) shows that *A* can be interpreted as a (normalized) weighted average of the income shares of the richest 50 per cent, where the weight increases with increasing rank from 0 for the median income to 4/3 for the highest income. The affluence measure *A* has itself a range $[0,1]$ and takes the value 0 if and only if all individuals receive the same income μ . At the other extreme, when total income is received by one sole individual, then *A* takes the value 1. Note that 3*A* becomes equal to the relative affluence gap $\left((\mu_U - \mu)/\mu = (1-\gamma)/(1+\gamma)\right)$ if individuals with higher income than the median income receive the same income μ_U . The estimation results for affluence, *A*, and the upper tail (above median) Gini coefficient G_U are displayed in Figure 6.

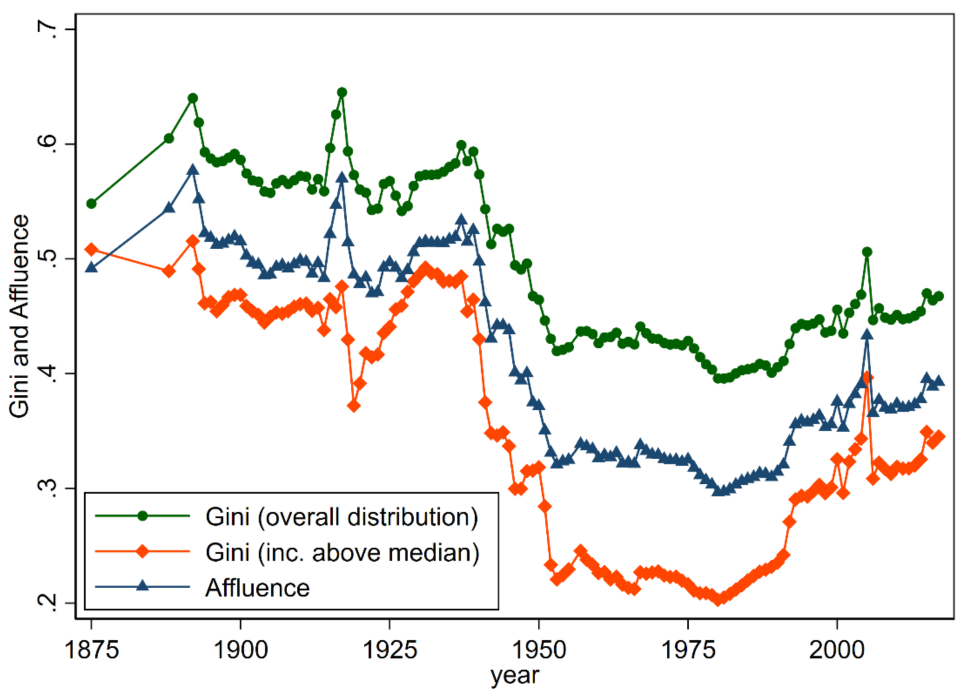
Since the available data provide a better basis for estimating affluence than overall inequality before World War II, it is reassuring that the affluence pattern largely captures the pattern of the

---

[17] As demonstrated by the following expression $\gamma = 2(\mu/\mu_U) - 1$ there is one to one correspondence between γ and μ/μ_U .

overall Gini. Note also that the reliability of affluence (and the upper tail Gini coefficient) to a large extent carries over to the estimated overall Gini series. This is because income distributions are normally skewed to the right, which means that the upper tail Gini contributes a significantly larger share of the overall Gini than the lower tail Gini. Aaberge and Atkinson (2016) demonstrated that the overall Gini is equal to $3(A+P)/4$, where *P* is the poverty counterpart of affluence (*A*). Thus, in 1900, with G = 0.586 and *A* = 0.515 (see Table A8), the contribution from affluence to the overall *G* was 66 per cent, while the contribution of *A* had declined only marginally, to 62 per cent, a hundred years later.

Figure 6. Gini coefficients for the overall distribution of income and the distribution of income for those with income above the median, and a measure of affluence, Norway 1875 - 2017



Note: The components are estimated using the same distribution as the one used for estimating the Gini coefficients in Figure 3. The graphs display the mean of upper and lower bound estimates. For sources, methods and assumptions, see text.

To complement the information on inequality given by the Gini coefficient, Appendix A provides estimates of two closely related rank-dependent measures of inequality. The results show to be in line with the evidence provided by this section, i.e. changes in inequality from 1939 to 1980 and from 1980 to 2017 largely concerned the upper part of the income distribution.

5. Long-run inequality in Norway: A series of episodes

The evolution of inequality in Norway is best characterized, we believe, as a series of episodes identified with sub-periods, which are summarized in Table 1. As demonstrated by the change in percentage points, the evolution of the overall Gini coefficient is closely related to the evolution of Gini-based affluence measure. In the same way, the upper tail Gini and the mean income ratio typically move in the opposite direction; rising (declining) upper tail Gini and declining (rising) mean income ratio. It is evident from Figure 6 that three episodic changes in income inequality deserve special mention. First, inequality was turbulent during World War I, but analysis is complicated by price and wage fluctuations during this period. Leaving aside this period, the Gini coefficient in the four decades from the 1890s to the end of the 1930s was measured in the range of 0.60 plus or minus 0.05. Second, the decline during World War II was swift and extensive. Third, the post-1989 reversal took the Gini from around 0.40 to over 0.45 in two decades. We turn now to consider the individual sub-periods in more detail.

Taken as a whole, the period from 1875 to 1939 shows unchanged overall inequality and affluence, whereas the upper tail Gini decreased by 5 percentage points. The different evolution of upper tail inequality and overall inequality (and affluence) corresponds to a significant decrease in the ratio between the mean incomes of the lower and upper halves of the population (see Figure A8 and Table A8). The increase in the Gini coefficient from 0.55 in 1875 to 0,64 in 1892 reflects an increase of the share of total income accruing to the highest income group. The growth rates were low during this period, and emigration to North America increased sharply from 1880. This was followed by high economic growth in the 1890s, which ended in the so-called "Kristiania crash" in 1899 leading to substantial drops in property values and stagnation for several years. In particular, there appears to have been a downward tendency in overall inequality from the late-1890s to around 1905, followed by remarkable stability from 1905 to 1914. The most dramatic short-run event occured during the First World War, where we observe the highest Gini coefficient of 0.65 and the smallest mean income ratio between the lower and upper half of the population occurred in 1917, when the mean income of the lower half was only 8.9 per cent of that of the upper half. The low mean income ratio for this period reflects the significant income growth for ship owners and the high speculative profits for wealthy people during a significant economic boom, which was followed by a recession with high inflation, trade deficits and currency depreciation and hardships such as rationing that affected wage earners. As a result, the income of the rich declined and the mean income ratio doubled from 1917 to 1923. However, inequality quickly returned to its pre-war level in the early 1920s and increased slightly during the 1930s.

The simultaneous substantial growth of the mean income ratio and decline of the upper tail Gini coefficient led to a substantial fall in the overall Gini coefficient from 1939 to 1953. Since the mean income ratio stayed fairly flat at around 1/4 since the early 1950s, the rise in overall inequality and affluence after the turning point in 1980 was largely due to rising upper tail inequality. This means that the richest became richer, as is also confirmed by the rising top income shares during this period. The concentration in time of the sharp decrease in the Gini coefficient between 1939 and 1953 is likely a combination of several factors. First, the manner of operation of labor market institutions changed significantly during the 1930s, where collective bargaining was introduced at the national level. Economic turbulence may have postponed the immediate effects of these reforms. Second, more than 40 per cent of the work force was still in agriculture in the 1930s, and rural-urban migration (and hence income equalization) was again constrained by high unemployment. Moreover, the Second World War was likely to have had an equalizing effect in itself, with more controls imposed on the economy where the German occupation led to increased labor demand for extensive construction projects and larger mean income for the bottom half of the population which resulted in increased mean income ratio. Moreover, the German command economy reduced the income opportunities of most capital owners, which might explain why the upper tail Gini coefficient sharply declined. The war experience might also have made Norwegians more receptive to the strict economic planning regime that was introduced during the early post-war period (Espeli, 2013).

Other sources support the finding of a significant fall in income inequality during this period. For example, the 1950 Wage statistics (NOS XI 092, p. 11, table A) compares wages for various occupations in 1939 and 1950. While high-paid groups such as senior public servants had experienced nominal wage growth of 69 per cent, the wage growth for sailors was 214 per cent, for forestry workers 264 per cent and for farm workers (servants) 380 per cent. For the lower-income groups wage data are also available for 1944; they show that wage compression was well underway during the war.

We observe a stable income Gini coefficient from 1950 onwards, with a further slight decrease in the early 1970s. The fall in income inequality was reversed in the early 1980s. The turning point was largely due to increased wage inequality and came shortly after oil began to flow from the North Sea (Aaberge and Mogstad, 2011). By 1990 production had been at a high level for a number of years. The 1990s show a recovery of the shares of top incomes, probably as a result of expanded opportunities to earn and lose money created by the oil sector, a major financial market reform in the mid-1980s, and the 1992 tax reform whereby taxes on capital incomes were significantly reduced. On top of that, a tax reform where taxes on capital incomes were significantly reduced was implemented in 1993. Over the period from 1980 to 2017, the Gini coefficient increased by approximately 20 per cent. The spike in income inequality in 2005 is largely due to the increased taxes on dividends in 2006. This tax reform gave owner-managers of closely held firms strong insentives to increase dividends in 2005. The effects of the reform discussed in further detail by Aaberge, Atkinson and Modalsli (2016) and Alstadsæter et al. (2016) suggest that the level of inequality might have been larger after dividend taxation was implemented in 2006 than what has been captured by the standard income statistics data.

Table 1. Changes in overall inequality, upper tail inequality, ratio of the mean income of the lower and upper 50 per cent and affluence (changes in percentage points in parentheses)

Period	Overall Gini coefficient	Gini-based affluence	Upper tail Gini coefficient	Mean income ratio $\gamma = \mu_L / \mu_U$
1875 - 1892	Increase (+9)	Increase (+9)	Slight increase (+1)	Decrease (-11)
1892 - 1914	Decrease (-8)	Decrease (-9)	Decrease (-8)	Increase (+6)
1914 - 1917	Increase (+9)	Increase (+9)	Increase (+4)	Decrease (-8)
1917 - 1923	Decrease (-10)	Decrease (-10)	Decrease (-6)	Increase (+9)
1923 - 1939	Increase (+5)	Increase (+5)	Increase (+5)	Slight decrease (-4)
1939 - 1953	Decrease (-17)	Decrease (-20)	Decrease (-24)	Increase (+11)
1953 - 1980	Slight decrease (-2)	Slight decrease (-2)	Slight decrease (-2)	Slight increase (+3)
1980 - 2017	Increase (+7)	Increase (+10)	Increase (+14)	Slight decrease (-4)

Note: The components are estimated by using the estimated overall means jointly with the same estimated Lorenz curves as were used for estimating the Gini coefficients in Figure 3. Changes are calculated on the basis of the average of upper and lower bound estimates. For sources, methods and assumptions, we refer to the text.

To get some sense of the magnitude of the changes in the Gini coefficient, note that the 22 percentage points fall in the Gini coefficient from 1892 to 1953 (see Table A8) corresponds to a 34 per cent decrease in the Gini coefficient. This corresponds to the redistributive effect of the following hypothetical tax/transfer intervention in 1892 (see Aaberge, 1997): introduce a flat tax with tax rate 34 per cent and allocate the collected tax as a fixed lump-sum transfer equal to the average tax of NOK 178. Then the 50 per cent poorest increase their income on average from NOK 104 to NOK 247, while the 50 per cent richest will get their mean income reduced from NOK 944 to NOK 801. Moreover, this hypothetical intervention would change the income of the poor from NOK 85 to NOK 234 and the 95 per cent quantile from NOK 1630 to NOK 1254.

6. Summary

While data on top income shares provide valuable information on the concentration of economic power, this paper demonstrates that available historic data souces make it feasible to examine the evolution of the income distribution as a whole over long time periods. By combining detailed tabulations with aggregate information on the incomes of municipal and central government taxpayers, as well as administrative data on poverty support, we are able to provide an estimate of the income distribution in 1875 and annually from 1892 to 1951. This is then supplemented with detailed tax tabulations and micro data from 1952 onwards in order to provide income distributions through to the most recently available data for 2017. From these income distributions we can then estimate Gini coefficients, as well as other relevant measures of income inequality and affluence, for consistent definitions of population and income throughout the entire period in question. The proposed method is likely also to be of relevance for other countries.

The empirical results provide three novel insights into the long-run evolution of income inequality in Norway. First, our findings suggest that at the end of the nineteenth century, the Gini coefficient for gross family income in Norway varied between 0.50 and 0.60. Such an apparently Latin American value casts some doubt on the claim made in the official publication for the Paris Exhibition of 1900 that "among civilised states, there is scarcely any that is so fortunate with regard to the equality of its social conditions as Norway. There is no nobility with political or economic privilege, no large estates, no capitalist class" (*Norway*, 1900, page 203). While Norway has exhibited low inequality from the 1940s till 1990s, we find no indication that this represents a continuation of an earlier egalitarian society.

Second, the movement of income inequality over time appears to be driven by episodic changes rather than predictable, secular cycles. Overall gross income inequality among families in Norway fell from 1892 to 1914, largely due to a fall in inequality in the upper half of the income distribution. There was an upward spike during World War I, and a moderate rise between 1923 and 1939, again largely due to changes in inequality in the upper half. Inequality fell substantially between 1939 and 1953 as a result of a decline in both upper tail inequality and the gap between upper and lower tail means. Income inequality was low and stable between 1953 and 1980 and has risen again since 1980.

Expressed in this way, the history of Norwegian income inequality is better seen as a series of episodes than as the expression of some long-run pattern. It can neither be summarized by an inverse U nor by a U. Moreover, the series of 143 years of income inequality estimates does neither point in the direction of any regular cycles of increasing and decreasing inequality.

Third, it should be noted that the turning point and the origin of the low post-war inequality in Norway was the significant decline in inequality starting in 1940 and continuing during the German occupation. The war experience might also have made Norwegians more receptive to the strict economic planning regime that was introduced in the early post-war period (Espeli, 2013).

A comparison of the levels of income inequality in Norway with previous estimates for other countries is challenging for several reasons: No countries have complete micro data far back in time; there is no universally agreed definition of population or income (as these in turn are dependent on the available data); and there are often breaks even within series for comparable countries. Despite the break in the series for Denmark provided by Atkinson and Søgaard (based on tabulations from income taxes) these series makes an exception and shows, as for Norway, increased income inequality during World War I and a substantial decrease in income inequality during the mid-20th century. The turning point with increasing income inequality arose in the early 1980s for both countries, but inequality has risen more for Norway than for Denmark.

References

Aaberge, R., 1997, "Interpretation of Changes in Rank-Dependent Measures of Inequality", *Economics Letters*, 55, 215-219.

Aaberge, R., 2007, "Gini's Nuclear Family." *Journal of Economic Inequality*, 5, 305-322.

Aaberge, R and A B, Atkinson, 2010, "Top incomes in Norway" in A B Atkinson and T Piketty, editors, *Top incomes in a global perspective*, Oxford University Press, Oxford.

Aaberge, R and A B, Atkinson, 2016, "Median as watershed", Discussion Paper No. 749, Statistics Norway.

Aaberge, R, Atkinson, A B and J Modalsli 2016, "The ins and outs of top income mobility." Mimeo.

Aaberge, R. and M. Mogstad, 2011, "Robust Inequality Comparisons", *Journal of Economic Inequality*, **9**, 353 – 371.

Alstadsæter A, Jacob M, Kopczuk W and Telle K, 2016, "Accounting for Business Income in Measuring Top Income Shares: Integrated Accrual Approach Using Individual and Firm Data from Norway", Discussion Paper No. 837, Statistics Norway.

Alvaredo, F., 2011. "A note on the relationship between top income shares and the Gini coefficient". *Economics Letters* **110**, 274-277

Atkinson, A.B., 2007. "Measuring Top Incomes: Methodological Issues." In: Atkinson, A., Piketty, T. (Eds.), Top Incomes over the Twentieth Century: A Contrast Between Continental European and English-Speaking Countries. Oxford University Press, Oxford.

Atkinson, A and Piketty. T, 2007. Top Incomes over the Twentieth Century: A Contrast Between Continental European and English-Speaking Countries. Oxford University Press, Oxford.

Atkinson, A and Piketty T., 2010. Top incomes in a global perspective, Oxford University Press, Oxford.

Atkinson, A B and Søgaard, J E, 2016, "The long run history of income inequality in Denmark", *Scandinavian Journal of Economics,* **118**, 264–291.

Atkinson, A B, Hasell, J, Morelli, S and Roser, M., 2017. The Chartbook of Economic Inequality. Institute for New Economic Thinking, Oxford.

Blanchet, T, Fournier, J and and Piketty, T, 2017. "Generalized Pareto curves: Theory and applications." Wid.world working paper 2017/3

Espeli, H, 2013. "Economic consequences of the German occupation of Norway, 1940-1945." *Scandinavian Journal of History*, **38** (4), 502-524.

Garbinti B, Goupille-Lebret J and and Piketty T, 2018. "Income inequality in France, 1900-2014: Evidence from Distributional National Accounts (DINA)". *Journal of Public Economics* **162**, 63-77.

Gerdrup, K R, 1998, "Skattesystem og skattestatistikk i et historisk perspektiv", Statistics Norway Report 98/6.

Kiær, A N, 1892-3, "Indtægtsforhold i Norge", *Tillæg Statsøkonomisk tidsskrift*. Aschehoug, Kristiania

Kopczuk W, Saez E and Song J, 2010: "Earnings inequality and mobility in the United States: Evidence from social security data since 1937". *Quarterly Journal of Economics*, **125**, 91-128.

Kuhn M, Schularick M and Steins U, 2017: "Income and wealth inequality in America, 1949-2013." CEPR Discussion Paper 12218

Kuznets, S, 1955, "Economic growth and income inequality", *American Economic Review,* **45,** 1-28.

Leigh, A, 2007. "How closely do top income shares track other measures of inequality?" *Economic Journal* 117, 619-633.

Lindert, P H, 2000, "Three centuries of inequality in Britain and America." In: A B Atkinson and F Bourguignon (Eds.), , *Handbook of Income Distribution*, *Volume* 1, Elsevier, Amsterdam.

Lindert, P H and Williamson, J G, 2016, Unequal gains: American growth and inequality since 1700, Princeton University Press, Princeton.

Merwin, C L, 1939, "American studies of the distribution of wealth and income by size", *Studies in Income and Wealth*, **3**, National Bureau of Economic Research, New York.

Milanovic, B, 2016. Global inequality: A new approach for the age of globalization. The Belknap Press of Harvard University Press

Morelli, S., T. Smeeding and J. Thompson (2015): "Post-1970 trends in within-country inequality and poverty: rich and middle-income countries". In Atkinson and Bourguignon (Eds.), *Handbook of Income Distribution*, *Volume 2A*, Elsevier, Amsterdam.

Norway. Official publication for the Paris exhibition, 1900. Kirke – og undervisningsdepartementet. http://www.archive.org/details/norwayofficialp00konogoog

Piketty, T, 2001, Les hauts revenus en France au 20ème siècle, Grasset, Paris.

Piketty, T, Saez, E and Zucman, G, 2018, "Distributional National Accounts: Methods and Estimates for the United States." *Quarterly Journal of Economics*, vol. 133 (2), p. 553-609.

Roine, J. and Waldenstrom, D., 2015. "Long-run trends in the distribution of income and wealth". In Atkinson and Bourguignon (Eds.), *Handbook of Income Distribution*, *Volume 2A*, Elsevier, Amsterdam.

Soltow, L, 1965, Toward Income Equality in Norway, Madison - Milwaukee: The University of Wisconsin Press.

Vecchi, G, 2017. Measuring Wellbeing: A history of Italian living standards. Oxford University Press.

Williamson, J G, 1985, Did British capitalism breed inequality?, Allen and Unwin, London.

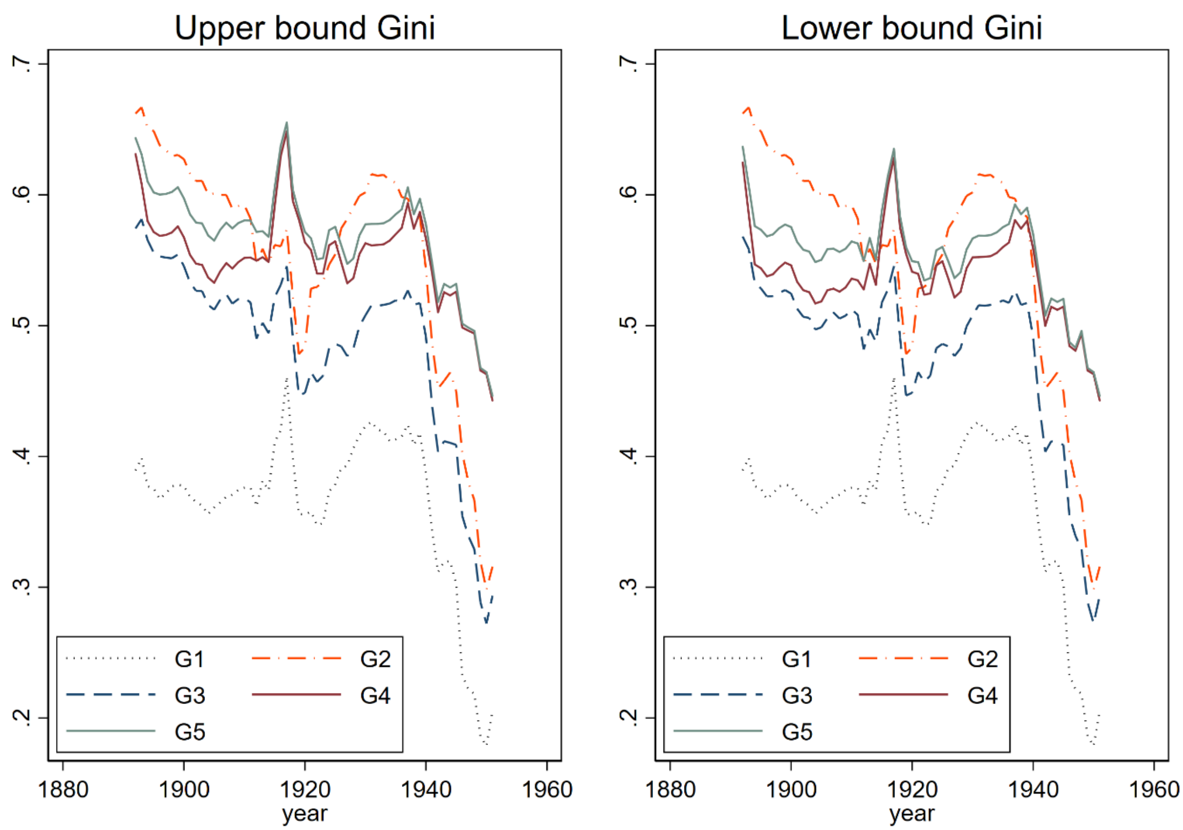
Williamson, J G and Lindert, P H, 1980, American inequality: A macroeconomic history, Academic Press, New York.

Appendix A: Robustness checks

Sensitivity of estimates to the use of additional data sources

Figure A1 shows how sensitive the estimated Gini coefficients in the 1892-1951 period are to the inclusion of additional sources of data and methodological assumptions. The lowest curve in each panel, G1, displays estimates of the Gini coefficent when the population is defined by taxpayers (MUN and CG) and only one interior point of the Lorenz curve has been identified. This gives a Gini coefficient of around 0.4 before 1939 and less than 0.2 in 1950. By including non-taxpayers and assigning them zero income, we get G2, which shows that the Gini coefficient rose to 0.65 in 1892 and 0.3 in 1950. The G3 curve is obtained by assigning an income to the non-taxpaying population according to the procedure described in Section 3. As a transition from the distribution underlying G2 to the distribution underlying G3 could be obtained (with the appropriate scaling of mean incomes) by means of regressive transfers, G3 is always lower than G2. By taking account of within-group inequality in the richest group, i.e. the CG taxpayers, we get the G4 curve. For years where detailed CG tax tabulations are not available, the closest available earlier or later distribution has been used (the higher one for the upper bound and the lower for the lower bound). Finally, in G5, we also allow for inequality within the group of those who pay municipal tax but not central government tax. This increases inequality moderately, and more so in years when this group is large. It is evident from this exercise that the steps we propose in Section 3 are crucial for a correct estimation of income inequality, though the assumptions on within-group inequality in some groups have only a minor effect on the estimates.

Figure A1. Estimates of the bounds of the Gini coefficients using five different sets of assumptions, 1892-1951.



Note: Definitions: G1:Inequality among taxpayers, no within-group dispersion. G2: Assuming zero income for non-taxpayers. G3: Baseline assumption about non-taxpayer income. G4: Within-group dispersion for CG taxpayers (using nearest year). G5: Within-group dispersion for MUN-CG taxpayers (our preferred estimate, as presented in Figure 3).

Sensitivity to choice of inequality measure

To complement the information on inequality provided by the Gini coefficient, we employ two closely related rank-dependent measures of inequality (C_1 and C_3) discussed by Aaberge (2007) and defined by

(A1) $$C_k = 1 - \frac{1}{\mu}\int_0^1 p_k(t)F^{-1}(t)dt,\ k = 1, 2, 3,$$

where

(A2) $$p_k(t) = \begin{cases} -\log t, & k = 1 \\ \dfrac{k}{k-1}\left(1 - t^{k-1}\right), & k = 2, 3. \end{cases}$$

and μ and $F^{-1}(u)$ denote the mean and the left inverse of F. As demonstrated by Aaberge (2007) the measures C_1, C_2 and C_3, denoted Gini's Nuclear Family, jointly provide a good summary of the information provided by the Lorenz curve. Whilst it can be shown that the Gini coefficient (C_2) tends to pay most attention to changes that occur in the middle part of a typical single peaked income distribution, the two other members of Gini's Nuclear Family are shown to be particularly sensitive to changes that occur in the lower part (C_1) and the upper part (C_3) of the income distribution.

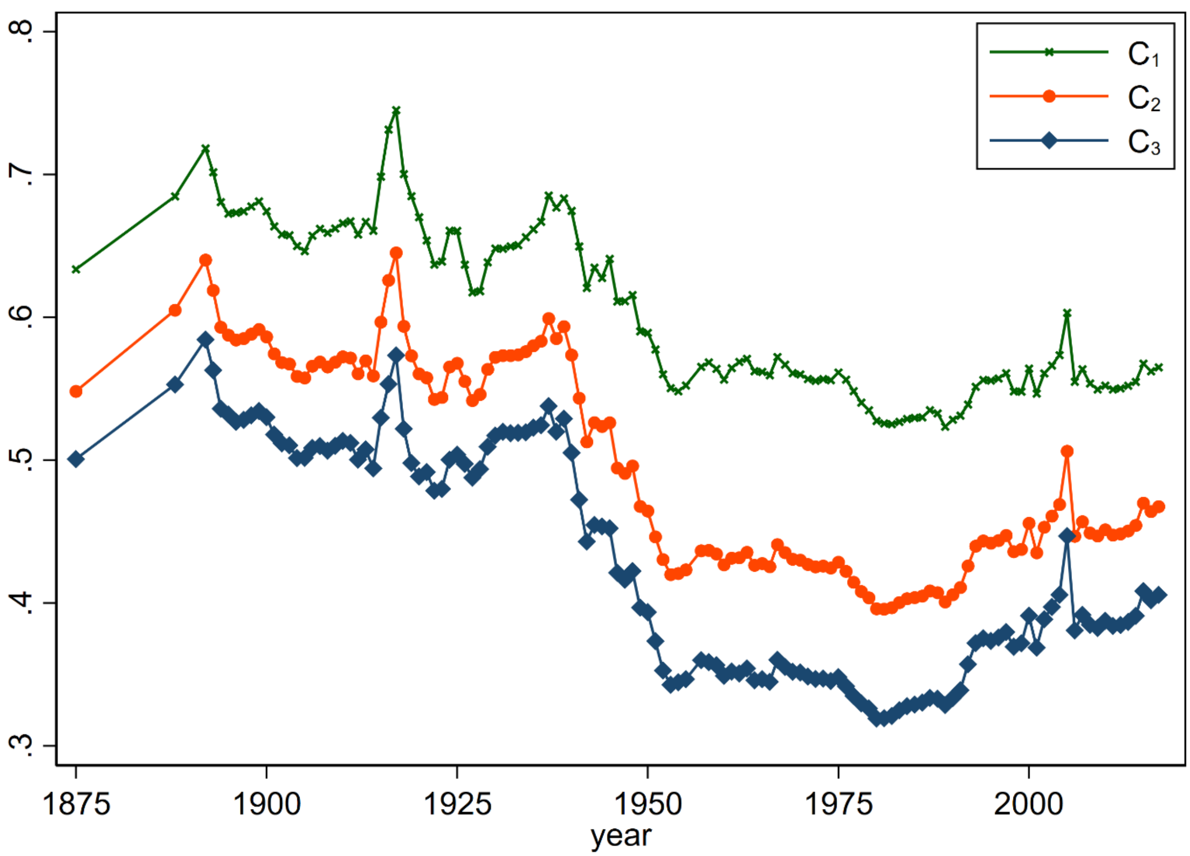
Note that the ratio of the second term of (A1) can be interpreted as the ratio between the social welfare attained under the observed distribution F and that attained under complete equality. As a contribution to the interpretation of the inequality aversion profiles exhibited by C_1, C_2 and C_3, Table A1 displays ratios of the weights – as defined by (A2) – given to the median individual and the 5 per cent poorest, the 30 per cent poorest and the 5 per cent richest individuals, respectively.

Table A1: Distributional weight profiles of the inequality measures C_1, C_2 and C_3

Relative weights	C_1 (Bonferroni)	C_2 (Gini)	C_3
p(.05)/p(.5)	4.32	1.90	1.33
p(.30)/p(.5)	1.74	1.40	1.21
p(.95)/p(.5)	0.07	0.10	0.13

As suggested by the table above, C_1 is more sensitive than C_2 to changes in the income distribution that concern the poor, whereas C_2 is more sensitive than C_3 to changes that occur in the lower part of the income distribution. For example, the weights in Table A1 demonstrate that the weight of an additional euro to a person located at the 5 per cent decile is 4.3 times the weight for the median income earner when C_1 is used as a measure of inequality, whereas it is only 1.3 times the weight for the median earner when C_3 is used as a measure of inequality. This means that C_1 is particularly sensitive to changes that take place in the lower part of the income distribution, and C_3 to changes in the upper part of the income distribution.

Figure A2. Long-run evolution of income inequality described by three alternative measures of income inequality, Norway 1875 - 2017



Note: The components are estimated using the same distribution as the one used for estimating the Gini coefficients in Figure 3. The graphs display the mean of the upper and lower bound estimates. For sources, methods and assumptions, see text.

The results displayed in Figure A2 show that the evolution of the Gini coefficient is largely reflected by the inequality measures that are particularly sensitive to changes in the lower and upper part of the income distribution, respectively. However, the magnitudes of the changes differ significantly for some periods. From 1939 to 1980, the C_1-coefficient decreased by 20 per cent, the Gini coefficient by 33 per cent and the C_3-coefficient by 40 per cent, whereas the percentage changes were almost equal from 1875 to 1939. From 1980 to 2017 C_1, C_2 *and* C_3 rose by 7, 20 and 27 per cent, respectively. These results, which are in line with the evidence provided by Section 4.3, show that changes in inequality from 1939 to 1980 and from 1980 to 2017 largely concern the upper part of the income distribution.

Online Appendix: data, estimates and supplementary results

Appendix B: Sources of Norwegian tax data

Sources of tabulated income tax data

The two income tax sources form the basis for the tabulation of taxpayers by income ranges from 1948 to 1966, which precede the micro-data available from 1967. As noted in the text, the number paying MUN tax exceeds that paying CG tax. In the tabulations (HS 1978, Table 314), income is equal to assessed income according to the central government tax assessment for the years 1948-51 and assessed income according to municipal tax assessment for the years 1952-55. This accounts for the jump in the number of taxpayers and amount of assessed income in 1952, from 947,842 CG taxpayers in 1951 to 1,412,873 MUN taxpayers in 1952 (an increase of 49 per cent), and from NOK 7,993 million to NOK 10,227 million in 1952 (an increase of 28 per cent). The smaller percentage increase in total income reflects the fact that those paying MUN but not CG the MUN-CG group), have lower average incomes. In the tabulations for the years 1957 to 1966, income is defined as assessed income according to the central government tax assessment if central government tax is levied. If not, income is defined as assessed income according to the municipal tax assessment.

The sources for pre-1948 years are listed in Table A2.

Table A2: Sources of tabulated income data

Year	Source	Taxpayer categories	Number of taxpayers	Number of groups
1875	Skattelikningen 1876 (A.N. Kiær - 1892-93, pp. 110-113, included tax free incomes and Ot.Prp. no. 11 for 1881 pp. 20-25	MUN	705 460	33
1888	St. Porp. Nr. 48. (1890), ppages 42 and 122	MUN	472 104	9
1892	Ot. Prop. No. 39	CG	176 142	8
1893	St. Prop. No 91	CG	102 542	6
1894	St. Prop. No. 112	CG	66 807	5
1895	St. Prop. No. 104	CG	68 233	14
1896	St. Prop. No. 89	CG	70 454	14
1897	Statsskattens fordeling 1892/93-1898/99	CG	75 578	14
1898	Statsskattens fordeling 1899/00-1905/06	CG	94 587	15
1899	Statsskattens fordeling 1899/00-1905/06	CG	91 422	14
1900	Statsskattens fordeling 1899/00-1905/06	CG	94 367	14
1901	Statsskattens fordeling 1899/00-1905/06	CG	95 767	14
1902	Statsskattens fordeling 1899/00-1905/06	CG	97 517	14
1903	Statsskattens fordeling 1899/00-1905/06	CG	96 431	14
1906	Rygg, 1910, pages 50 and 69		677 487	17
1913	NOS VI.57, page 30*		774 308	12
1938	Stat Medd 1941, nos 11 and 12, page 333		410 020	26
1948-1951	HS1978, Table 314, page 572-573	CG	Lowest: 954 524 Highest: 1 047 017	25
1952-1955	HS1978, Table 314, page 572-573	MUN	Lowest: 1 396 738 Highest: 1 439 770	25
1957-1966	HS1978, Table 314, page 572-573	MUN and CG	Lowest: 1 372 298 Highest: 1 543 022	25
1967 -	Administrative microdata			

Sources of aggregate statistics on taxpayers and the poor

The source of the aggregate numbers of taxpayers and total assessed income (before the adjustment from all taxpayers to personal taxpayers) is displayed in Table A3.

Table A3. Sources of data on municipal and central government taxpayers

Years	Municipal tax aggregates	Central government tax aggregates
1875 and 1888	See Appendix B (Detailed sources)	Not applicable
1892-1899	Historical Statistics 1948, Table 220	Statistical Yearbook 1902, Table 99
1900	Historical Statistics 1948, Table 220	Statistical Yearbook 1906, Table 104
1901	Historical Statistics 1948, Table 220	Statistical Yearbook 1907, Table 104
1902	Historical Statistics 1948, Table 220	Statistical Yearbook 1908, Table 108
1903-1908	Historical Statistics 1948, Table 220	Statistical Yearbook 1909, Table 108
1909-1914	Historical Statistics 1948, Table 220	Statistical Yearbook 1915, Table 112A
1915-1916	Historical Statistics 1948, Table 220	Statistical Yearbook 1918, Table 124
1917-1919	Historical Statistics 1948, Table 220	Statistical Yearbook 1920, Table 143a
1920	Historical Statistics 1948, Table 220	Statistical Yearbook 1921, Table 160
1921	Historical Statistics 1948, Table 220	Statistical Yearbook 1924, Table 179
1922-1923	Historical Statistics 1948, Table 220	Statistical Yearbook 1926/1927, Table 178
1924-1926	Historical Statistics 1948, Table 220	Statistical Yearbook 1929, Table 199
1927-1936	Historical Statistics 1948, Table 220	Statistical Yearbook 1940, Table 267
1937-1945	Historical Statistics 1948, Table 220	NOS Tax Statistics *
1946-1953	Historical Statistics 1958, Table 200	NOS Tax Statistics *

* NOS tax statistics are annual publications; numbers for year t are reported in the publication with the title "t+1/t+2" i.e. "NOS tax statistics for the fiscal year 1938/39" have data for 1937 and so on.

One problem in using these statistics is to restrict the coverage to personal taxpayers by excluding non-personal taxpayers, a group that "comprises joint-stock companies, co-operative societies and other corporations" (HS 1968, page 428). This applies to the tax data between 1921 and 1947 (from 1948 onwards we have separate reports on personal taxpayers and total taxpayers). For most years between 1937 and 1947, we have separate reports of the totals and interpolate the missing years using the ratio between personal and all taxpayers. There is little year-to-year variation in this ratio. For this reason, we use the 1937 ratio to impute the share of personal taxpayers (and their income) for the period 1921-1936. For municipal taxpayers, this amounts to multiplying the total number of taxpayers by 0.937 and total income by 0.855. For central government taxpayers, the corresponding numbers are 0.973 and 0.848.

Data on the number of supported poor and total poverty support 1875-1951 are obtained from the annually published poverty statistics. An overview of data for every fifth year is found in Historical Statistics 1994, Table 7.8.

Appendix C: Controls for total tax units and total income

Control totals: adults and tax units

The adult population is defined as those aged 16 and over. The data from 1948 onwards were supplied by Statistics Norway. Pre-1948 data on the population by age are available from *Historical Statistics 1994*, Table 3.5, for 5 year intervals. We took the data for 31 December of year (t-1) as applying to year t, so that the data cover years ending in 1 or 6. From these, we calculated the proportion of the population aged 16 and over, and interpolated linearly for the intervening years. The percentages were then applied to the mean annual population figures given in *Historical Statistics 1978*, Table 9.

Total tax units are obtained by subtracting the number of married women. The numbers of married women are given at 5 year intervals in Historical Statistics 1994, Table 3.7. They are expressed as percentages of the adult population and the percentages linearly interpolated. The results are shown in Table A5.

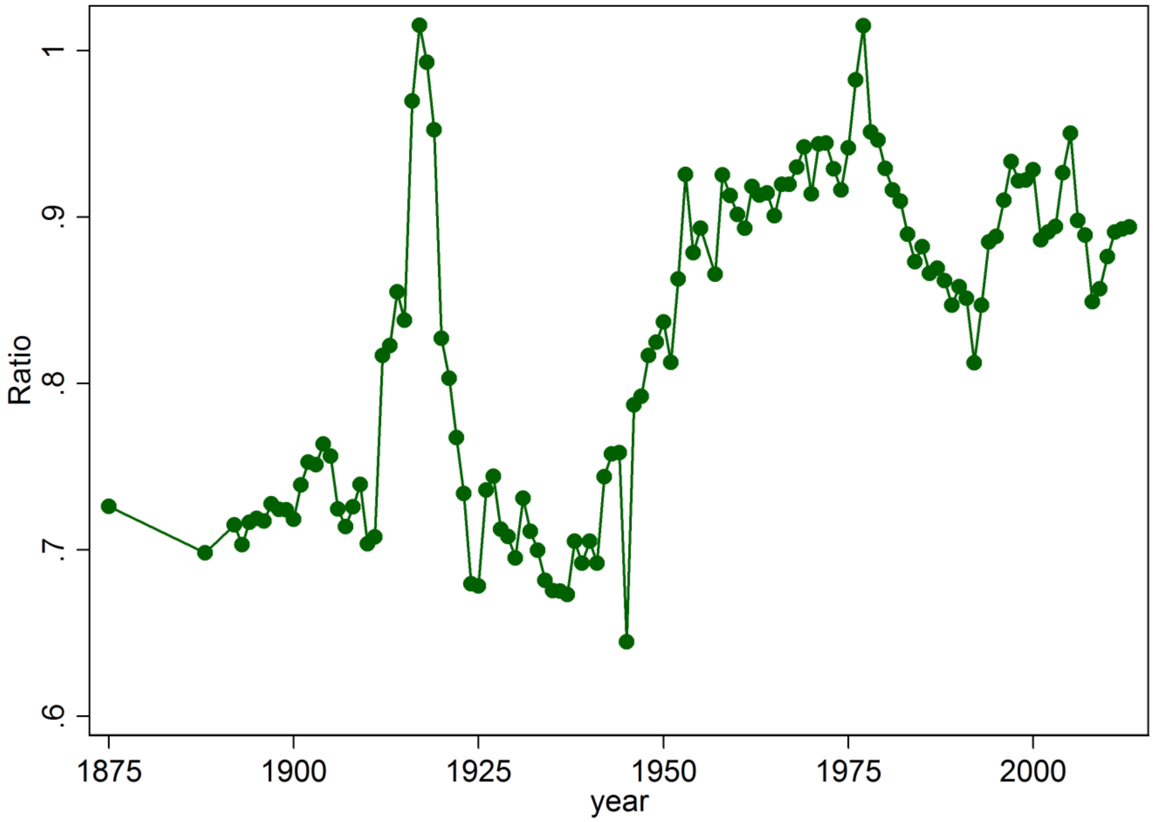
Control totals: household income

The starting point for total income is a series for total household income as measured in the national accounts for 1978 to 2006 provided by Statistics Norway. Total household income is made up of (i) compensation of employees (not including employers' social security contributions), (ii) operating surplus of self-employed businesses, (iii) property income, (iv) transfers from government and from abroad, and (v) income not elsewhere classified. In order to extrapolate this series backwards, we have made use of series that are as comparable as possible, given the available material from HS 1994 and earlier editions. In each case, the series have been linked at years where the estimates seem most comparable (for this reason we have started with 1979, rather than 1978). If the 1979 value from the Statistics Norway series is A1979, and the first linked series is for 1975 to 1979, given by B1975, …, B1979, then for 1978 we take the value of B1978 multiplied by A1979/B1979.

Working backwards to 1950, we have used the *Nasjonalregnskap 1968-1979*, Table 33, pages 138-139 for the New Definition of Private Income for 1968 to 1978. For 1950 to 1968, we have used the Old Definition of Private Income from *Historisk statistikk 1978* (Statistics Norway, 1978), Table 59 (page 104) for 1965 to 1968 and from *Historisk statistikk 1968*, Table 70 (pages 110-111) for 1950 to 1964. In each case employers' social security contributions were subtracted from the total of private income; these were taken from *Nasjonalregnskap 1969-1980*, Table 30 (for 1969 to 1974), *Nasjonalregnskap 1962-1978*, Table 29 (for 1962 to 1968), *Nasjonalregnskap 1953-1969*, Table 14 (for 1953 to 1961), and *Nasjonalregnskap 1968-1979*, Table 14 (for 1950 to 1952).

For years prior to 1950, we use for 1930 to 1950 *Nasjonalregnskap 1865-1960* (NOS XII 163), Table 24, adding Direct taxes paid to Private disposable income. This source does not give figures for 1940 to 1945, and we have interpolated for 1940 to 1943 using the net real income figure in Table 35 of *Statistiske oversikter 1948* (NOS X 178). No figures are given for 1944 and 1945. For years prior to 1930, the main source is *Langtidslinjer i Norsk Økonomi 1865-1960,* Table VIII, where we have taken the sum of Private income from labour and capital and Transfers from government and Transfers from abroad. This source provides annual estimates from 1865 to 1900. For the period 1900 to 1930, the estimates are given at 5 yearly intervals. The figures for intermediate years have been interpolated using the series for "private gross income" from *Nasjonalregnskap 1900-1929* (NOS XI 143), Table 7.

Figure A3. Internal total as a share of control total

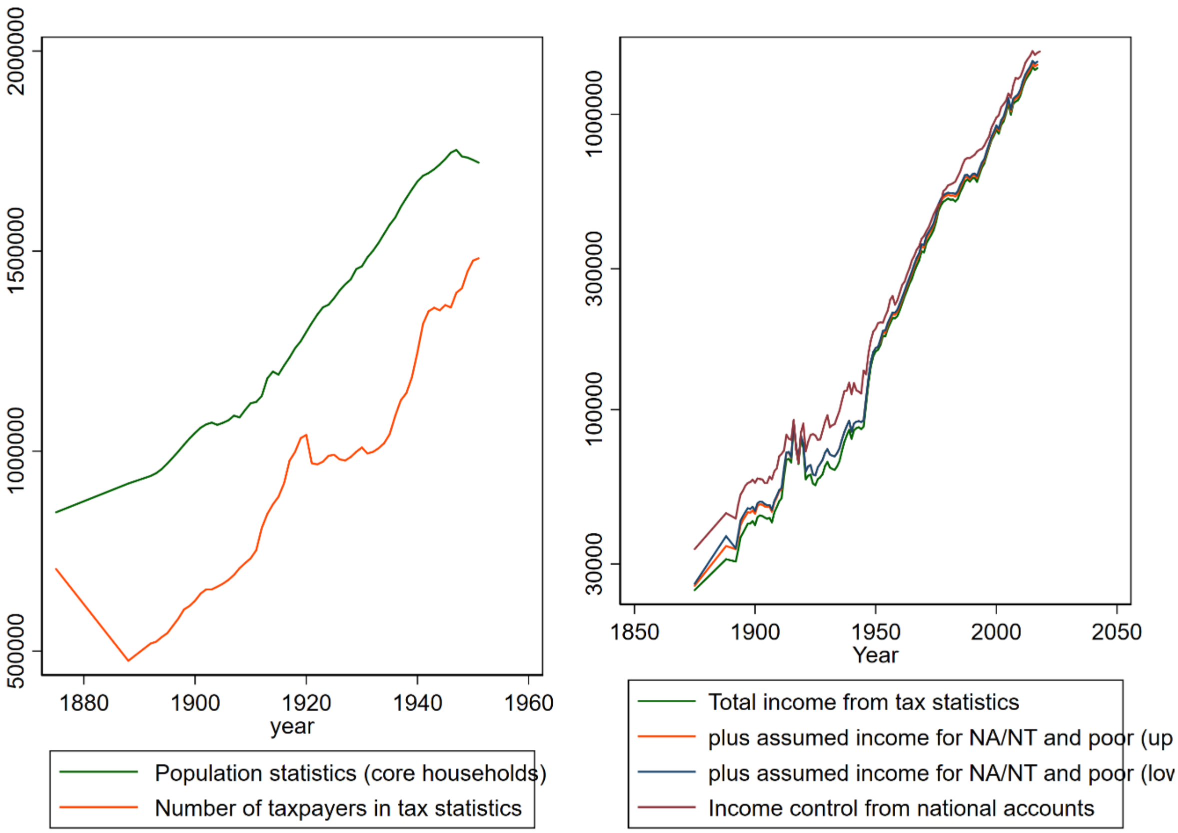


Note: The figure shows the internal total (total income calculated from tax statistics) relative to the control total calculated from (reconstructed) national accounts. Source: National accounts (control total) and calculations from tax and poverty statistics (internal total)

Figure A4 shows control totals for income and population, as discussed in Section 2.2.

If we subtract the income totals for taxpayers, non-assisted/non-taxed and poor from the control total, we get an "excess income" that in our method is allocated to taxpayers, as discussed in Section 3. We can compare this excess to the control total to get an idea of its magnitude. This ratio is between 0.10 and 0.27 in all years 1875-1951, with the exception of the years between 1915 and 1920 when it is lower and 1945 when it is higher (0.32).

Figure A4. Control totals for income and population



Note: The left panel shows the total population (age 16 and over, married couples counted as one unit) from population statistics and the total number of units in the tax statistics. The right panel shows the total income from tax statistics, as well as the total income obtained using the assumptions explained in Section 3, as well as the total income from (reconstructed) national accounts.

Appendix D: Within-group distributions

The Gini coefficients discussed in the main paper are not dependent on assumptions on within-group inequality per se. Rather, they can be construed (in the years where there are no data on within-group dispersion) as interpolations based on within-group Gini coefficients. However, in some cases it is desirable to draw Lorenz curves for illustrative purposes or to estimate other inequality measures beside the Gini coefficient. In these cases, the following within-group distributions are one example of function forms that are consistent with the within-group Gini coefficients. Moreover, the calculations here verify that the within-group Gini coefficients are consistent across groups: i.e., that the lowest-income individuals in the higher groups do not have lower incomes than the highest-income individuals in poorer groups.

CG group (highest incomes)

For the three-group case, consider a Pareto distribution for the CG group with the probability density function

$$f(y) = \frac{\alpha d^{\alpha}}{y^{\alpha-1}}$$

with mean income $\frac{\alpha d}{\alpha-1}$ and lower bound d. We set the parameter d to make the mean correspond to the mean income of the CG group, $\mu^{CG} = (1-g)/c$. This gives

$$d = \frac{\alpha - 1}{\alpha}\mu^{CG}$$

The within-group Gini coefficient of the GC group is

$$G^{*CG} = \frac{1}{2\alpha - 1}$$

Aaberge and Atkinson (2010) provide values of the Pareto coefficient α for the relevant periods (1892-1903 and 1948-1951), which correspond to within-group Gini coefficients between 0.33 and 0.5.

MUN-CG groups

For the individuals who pay municipal tax but not state tax, we use a uniform distribution with probability density function

$$f(y) = \frac{1}{b-a}, y \in [a, b]$$

with mean income $\mu = (a+b)/2$, lower bound a and upper bound b.

The Lorenz curve for a uniformly distributed population is

$$L(F) = \frac{1}{a+b}\left((b-a)F^2 + 2aF\right)$$

and the corresponding Gini coefficient is

$$G^* = 1 - 2\int_0^1 L(F)dF = \frac{1}{3}\frac{b-a}{a+b}$$

For our purposes, it is convenient to rephrase the uniform distribution using the mean m and a spread parameter z giving the relative distance of the lower and upper bounds from the mean: $a = (1-z)\mu$ and $b = (1+z)\mu$. This gives a Gini coefficient of $G^{**} = z/3$.

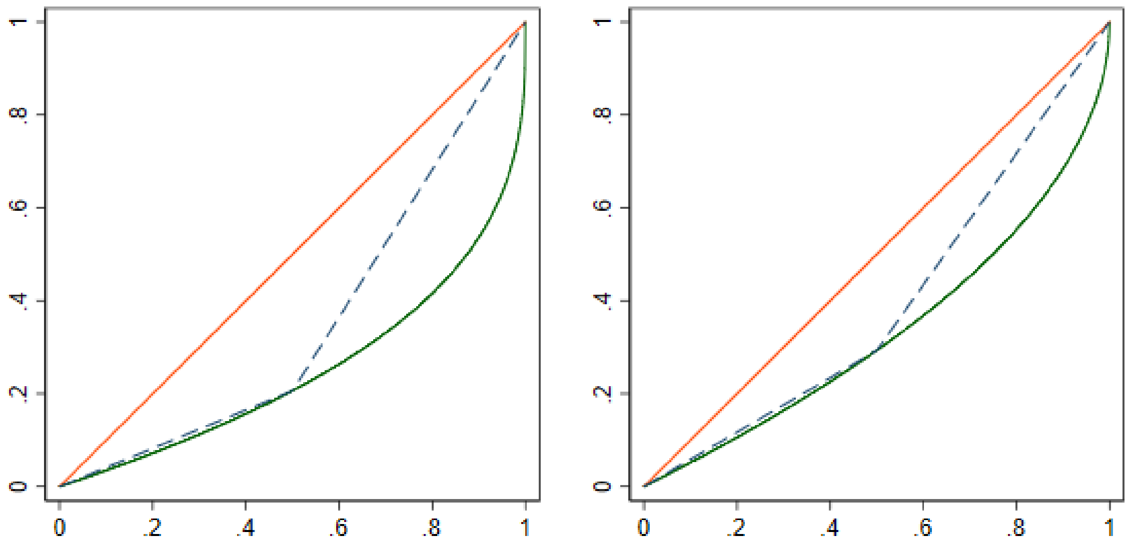
To respect the assumption that the highest-income individual in the MUN-CG group should not have higher income than the lowest-income individual in the CG group, b^{MUN-CG} must be lower than or equal to d. Using the known means and inserting for the above equations, we get

$$(1 + z^{MUN-CG}) \leq \frac{\alpha - 1}{\alpha}\frac{\mu^{CG}}{\mu^{MUN-CG}}$$

An illustration of the Lorenz curve using Pareto distributions

Figure A5 illustrates the general point that Lorenz curves of left-skewed income distributions lean heavily to the right. The two Lorenz curves are plotted for Pareto distributions with shape parameters $\alpha = 1.5$ (left panel and $\alpha = 2$ (right panel, which correspond to Gini equal to 0.500 and 0.333). It is evident from Figure A5 that there is clearly larger potential for within-group inequality for the half of the population above the median than for the half below the median.

Figure A5. Illustration of Lorenz curves for Pareto distributions with shape parameter $\alpha = 1.5$ (left) and $\alpha = 2$ (right)



Appendix E: Adjustments to tabulated data 1952 to 1966

From 1952 to 1966, the income distributions used in this paper are obtained from detailed tables in Historical Statistics 1978 (HS1978). (There are also tables for 1948-1951 on the same pages, but these are CG taxpayers only and hence cover a lower share of the population. They are used to calculate G* for these years in the 4-class tables). Some adjustments to these data are required to make the time series consistent with the period up until 1951 and the micro data from 1967 onwards.

Adjustment for sailor taxation

A separate sailor tax, based on a law from 1947, was introduced in 1948. Sailors are not included in the HS1978 detailed tables. We add sailors to these tables. From 1956 onward we have the number of sailors and their mean income from HS1978 Table 308. Before 1956, we use the tax statistics, or HS1978 Table 307, which shows total sailor taxes paid, and deduce the numbers from that. We use a uniform distribution on (0, 2*sailor mean inc) for sailor incomes and add these to the tables for 1951-1966.

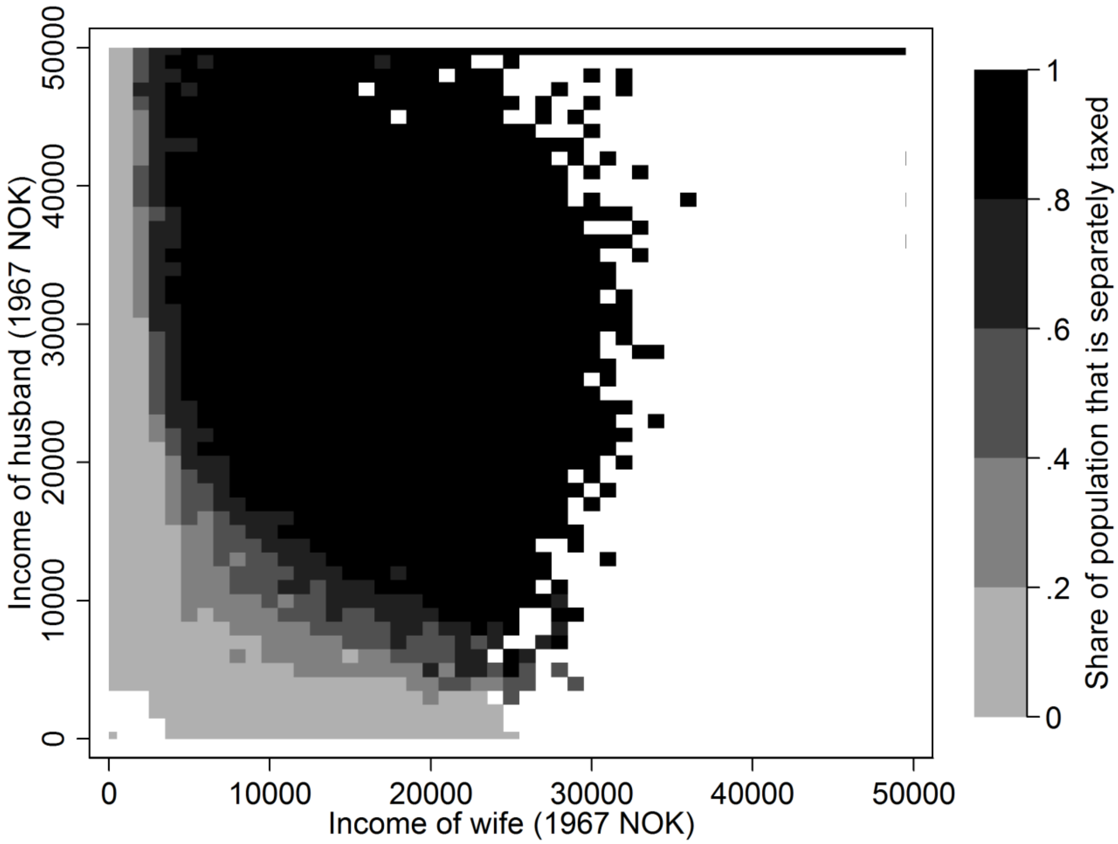
To apply the changes, the tabular data are re-grouped into 100 percentiles with mean incomes and population sizes. The same is done for the sailors; the tables are then added. For the spouses, a transformation algorithm based on registry data from 1967 is applied whereby spouses are observed individually, as well as a variable informing us whether they chose to be taxed separately or not. This is described in detail in the next paragraph.

Treatment of married couples

Up until 1960, married women are always taxed with their husbands. From 1960, married couples could elect to be taxed separately. They are then included as two separate individuals in the tabulations. In the registry data (available from 1967) we can identify, on the individual level, which individuals were taxed separately. Hence, we can construct tabulation of units both by taxation status (as in HS1978) and by couples jointly (our preferred population, and the one used in tabulations before 1960). In this section we describe how we use information from the registry data to construct a conversion algorithm that we apply to the 1960-1966 tabulations, and in this way increase the comparability of the data.

In the 1967 income file, we observe 115,753 couples that are definitelytaxed separately. These are mainly couples where both have high incomes, as shown in Figure A6, which gives the share of couples that are taxed separately by wife's and husband's incomes. The darker shade denotes that more than 80% with this income combination are taxed separately, while the lighter shade denotes less than 20%. White means that there are few individuals with this income combination.

Figure A6. Percentage of couples taxed separately in Norway, by husband and wife's income, 1967



Note: The figure shows the incidence of joint taxation, by husbands' and wives' incomes in 1967 (when joint taxation was optional). Source: Individual-level income data from Statistics Norway. For details, see text.

In 1960, we have no registry data on incomes, but the Census of 1960 has information on the "main source of livelihood" for individuals and is available in registry form. The variable "main source of livelihood" has three possible values:

1. Income from own work
2. Pensions / transfers / income from wealth / loan / scholarships etc
3. Income from someone else's work (supported)

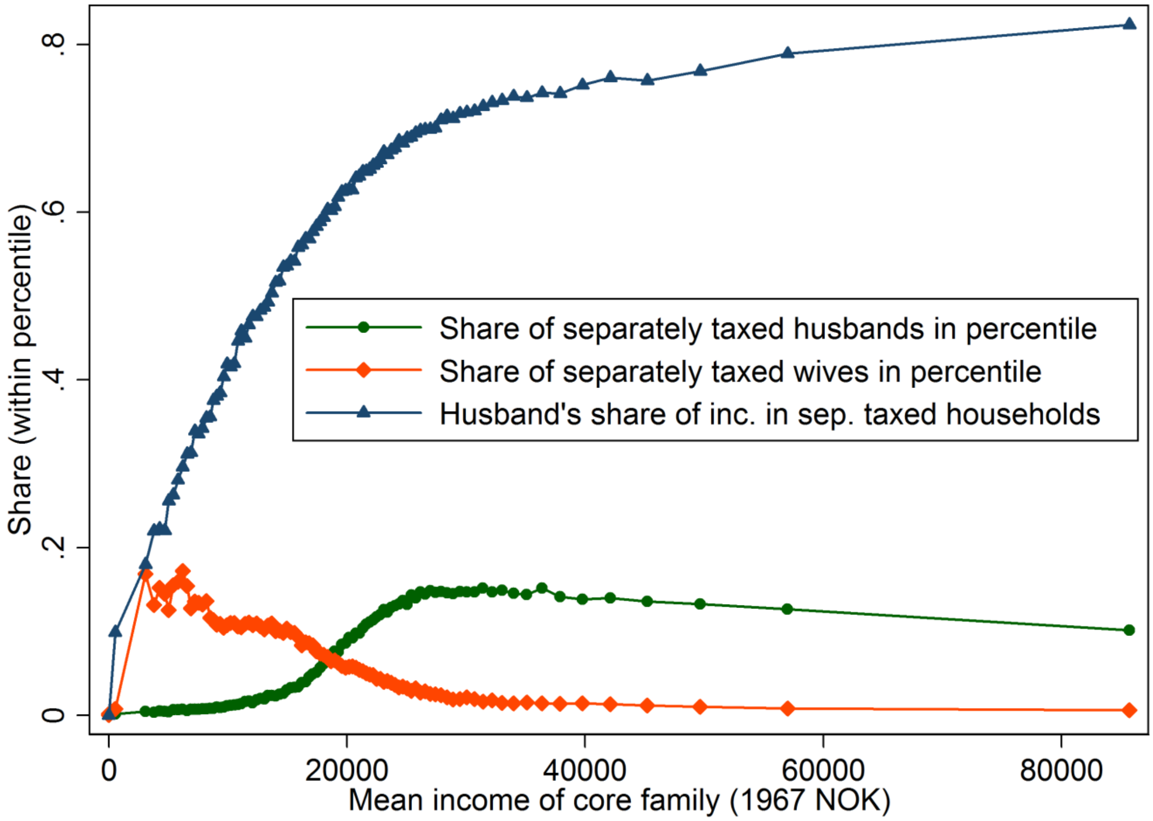
Most married women are in category 3, while most married men are in category 1 (all combinations exist). There are around 45 000 married couples where both husband and wife are in category 1. Hence, we assume that 45 000 couples were taxed separately in 1960. This corresponds well with the increase in the number of units in the tax statistics of 67 509 (from 1 372 298 to 1 439 807) from 1959 to 1960, allowing for some growth in the general population in addition to the results of the tax law change. In the absence of more data, we use a linear interpolation for the numbers of those who were separately taxed between 1960 and 1967.

The approach adopted to transform the data from 1960 to 1966 is as follows:

- Construct a file of the 1967 population that corresponds to the tabular definitions from 1960 onwards. That is, merge married couples into one unit with income = (husband's income + wife's income) only if they are jointly taxed. If they are separately taxed, keep them as two units. Separate taxation usually takes place if both spouses have non-neglible incomes. In the file, each unit can be either
  - An unmarried (or widowed, etc) man
  - An unmarried (or widowed, etc) woman
  - A married couple with joint taxation (where at least one spouse is marked as not filing separately)
  - A married man with separate filing (whose wife also files separately)
  - A married woman with separate filing (whose husband also files separately)
- Divide this population into 100 percentiles, sorted by income.
- For each percentile, calculate
  - The share of units that is a married man with separate filing
  - The share of units that is a married woman with separate filing
  - Among units that are married men with separate filings, the income share of the husband in the marriage (ie (income of husband / (income of husband+income of wife))

These shares (from now on interpreted as probabilities contingent on income percentile) are shown in Figure A7.

Figure A7. Prevalence of separate taxation in 1967, by income percentile



Note: The figure shows the incidence of joint taxation, by household income, in 1967 (when joint taxation was optional). Source: Individual-level income data from Statistics Norway. For details, see text.

We now apply the transformation to each of the years 1960 to 1966 (year *t*) as follows:

- Adjust the "is husband" and "is wife" probabilities down by the factor (number of separately taxed couples in *t*)/(number of separately taxed couples in 1967)
- Divide the tabular population of year *t* into 100 percentiles, sorted by income. (Many of the percentiles will have equal incomes, as the tables have less than 100 categories. This is not a problem.)
- Divide each percentile group with a given mean income y and population N into three groups:

- Separately taxed husbands: N* (probabilitiy of being separately taxed husband). We return to this group below
- Separately taxed wives: N* (probability of being separately taxed wife). We delete these observations as we want to consider them together with their husbands
- The remaining population N* (the sum of the two above probabilities) consists of either single individuals or jointly taxed couples and is left as is.

- Divide the income of the separately taxed husbands by the mean share of separately taxed husbands in the percentile. As we divide by a number between 0 and 1, these incomes are inflated. This step converts the separately taxed husbands' incomes into couples' incomes.
- Finally, re-group the observations into 100 percentiles again. We will now have a smaller population as we have created "pseudo-couples" that closely resemble couples in the underlying population.

For 1967, the procedure gives near-perfect results. For earlier years, we cannot test the procedure directly. However, the sums of the imputed incomes are very close to the sums of original incomes (largest difference is 0.6 %), which is a sign that the interpolation is relativelyaccurate.

Adjustments for years before 1892

Some special adjustments have been made for the pre- 1892 period, as no state tax was collected in this period. We do have the total number of poor and their total support, from which we obtain the mean income of the poor. For the NA/NT group, we use the income of the poor (NOK 85 in 1875 and NOK 76 in 1888) for the upper bound Gini and NOK 150 as the lower bound Gini. We take the lowest income group in the tabulations, which contains 74 per cent of tabulated individuals in 1875 and 86 per cent of tabulated individuals in 1888, and treat these similarly to the MUN-CG groups in 1892 and thereafter. The remaining 26 and 14 per cent are treated similarly to the CG groups in later years, and the G* figures for these years are estimated on this population, with incomes above NOK 400 (in 1875) and NOK 1000 (in 1888).

Appendix F: Estmates of the Gini coefficient and other key variables for Norway 1875-2017

Five tables are shown below:

- Overall Gini coefficient and other measures of income dispersion (Table A4)
- Income and population totals for the four groups before 1951 (Table A5) and population and income totals after 1951 (Table A6)
- Parameters used for calculating the "four-group" Gini coefficient (Table A7)
- Alternative measures of income dispersion (Table A8)

The information is also available as an online Appendix in Excel format (on request). Moreover, plots of Lorenz curves for all years are available as a PDF file.

Table A4. Upper and lower bounds of overall Gini coefficient, upper tail Gini coefficient, ratio of upper tail mean income and overall mean income and affluence, Norway 1875-2017

	Upper bound assumptions				Lower bound assumptions			
Year	Gini coefficient	$\frac{\mu_U}{\mu}$	Gini (above median)	Affluence	Gini coefficient	$\frac{\mu_U}{\mu}$	Gini (above median)	Affluence
1875	0.556	1.651	0.508	0.496	0.540	1.651	0.508	0.496
1888	0.641	1.828	0.489	0.574	0.569	1.763	0.489	0.542
1892	0.643	1.807	0.515	0.580	0.637	1.796	0.515	0.574
1893	0.631	1.801	0.491	0.562	0.607	1.761	0.491	0.542
1894	0.610	1.785	0.461	0.536	0.576	1.729	0.461	0.509
1895	0.602	1.770	0.462	0.530	0.573	1.724	0.462	0.507
1896	0.600	1.770	0.454	0.525	0.568	1.718	0.454	0.500
1897	0.601	1.765	0.460	0.525	0.570	1.716	0.460	0.501
1898	0.602	1.760	0.467	0.527	0.574	1.716	0.467	0.506
1899	0.606	1.764	0.469	0.530	0.577	1.720	0.469	0.508
1900	0.597	1.751	0.469	0.524	0.575	1.716	0.469	0.507
1901	0.585	1.736	0.459	0.511	0.564	1.703	0.459	0.495
1902	0.579	1.728	0.454	0.504	0.558	1.696	0.454	0.489
1903	0.578	1.729	0.451	0.503	0.557	1.696	0.451	0.487
1904	0.569	1.717	0.445	0.494	0.549	1.686	0.444	0.478
1905	0.565	1.707	0.450	0.492	0.550	1.686	0.449	0.481
1906	0.573	1.717	0.453	0.498	0.559	1.694	0.453	0.487
1907	0.579	1.721	0.458	0.504	0.559	1.699	0.446	0.486
1908	0.574	1.710	0.461	0.500	0.556	1.693	0.447	0.483
1909	0.578	1.714	0.466	0.504	0.559	1.695	0.451	0.486
1910	0.580	1.716	0.469	0.507	0.564	1.703	0.452	0.491
1911	0.580	1.713	0.470	0.506	0.563	1.698	0.452	0.489
1912	0.572	1.698	0.471	0.499	0.549	1.686	0.439	0.475

1913	0.572	1.712	0.457	0.498	0.567	1.704	0.457	0.495
1914	0.568	1.708	0.450	0.493	0.550	1.699	0.426	0.474
1915	0.606	1.754	0.480	0.532	0.588	1.748	0.449	0.511
1916	0.637	1.814	0.479	0.561	0.614	1.809	0.437	0.533
1917	0.655	1.836	0.498	0.583	0.635	1.836	0.454	0.557
1918	0.604	1.779	0.452	0.528	0.584	1.779	0.407	0.501
1919	0.586	1.792	0.401	0.504	0.560	1.792	0.343	0.469
1920	0.572	1.749	0.417	0.493	0.549	1.749	0.366	0.463
1921	0.567	1.729	0.439	0.496	0.549	1.729	0.397	0.472
1922	0.551	1.704	0.433	0.481	0.535	1.704	0.395	0.459
1923	0.552	1.703	0.435	0.481	0.536	1.703	0.399	0.461
1924	0.573	1.726	0.453	0.503	0.558	1.726	0.417	0.482
1925	0.576	1.729	0.459	0.507	0.560	1.729	0.423	0.487
1926	0.561	1.701	0.471	0.501	0.549	1.701	0.441	0.484
1927	0.547	1.679	0.472	0.490	0.536	1.679	0.446	0.476
1928	0.551	1.679	0.484	0.497	0.541	1.679	0.459	0.483
1929	0.569	1.700	0.493	0.513	0.558	1.700	0.468	0.499
1930	0.577	1.709	0.499	0.521	0.567	1.709	0.474	0.506
1931	0.578	1.706	0.503	0.521	0.569	1.706	0.482	0.509
1932	0.578	1.709	0.498	0.520	0.569	1.709	0.477	0.508
1933	0.578	1.710	0.497	0.520	0.569	1.710	0.476	0.508
1934	0.581	1.717	0.491	0.520	0.571	1.717	0.469	0.507
1935	0.585	1.722	0.492	0.523	0.575	1.722	0.469	0.510
1936	0.589	1.728	0.493	0.526	0.578	1.728	0.467	0.512
1937	0.606	1.751	0.500	0.542	0.593	1.751	0.470	0.525
1938	0.585	1.750	0.454	0.515	0.585	1.750	0.454	0.515
1939	0.597	1.758	0.472	0.530	0.590	1.758	0.457	0.520
1940	0.577	1.743	0.438	0.502	0.570	1.743	0.422	0.493
1941	0.548	1.735	0.386	0.469	0.538	1.735	0.364	0.455
1942	0.518	1.700	0.360	0.437	0.508	1.700	0.336	0.424

1943	0.532	1.728	0.359	0.450	0.521	1.728	0.333	0.435
1944	0.529	1.725	0.362	0.449	0.518	1.725	0.336	0.435
1945	0.532	1.730	0.350	0.445	0.520	1.730	0.324	0.430
1946	0.502	1.697	0.315	0.410	0.487	1.694	0.284	0.392
1947	0.498	1.682	0.315	0.404	0.483	1.676	0.284	0.384
1948	0.496	1.674	0.315	0.400	0.496	1.674	0.315	0.400
1949	0.468	1.615	0.315	0.375	0.468	1.615	0.315	0.375
1950	0.464	1.604	0.318	0.372	0.464	1.604	0.318	0.372
1951	0.446	1.597	0.284	0.350	0.446	1.597	0.284	0.350
1952	0.440	1.628	0.233	0.336	0.421	1.604	0.233	0.326
1953	0.429	1.619	0.221	0.326	0.411	1.596	0.221	0.316
1954	0.431	1.624	0.224	0.329	0.410	1.595	0.224	0.318
1955	0.433	1.619	0.230	0.330	0.413	1.592	0.230	0.319
1957	0.446	1.632	0.245	0.344	0.427	1.606	0.245	0.333
1958	0.446	1.636	0.238	0.342	0.427	1.611	0.238	0.331
1959	0.445	1.638	0.233	0.340	0.423	1.609	0.233	0.328
1960	0.438	1.628	0.226	0.332	0.416	1.599	0.226	0.320
1961	0.441	1.633	0.227	0.334	0.421	1.607	0.227	0.324
1962	0.440	1.635	0.221	0.332	0.423	1.612	0.221	0.322
1963	0.445	1.643	0.223	0.336	0.425	1.616	0.223	0.325
1964	0.436	1.628	0.216	0.327	0.417	1.603	0.216	0.317
1965	0.439	1.638	0.213	0.329	0.416	1.606	0.213	0.316
1966	0.437	1.636	0.212	0.328	0.414	1.604	0.212	0.315
1967	0.453	1.657	0.227	0.344	0.429	1.624	0.227	0.331
1968	0.447	1.646	0.226	0.339	0.424	1.615	0.226	0.326
1969	0.442	1.636	0.227	0.336	0.419	1.606	0.227	0.323
1970	0.441	1.635	0.228	0.336	0.418	1.603	0.228	0.323
1971	0.439	1.631	0.224	0.332	0.415	1.599	0.224	0.319
1972	0.437	1.629	0.223	0.331	0.414	1.598	0.223	0.318
1973	0.437	1.629	0.223	0.331	0.414	1.599	0.223	0.319

1974	0.436	1.630	0.220	0.329	0.413	1.599	0.220	0.317
1975	0.440	1.639	0.217	0.331	0.417	1.608	0.217	0.319
1976	0.433	1.628	0.211	0.324	0.412	1.599	0.211	0.312
1977	0.425	1.615	0.209	0.317	0.404	1.586	0.209	0.306
1978	0.419	1.604	0.209	0.313	0.397	1.575	0.209	0.301
1979	0.414	1.596	0.207	0.309	0.393	1.568	0.207	0.298
1980	0.406	1.584	0.203	0.302	0.386	1.558	0.203	0.291
1981	0.406	1.584	0.205	0.303	0.385	1.556	0.205	0.291
1982	0.408	1.585	0.208	0.305	0.386	1.556	0.208	0.293
1983	0.413	1.592	0.212	0.310	0.388	1.559	0.212	0.296
1984	0.415	1.595	0.216	0.313	0.390	1.562	0.216	0.300
1985	0.416	1.594	0.220	0.315	0.392	1.562	0.220	0.302
1986	0.417	1.593	0.224	0.316	0.393	1.561	0.224	0.303
1987	0.419	1.595	0.227	0.319	0.398	1.566	0.227	0.307
1988	0.418	1.591	0.229	0.318	0.397	1.563	0.229	0.307
1989	0.410	1.578	0.232	0.314	0.392	1.556	0.232	0.305
1990	0.415	1.584	0.235	0.319	0.397	1.562	0.235	0.310
1991	0.420	1.592	0.242	0.326	0.401	1.567	0.242	0.316
1992	0.437	1.605	0.271	0.346	0.415	1.577	0.271	0.335
1993	0.450	1.615	0.290	0.361	0.430	1.589	0.290	0.350
1994	0.453	1.618	0.294	0.364	0.434	1.594	0.294	0.354
1995	0.450	1.614	0.293	0.362	0.433	1.592	0.293	0.353
1996	0.452	1.612	0.297	0.364	0.435	1.592	0.297	0.355
1997	0.455	1.613	0.303	0.368	0.440	1.595	0.303	0.359
1998	0.444	1.600	0.296	0.358	0.428	1.580	0.296	0.349
1999	0.446	1.600	0.301	0.360	0.429	1.579	0.301	0.351
2000	0.464	1.615	0.325	0.380	0.447	1.594	0.325	0.371
2001	0.444	1.600	0.296	0.358	0.426	1.577	0.296	0.348
2002	0.462	1.614	0.323	0.378	0.444	1.591	0.323	0.368
2003	0.470	1.621	0.334	0.387	0.451	1.597	0.334	0.377

2004	0.478	1.629	0.343	0.396	0.460	1.605	0.343	0.385
2005	0.516	1.658	0.397	0.439	0.497	1.635	0.397	0.428
2006	0.457	1.615	0.308	0.371	0.437	1.590	0.308	0.360
2007	0.467	1.624	0.322	0.383	0.447	1.599	0.322	0.371
2008	0.460	1.616	0.317	0.376	0.438	1.588	0.317	0.364
2009	0.460	1.621	0.312	0.376	0.434	1.588	0.312	0.361
2010	0.465	1.626	0.319	0.381	0.438	1.591	0.319	0.366
2011	0.460	1.618	0.317	0.377	0.435	1.586	0.317	0.363
2012	0.461	1.620	0.317	0.378	0.435	1.587	0.317	0.363
2013	0.464	1.623	0.320	0.381	0.437	1.589	0.320	0.366
2014	0.468	1.627	0.325	0.385	0.441	1.591	0.325	0.370
2015	0.484	1.638	0.349	0.403	0.456	1.603	0.349	0.387
2016	0.478	1.635	0.340	0.397	0.450	1.598	0.340	0.380
2017	0.482	1.637	0.345	0.401	0.453	1.601	0.345	0.384

Table A5. Number of individuals and mean incomes, by group, and control totals. Nominal values in NOK (not CPI-adjusted)

			Number of individuals				Mean income				
									NA/NT		
Year	Population control	Income control (Mill NOK)	CG[18] taxpayers	MUN-CG taxpayers	NA/NT	Poor	CG taxpayers	MUN-CG taxpayers	(Upper Gini bound)	(Lower Gini bound)	Poor
1875	847 000	476	184 053	521 407	83 730	57 810	1 225	230	85	150	85
1888	919 000	442	65 845	410 098	365 502	77 555	2 537	345	76	150	76
1892	937 870	490	176 075	342 860	340 254	78 681	1 450	277	85	92	85
1893	944 840	500	102 542	421 050	340 318	80 930	2 039	338	83	113	83
1894	955 117	503	66 807	468 322	339 239	80 749	2 709	383	85	128	85
1895	968 945	514	68 227	476 723	345 533	78 462	2 712	387	93	129	93
1896	983 818	538	70 454	492 729	340 393	80 242	2 696	397	90	132	90
1897	999 315	559	75 578	505 845	337 693	80 199	2 700	401	91	134	91
1898	1 015 808	606	83 933	520 295	331 403	80 177	2 698	408	94	136	94
1899	1 031 501	639	91 422	521 528	337 821	80 730	2 669	419	95	140	95
1900	1 045 420	667	94 367	531 711	341 090	78 252	2 683	425	106	142	106
1901	1 058 452	657	95 767	548 176	332 368	82 141	2 625	427	108	142	108
1902	1 066 877	652	97 517	556 891	329 077	83 392	2 588	428	108	143	108
1903	1 071 397	648	96 431	557 972	330 361	86 634	2 574	428	107	143	107
1904	1 065 571	638	100 380	560 810	316 199	88 182	2 464	428	109	143	109
1905	1 070 722	654	99 463	569 071	318 628	83 560	2 537	426	118	142	118
1906	1 077 000	708	105 145	573 059	315 190	83 606	2 535	430	116	143	116
1907	1 088 673	749	113 288	577 144	315 810	82 431	2 519	433	116	144	116

[18] The Mun-CG division for 1875 and 1888 has been discussed in the section "Using data on the assisted poor".

1908	1 084 270	770	124 027	583 412	294 056	82 775	2 462	435	119	145	119
1909	1 102 688	784	130 278	589 821	298 630	83 959	2 472	437	117	146	117
1910	1 119 676	866	140 864	591 294	307 841	79 677	2 459	445	127	148	127
1911	1 122 989	920	153 321	599 031	290 504	80 133	2 462	457	127	152	127
1912	1 138 014	1 017	237 787	570 380	252 307	77 540	2 292	501	139	167	139
1913	1 181 740	1 130	256 299	586 611	261 662	77 168	2 486	498	149	166	149
1914	1 198 991	1 165	277 668	589 343	254 034	77 947	2 491	517	147	172	147
1915	1 191 118	1 590	315 126	570 701	230 296	74 995	3 241	546	159	182	159
1916	1 213 725	2 344	403 017	516 477	220 729	73 502	4 842	624	178	208	178
1917	1 234 220	2 785	414 844	561 408	188 332	69 636	6 040	574	224	224	224
1918	1 257 369	3 196	448 653	549 757	192 108	66 851	6 001	875	298	298	298
1919	1 274 625	3 890	556 348	476 176	179 495	62 606	5 959	819	395	395	395
1920	1 297 828	4 702	512 180	528 326	201 062	56 260	6 439	1 120	494	494	494
1921	1 320 416	3 512	448 155	521 486	285 654	65 122	5 161	974	507	507	507
1922	1 341 487	3 170	424 732	542 246	301 006	73 503	4 506	957	492	492	492
1923	1 359 382	3 140	412 585	561 104	309 925	75 768	4 317	933	469	469	469
1924	1 366 009	3 468	407 816	580 535	296 892	80 766	4 543	868	462	462	462
1925	1 382 259	3 491	409 671	581 312	305 945	85 331	4 578	848	480	480	480
1926	1 401 352	2 869	360 762	618 748	324 945	96 897	4 444	821	482	482	482
1927	1 416 889	2 587	332 276	644 336	332 594	107 683	4 248	797	485	485	485
1928	1 429 250	2 583	323 486	661 957	337 679	106 127	4 180	738	492	492	492
1929	1 455 069	2 656	330 210	668 110	348 444	108 305	4 252	713	451	451	451
1930	1 462 006	2 701	328 673	680 495	342 738	110 100	4 279	693	434	434	434
1931	1 484 265	2 331	290 127	704 092	364 673	125 373	4 220	682	371	371	371
1932	1 500 824	2 324	297 978	700 192	357 232	145 423	3 989	663	359	359	359
1933	1 520 458	2 323	300 982	706 177	360 450	152 850	3 878	649	352	352	352
1934	1 543 222	2 450	315 183	704 531	366 504	157 003	3 820	662	347	347	347
1935	1 565 806	2 627	330 193	712 185	365 219	158 209	3 922	673	354	354	354
1936	1 583 790	2 919	365 267	723 020	339 260	156 242	4 053	679	371	371	371
1937	1 610 577	3 372	409 369	717 636	334 786	148 786	4 366	672	378	378	378

1938	1 632 718	3 497	444 099	700 914	347 219	140 486	4 424	716	387	387	387
1939	1 654 129	3 755	471 654	712 892	331 743	137 840	4 498	670	406	406	406
1940	1 674 238	4 019	517 468	730 297	275 149	151 324	4 388	771	377	377	377
1941	1 688 313	5 134	664 652	653 608	278 337	91 716	4 530	828	523	523	523
1942	1 695 121	5 137	711 786	637 713	281 683	63 939	4 505	965	576	576	576
1943	1 704 634	5 223	739 956	618 432	291 888	54 358	4 623	868	555	555	555
1944	1 716 464	5 198	739 897	612 073	315 106	49 388	4 619	857	613	613	613
1945	1 730 001	6 330	768 327	596 787	315 131	49 756	4 599	918	576	576	576
1946	1 746 103	6 303	917 116	442 106	336 076	50 805	4 908	1 041	662	662	662
1947	1 752 946	7 456	979 409	416 528	308 940	48 069	5 572	1 081	733	733	733
1948	1 736 464	8 209	1 006 112	401 233	284 421	44 698	6 220	1 114	795	795	795
1949	1 733 690	8 800	1 076 360	372 694	243 250	41 386	6 266	1 382	835	835	835
1950	1 727 813	9 463	1 125 158	351 160	212 229	39 266	6 621	1 338	845	845	845
1951	1 721 099	11 472	1 026 214	455 814	199 979	39 092	8 156	2 091	908	908	908

Table A6. Population and income control totals 1951 and later

Year	Population control	Income control	Year	Population control	Income control	Year	Population control	Income control
1951	1 721 099	11 472	1974	1 989 957	74 603	1996	2 590 583	526 145
1952	1 723 350	12 556	1975	2 009 594	86 418	1997	2 608 585	558 102
1953	1 723 163	12 760	1976	2 032 203	98 343	1998	2 629 277	612 113
1954	1 723 981	14 055	1977	2 052 434	108 545	1999	2 652 168	649 219
1955	1 725 450	14 826	1978	2 076 830	128 727	2000	2 668 561	697 332
1957	1 741 998	17 685	1979	2 100 389	137 116	2001	2 683 319	731 486
1958	1 748 932	17 301	1980	2 126 458	156 663	2002	2 705 535	789 216
1959	1 758 814	18 382	1981	2 158 775	178 977	2003	2 726 116	828 107
1960	1 771 109	19 601	1982	2 190 717	201 213	2004	2 752 110	854 319
1961	1 788 908	21 349	1983	2 222 341	221 096	2005	2 786 213	918 836
1962	1 809 911	22 996	1984	2 254 414	244 354	2006	2 825 535	909 197
1963	1 833 869	24 916	1985	2 293 666	268 342	2007	2 879 690	1 007 117
1964	1 854 113	27 564	1986	2 330 892	303 474	2008	2 933 108	1 109 269
1965	1 872 800	30 590	1987	2 367 549	343 704	2009	2 979 896	1 124 151
1966	1 889 704	32 847	1988	2 402 329	370 905	2010	3 035 102	1 171 878
1967	1 904 805	35 865	1989	2 425 794	387 500	2011	3 079 228	1 244 480
1968	1 885 438	38 272	1990	2 450 457	408 447	2012	3 135 208	1 329 525
1969	1 900 571	41 543	1991	2 480 929	428 316	2013	3 187 678	1 399 572
1970	1 914 912	47 014	1992	2 508 283	449 394	2014	3 239 492	1 463 425
1971	1 934 029	52 095	1993	2 533 015	466 137	2015	3 281 505	1 553 267
1972	1 950 723	57 432	1994	2 553 029	475 796	2016	3 321 031	1 562 025
1973	1 970 938	64 933	1995	2 571 878	500 651	2017	3 358 711	1 616 731

Table A7: Parameters used for calculating four-class Gini. For calculation, see text. Note: G**=0.1333 (from z=0.4) for all years.

	Same for upper and lower bounds					Upper bound				Lower bound			
Year	p	n	m	c	a	h	g	g'	G*	h	g	g'	G*
1875	0.068	0.099	0.616	0.217	0.010	0.025	0.364	0.338	0.494	0.037	0.371	0.335	0.494
1888	0.084	0.398	0.446	0.072	0.013	0.076	0.499	0.424	0.461	0.137	0.533	0.395	0.461
1892	0.084	0.363	0.366	0.188	0.014	0.072	0.324	0.252	0.448	0.078	0.328	0.250	0.448
1893	0.086	0.360	0.446	0.109	0.013	0.070	0.447	0.377	0.452	0.090	0.459	0.369	0.452
1894	0.085	0.355	0.490	0.070	0.014	0.071	0.533	0.463	0.447	0.100	0.548	0.448	0.447
1895	0.081	0.357	0.492	0.070	0.014	0.077	0.538	0.461	0.485	0.101	0.550	0.449	0.485
1896	0.082	0.346	0.501	0.072	0.013	0.070	0.542	0.472	0.480	0.097	0.555	0.458	0.480
1897	0.080	0.338	0.506	0.076	0.013	0.068	0.532	0.465	0.488	0.094	0.545	0.452	0.488
1898	0.079	0.326	0.512	0.083	0.012	0.064	0.517	0.453	0.494	0.087	0.529	0.442	0.494
1899	0.078	0.328	0.506	0.089	0.012	0.062	0.505	0.443	0.492	0.086	0.518	0.432	0.492
1900	0.075	0.326	0.509	0.090	0.012	0.067	0.507	0.440	0.495	0.085	0.517	0.432	0.495
1901	0.078	0.314	0.518	0.090	0.013	0.068	0.517	0.449	0.481	0.085	0.526	0.441	0.481
1902	0.078	0.308	0.522	0.091	0.014	0.069	0.521	0.453	0.473	0.086	0.530	0.444	0.473
1903	0.081	0.308	0.521	0.090	0.014	0.069	0.525	0.456	0.469	0.087	0.534	0.447	0.469
1904	0.083	0.297	0.526	0.094	0.015	0.069	0.527	0.458	0.469	0.086	0.536	0.450	0.456
1905	0.078	0.298	0.531	0.093	0.015	0.073	0.527	0.455	0.469	0.084	0.533	0.449	0.456
1906	0.078	0.293	0.532	0.098	0.014	0.065	0.514	0.449	0.456	0.077	0.520	0.443	0.456
1907	0.076	0.290	0.530	0.104	0.013	0.062	0.500	0.438	0.456	0.074	0.506	0.432	0.353
1908	0.076	0.271	0.538	0.114	0.013	0.058	0.486	0.428	0.456	0.068	0.491	0.423	0.353
1909	0.076	0.271	0.535	0.118	0.013	0.057	0.476	0.419	0.456	0.068	0.482	0.414	0.353
1910	0.071	0.275	0.528	0.126	0.012	0.057	0.464	0.407	0.456	0.064	0.468	0.404	0.353
1911	0.071	0.259	0.533	0.137	0.011	0.051	0.450	0.399	0.456	0.059	0.455	0.395	0.353

1912	0.068	0.222	0.501	0.209	0.011	0.045	0.374	0.328	0.456	0.052	0.378	0.326	0.353
1913	0.065	0.221	0.496	0.217	0.010	0.045	0.345	0.300	0.353	0.049	0.348	0.299	0.353
1914	0.065	0.212	0.492	0.232	0.010	0.042	0.335	0.293	0.353	0.047	0.338	0.291	0.284
1915	0.063	0.193	0.479	0.265	0.008	0.031	0.257	0.226	0.353	0.034	0.260	0.226	0.284
1916	0.061	0.182	0.426	0.332	0.006	0.022	0.161	0.138	0.353	0.025	0.163	0.138	0.284
1917	0.056	0.153	0.455	0.336	0.006	0.021	0.132	0.112	0.353	0.021	0.132	0.112	0.284
1918	0.053	0.153	0.437	0.357	0.006	0.024	0.172	0.148	0.353	0.024	0.172	0.148	0.284
1919	0.049	0.141	0.374	0.436	0.006	0.025	0.127	0.103	0.353	0.025	0.127	0.103	0.284
1920	0.043	0.155	0.407	0.395	0.006	0.027	0.175	0.148	0.353	0.027	0.175	0.148	0.284
1921	0.049	0.216	0.395	0.339	0.009	0.051	0.222	0.171	0.353	0.051	0.222	0.171	0.284
1922	0.055	0.224	0.404	0.317	0.011	0.058	0.259	0.201	0.353	0.058	0.259	0.201	0.284
1923	0.056	0.228	0.413	0.304	0.011	0.058	0.272	0.214	0.353	0.058	0.272	0.214	0.284
1924	0.059	0.217	0.425	0.299	0.011	0.050	0.253	0.203	0.353	0.050	0.253	0.203	0.284
1925	0.062	0.221	0.421	0.296	0.012	0.054	0.251	0.197	0.353	0.054	0.251	0.197	0.284
1926	0.069	0.232	0.442	0.257	0.016	0.071	0.294	0.224	0.353	0.071	0.294	0.224	0.284
1927	0.076	0.235	0.455	0.235	0.020	0.083	0.327	0.245	0.353	0.083	0.327	0.245	0.284
1928	0.074	0.236	0.463	0.226	0.020	0.084	0.327	0.243	0.353	0.084	0.327	0.243	0.284
1929	0.074	0.239	0.459	0.227	0.018	0.078	0.311	0.234	0.353	0.078	0.311	0.234	0.284
1930	0.075	0.234	0.465	0.225	0.018	0.073	0.306	0.233	0.353	0.073	0.306	0.233	0.284
1931	0.084	0.246	0.474	0.195	0.020	0.078	0.338	0.260	0.353	0.078	0.338	0.260	0.284
1932	0.097	0.238	0.467	0.199	0.022	0.078	0.337	0.259	0.353	0.078	0.337	0.259	0.284
1933	0.101	0.237	0.464	0.198	0.023	0.078	0.338	0.260	0.353	0.078	0.338	0.260	0.284
1934	0.102	0.237	0.457	0.204	0.022	0.074	0.333	0.258	0.353	0.074	0.333	0.258	0.284
1935	0.101	0.233	0.455	0.211	0.021	0.071	0.322	0.251	0.353	0.071	0.322	0.251	0.284
1936	0.099	0.214	0.457	0.231	0.020	0.063	0.296	0.233	0.353	0.063	0.296	0.233	0.284
1937	0.092	0.208	0.446	0.254	0.017	0.054	0.255	0.201	0.353	0.054	0.255	0.201	0.284
1938	0.086	0.213	0.429	0.272	0.016	0.054	0.246	0.192	0.284	0.054	0.246	0.192	0.284
1939	0.083	0.201	0.431	0.285	0.015	0.051	0.225	0.174	0.315	0.051	0.225	0.174	0.284
1940	0.090	0.164	0.436	0.309	0.014	0.040	0.231	0.191	0.315	0.040	0.231	0.191	0.284
1941	0.054	0.165	0.387	0.394	0.009	0.038	0.184	0.147	0.315	0.038	0.184	0.147	0.284

1942	0.038	0.166	0.376	0.420	0.007	0.039	0.193	0.155	0.315	0.039	0.193	0.155	0.284
1943	0.032	0.171	0.363	0.434	0.006	0.037	0.167	0.131	0.315	0.037	0.167	0.131	0.284
1944	0.029	0.184	0.357	0.431	0.006	0.043	0.170	0.127	0.315	0.043	0.170	0.127	0.284
1945	0.029	0.182	0.345	0.444	0.005	0.033	0.163	0.130	0.315	0.033	0.163	0.130	0.284
1946	0.029	0.192	0.253	0.525	0.005	0.041	0.130	0.089	0.315	0.041	0.130	0.089	0.284
1947	0.027	0.176	0.238	0.559	0.005	0.035	0.109	0.074	0.315	0.035	0.109	0.074	0.284
1948	0.026	0.164	0.231	0.579	0.004	0.032	0.096	0.065	0.315	0.032	0.096	0.065	0.315
1949	0.024	0.140	0.215	0.621	0.004	0.027	0.096	0.069	0.315	0.027	0.096	0.069	0.315
1950	0.023	0.123	0.203	0.651	0.004	0.022	0.080	0.058	0.318	0.022	0.080	0.058	0.318
1951	0.023	0.116	0.265	0.596	0.003	0.019	0.119	0.100	0.284	0.019	0.119	0.100	0.284

Table A8. Mean of the upper and lower bounds of the Gini coefficient, the upper tail (above median) Gini coefficient, the ratio between the lower tail (above the median) and upper tail (below the median) mean incomes, affluence, and the three alternative measures of overall inequality

year	Gini coefficient (average)	Gini (above median)	$\frac{\mu_L}{\mu_U}$	Affluence	C_1	C_2	C_3
1875	.548	.508	.219	.492	.634	.548	.501
1888	.605	.489	.132	.544	.685	.605	.553
1892	.640	.515	.110	.577	.718	.640	.584
1893	.619	.491	.123	.552	.702	.619	.563
1894	.593	.461	.138	.522	.681	.593	.536
1895	.588	.462	.145	.518	.673	.588	.532
1896	.584	.454	.147	.512	.673	.584	.527
1897	.585	.460	.149	.513	.674	.585	.528
1898	.588	.467	.151	.516	.678	.588	.531
1899	.591	.469	.148	.519	.681	.591	.534
1900	.586	.469	.154	.515	.674	.586	.530
1901	.574	.459	.163	.503	.664	.574	.518
1902	.568	.454	.168	.496	.658	.568	.512
1903	.567	.451	.168	.495	.657	.567	.510
1904	.559	.444	.176	.486	.650	.559	.501
1905	.558	.450	.179	.486	.646	.558	.502
1906	.566	.453	.173	.493	.657	.566	.509
1907	.569	.452	.169	.495	.662	.569	.510
1908	.565	.454	.175	.492	.659	.565	.507
1909	.569	.458	.173	.495	.662	.569	.510
1910	.572	.461	.170	.499	.666	.572	.513
1911	.572	.461	.173	.497	.667	.572	.512
1912	.561	.455	.182	.487	.658	.561	.500
1913	.570	.457	.171	.496	.667	.570	.507

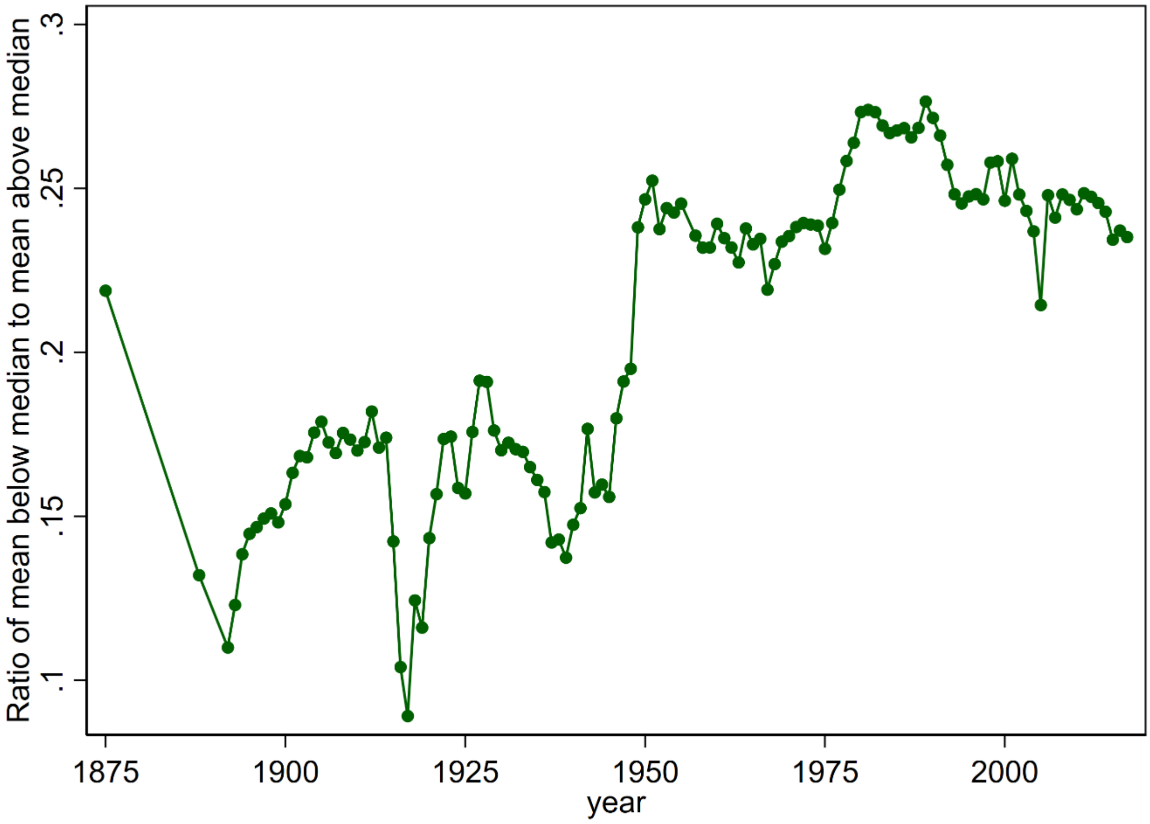
1914	.559	.438	.174	.483	.661	.559	.494
1915	.597	.465	.142	.522	.698	.597	.530
1916	.626	.458	.104	.547	.731	.626	.553
1917	.645	.476	.089	.570	.745	.645	.573
1918	.594	.430	.124	.514	.700	.594	.522
1919	.573	.372	.116	.486	.685	.573	.498
1920	.560	.391	.143	.478	.670	.560	.489
1921	.558	.418	.157	.484	.654	.558	.492
1922	.543	.414	.174	.470	.637	.543	.479
1923	.544	.417	.174	.471	.639	.544	.480
1924	.565	.435	.159	.492	.661	.565	.500
1925	.568	.441	.157	.497	.661	.568	.504
1926	.555	.456	.176	.492	.637	.555	.497
1927	.542	.459	.191	.483	.617	.542	.488
1928	.546	.471	.191	.490	.618	.546	.494
1929	.564	.480	.176	.506	.639	.564	.509
1930	.572	.486	.170	.513	.648	.572	.517
1931	.573	.492	.172	.515	.648	.573	.520
1932	.573	.488	.171	.514	.650	.573	.519
1933	.574	.487	.170	.514	.651	.574	.519
1934	.576	.480	.165	.514	.656	.576	.519
1935	.580	.481	.161	.517	.662	.580	.523
1936	.583	.480	.157	.519	.667	.583	.524
1937	.599	.485	.142	.533	.685	.599	.538
1938	.585	.454	.143	.515	.677	.585	.520
1939	.594	.464	.137	.525	.683	.594	.529
1940	.574	.430	.147	.498	.675	.574	.505
1941	.543	.375	.152	.462	.650	.543	.472
1942	.513	.348	.177	.431	.621	.513	.443
1943	.526	.346	.157	.442	.635	.526	.455

1944	.524	.349	.160	.442	.628	.524	.454
1945	.526	.337	.156	.438	.641	.526	.452
1946	.494	.300	.180	.401	.611	.494	.421
1947	.491	.300	.191	.394	.611	.491	.416
1948	.496	.315	.195	.400	.616	.496	.422
1949	.468	.315	.238	.375	.590	.468	.397
1950	.464	.318	.247	.372	.589	.464	.394
1951	.446	.284	.252	.350	.577	.446	.373
1952	.430	.233	.238	.331	.560	.430	.353
1953	.420	.221	.244	.321	.550	.420	.343
1954	.421	.224	.243	.324	.548	.421	.345
1955	.423	.230	.245	.325	.552	.423	.347
1957	.437	.245	.236	.339	.565	.437	.360
1958	.437	.238	.232	.336	.568	.437	.359
1959	.434	.233	.232	.334	.564	.434	.356
1960	.427	.226	.239	.326	.556	.427	.349
1961	.432	.227	.235	.329	.565	.431	.352
1962	.432	.221	.232	.327	.569	.432	.351
1963	.436	.223	.227	.331	.571	.436	.354
1964	.426	.216	.238	.322	.562	.426	.346
1965	.428	.213	.233	.323	.562	.428	.347
1966	.425	.212	.235	.321	.559	.425	.345
1967	.441	.227	.219	.338	.572	.441	.360
1968	.435	.226	.227	.333	.567	.435	.355
1969	.431	.227	.234	.330	.561	.431	.352
1970	.430	.228	.235	.329	.560	.430	.351
1971	.427	.224	.238	.326	.557	.427	.349
1972	.425	.223	.239	.324	.556	.425	.347
1973	.426	.223	.239	.325	.557	.426	.347
1974	.425	.220	.239	.323	.556	.425	.345

1975	.429	.217	.232	.325	.561	.429	.348
1976	.422	.211	.239	.318	.556	.422	.342
1977	.414	.209	.250	.311	.548	.415	.335
1978	.408	.209	.258	.307	.540	.408	.330
1979	.404	.207	.264	.303	.535	.404	.326
1980	.396	.203	.273	.296	.527	.396	.319
1981	.396	.205	.274	.297	.526	.396	.319
1982	.397	.208	.273	.299	.525	.397	.321
1983	.400	.212	.269	.303	.527	.400	.325
1984	.403	.216	.267	.306	.529	.403	.328
1985	.404	.220	.268	.308	.530	.404	.329
1986	.405	.224	.268	.310	.530	.405	.330
1987	.409	.227	.266	.313	.535	.409	.334
1988	.407	.229	.268	.313	.533	.407	.333
1989	.401	.232	.277	.310	.523	.401	.329
1990	.406	.235	.272	.314	.528	.406	.333
1991	.411	.242	.266	.321	.531	.411	.339
1992	.426	.271	.257	.341	.539	.426	.357
1993	.440	.290	.248	.356	.551	.440	.372
1994	.443	.294	.245	.359	.556	.444	.375
1995	.442	.293	.248	.357	.556	.442	.373
1996	.444	.297	.248	.360	.557	.444	.376
1997	.447	.303	.247	.364	.561	.447	.380
1998	.436	.296	.258	.353	.548	.436	.369
1999	.437	.301	.258	.356	.548	.438	.372
2000	.456	.325	.246	.376	.564	.456	.391
2001	.435	.296	.259	.353	.547	.435	.369
2002	.453	.323	.248	.373	.561	.453	.389
2003	.461	.334	.243	.382	.566	.461	.397
2004	.469	.343	.237	.391	.574	.469	.406

2005	.506	.397	.214	.433	.603	.506	.447
2006	.447	.308	.248	.366	.555	.447	.381
2007	.457	.322	.241	.377	.564	.457	.392
2008	.449	.317	.248	.370	.554	.449	.385
2009	.447	.312	.247	.369	.550	.447	.383
2010	.451	.319	.244	.374	.552	.451	.388
2011	.448	.317	.249	.370	.550	.448	.384
2012	.448	.317	.247	.371	.550	.448	.385
2013	.451	.320	.246	.373	.552	.451	.387
2014	.454	.325	.243	.378	.555	.454	.391
2015	.470	.349	.234	.395	.568	.470	.408
2016	.464	.340	.237	.389	.562	.464	.402
2017	.468	.345	.235	.393	.565	.468	.406

Figure A8. Mean income below the median relative to mean income above the median



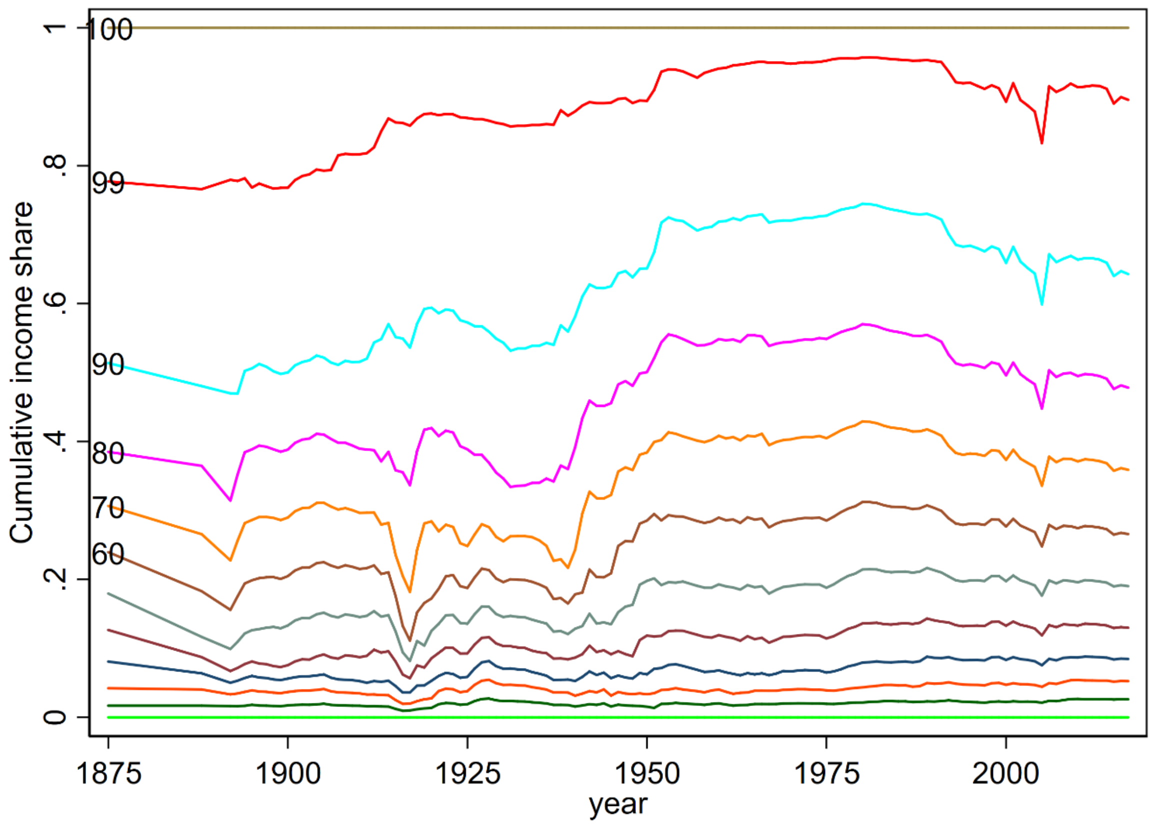
Note: The components are estimated using the same distribution as the one used for obtaining Gini coefficients in Figure 3. The calculation is based on the mean of upper and lower bound estimates. For sources, methods and assumptions, see text.

Appendix G: Income shares of income groups and decomposition of the top 10 per cent

Income shares

Figure A9 shows the development of cumulative income shares between 1875 and 2017, based on the average of the upper and lower bound income distributions. The interval between the topmost line at 100 per cent and the next line at 99 per cent shows the share of the top 1 per cent. We observe that this decreases from more than 20 per cent at the beginning of the period to around 10 per cent at the end. The second interval shows the income share of the 90-99 per cent, the next the share of the 80-90 per cent, and so on. The income share of the lowest 10 per cent is largely based on data on recipients of poverty support, and ranges between just less than 1 per cent in the year with the lowest share to around 2.7 per cent in the year with the highest share. A detailed table of estimated income shares is available as an online appendix.

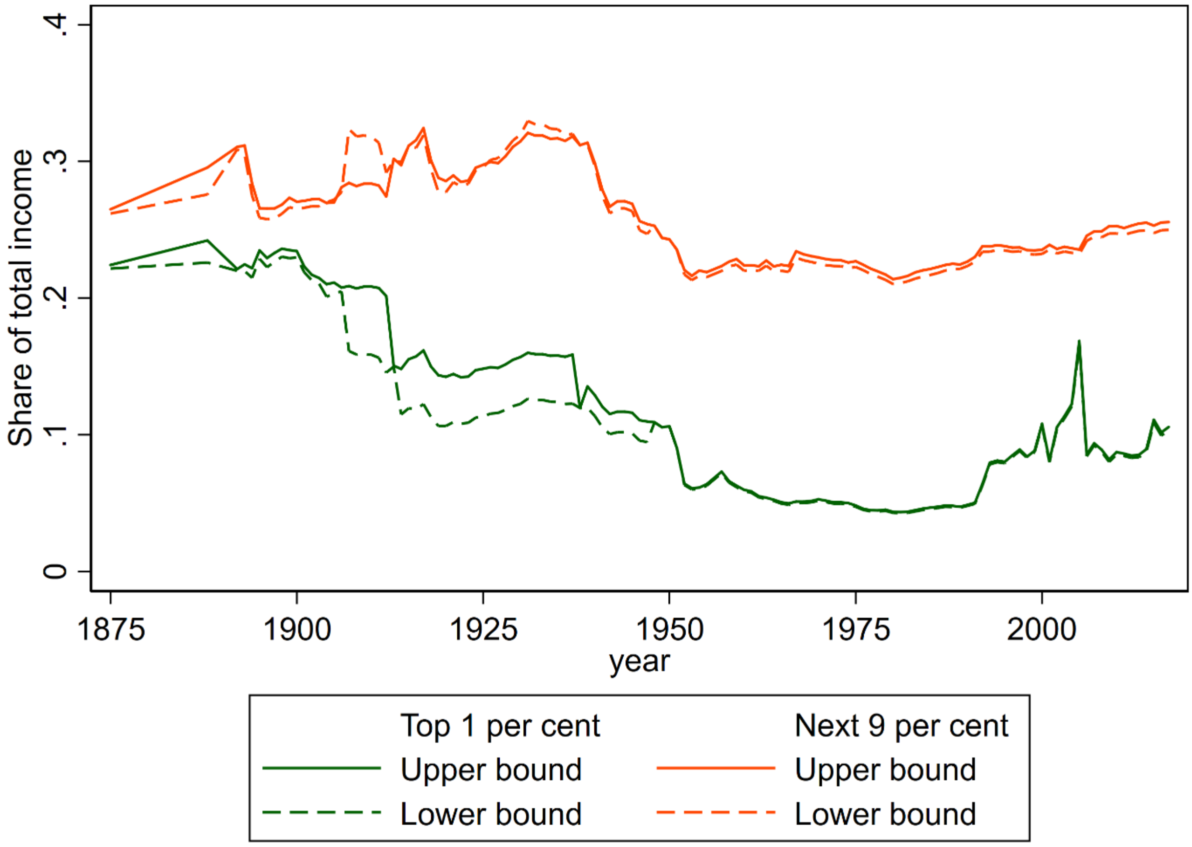
Figure A9. Top 1 and decile-specific shares of income in Norway, 1875-2017



Note: Each line in Figure A3 (labelled with the a number *i* shows the cumulative share of income of the lowest *i* per cent of the population, e.g. the gap between the line labelled 99 and the line labelled 100 is the top 1% income share, the gap between 90 and 99 is the 90-99% income share, the distance between 80 and 90 is the 80-90% income share, and so on. The components are estimated using the same distribution as the one used for obtaining Gini coefficients in Figure 3. The graphs display the mean of the upper and lower bound estimates. For sources, methods and assumptions, see text.

Figure A10 shows the development of the top 1 per cent income share together with the development of the income share of the top 90-99 per cent (e.g. the next 9 per cent) for our lower and upper bound assumptions. We observe that the assumptions matter slightly for the calculation of upper and lower bounds in the prewar period, due to the different Pareto distributions used for imputation of the top of the income distribution in years for which detailed data is not available. The income share of the top 90-99 per cent is shown to be high and moderately increasing until around 1940, to decrease over a 10-year period. In the postwar period, this share is low and stable but it increases slightly from 1980 onward.

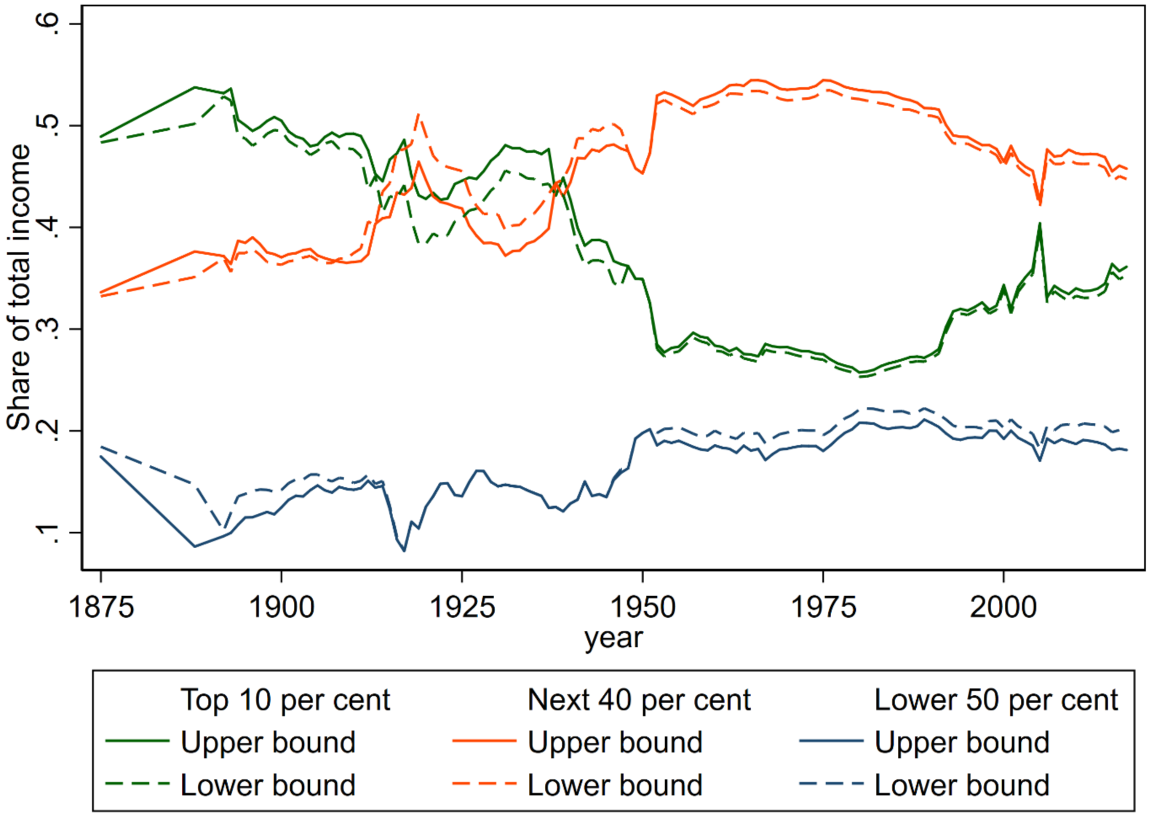
Figure A10. Top 1 and 90-99 per cent income shares



Note: The components are estimated using the same distribution as the one used for obtaining Gini coefficients in Figure 3. For sources, methods and assumptions, see text.

Figure A11 shows the development of the income shares of the top 10 per cent, top 50-90 per cent and the bottom 90 per cent over time.

Figure A11. Top 10, 50-90 and 0-50 per cent income shares



Note: The components are estimated using the same distribution as the one used for obtaining Gini coefficients in Figure 3. For sources, methods and assumptions, see text.

Decomposition of the overall Gini coefficient with respect to the income share of the top 10 per cent and the Gini coefficient of the remaing 90 per cent

In this appendix section, we examine how the results in Section 4.1 on the relationship between the top 1 per cent income share and the Gini coefficient correspons to the relationship between the top 10 per cent income share and the Gini coefficient.

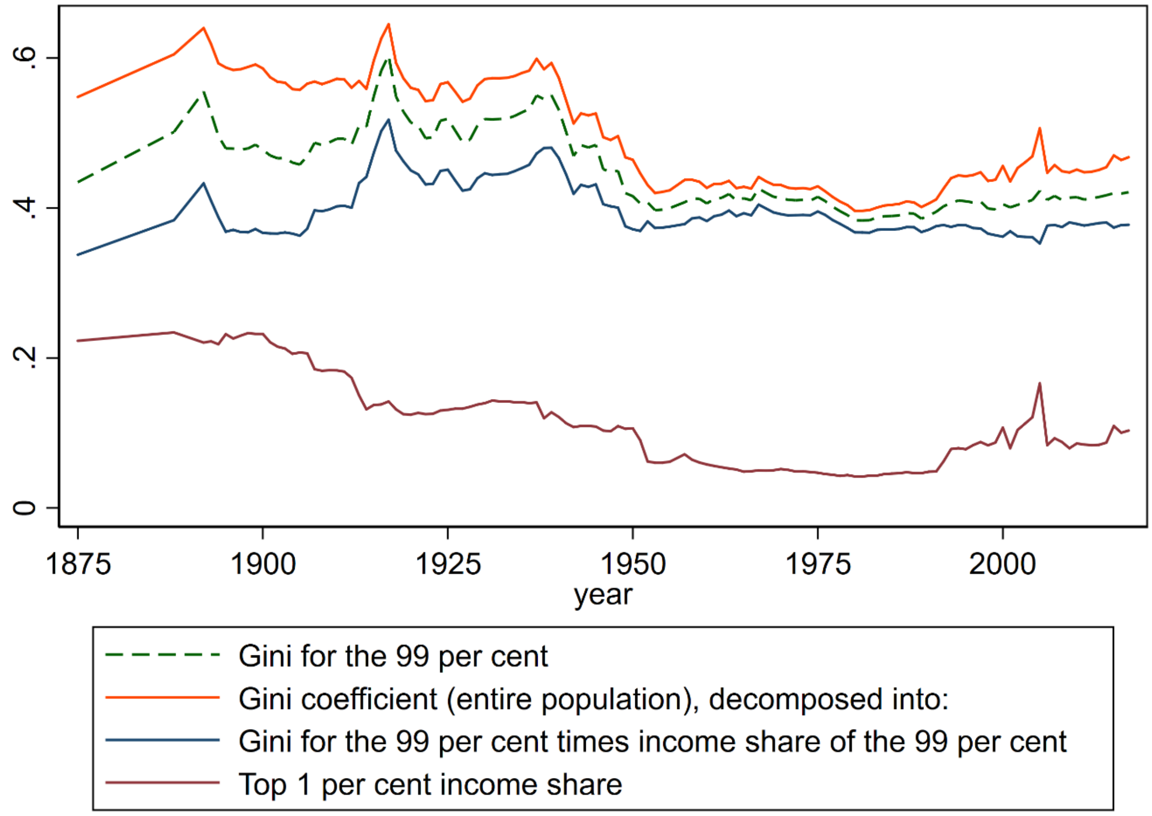
To this end, we rely on the decomposition

$$G = x \cdot S \cdot G^{top\, x\%} + (1 - S) \cdot G^{bottom\,(100-x)\%} - 0.01 \cdot (1 - S) \cdot G^{bottom\,(100-x)\%} + S - x/100$$

which can, for low x, be simplified to

$$G \approx (1 - S) \cdot G^{bottom(100-x)\%} + S$$

Figure A12. Decomposition of Gini coefficient into top 1 per cent income share and Gini coefficient components

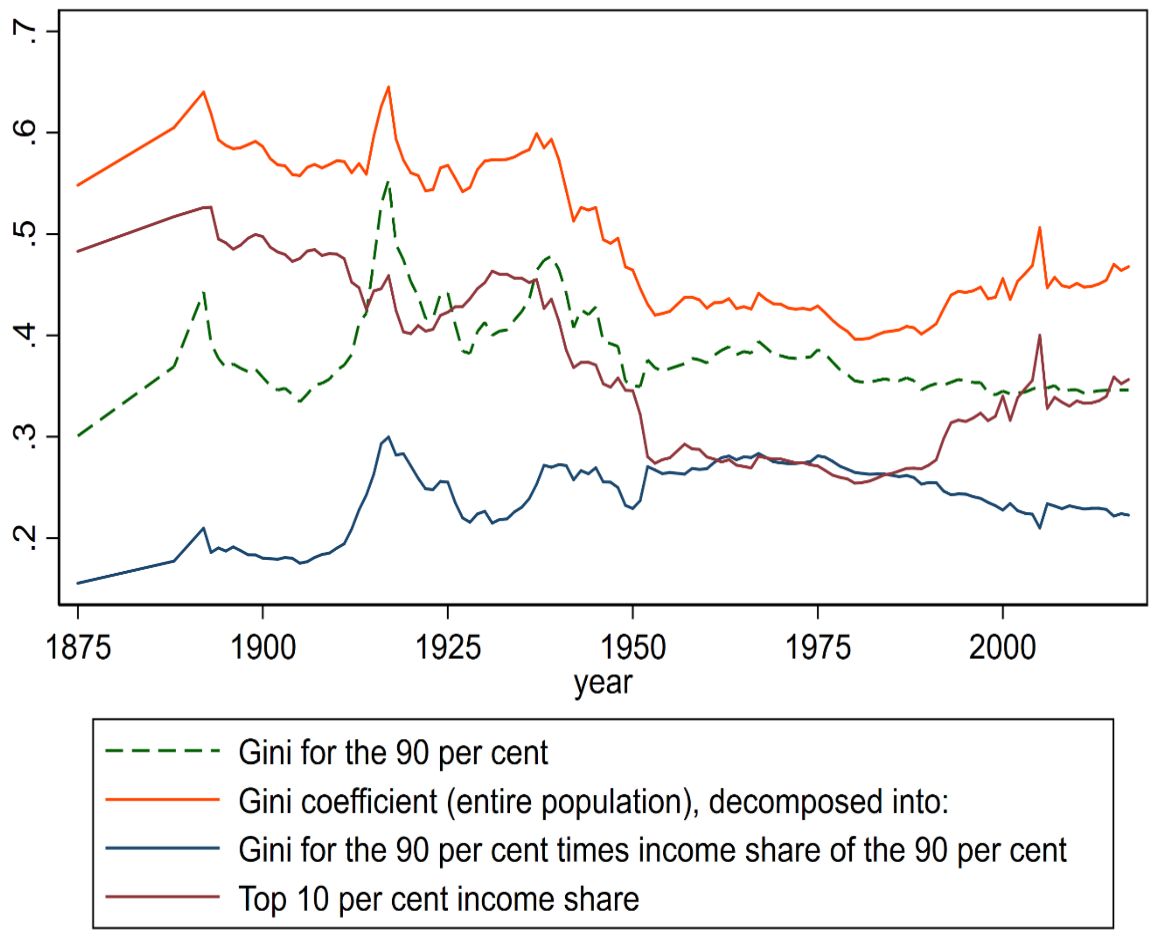


Note: The components are estimated using the same distribution as the one used for obtaining Gini coefficients in Figure 3. The graphs display the mean of the upper and lower bound estimates. For sources, methods and assumptions, see text.

Figure A12 for x=1 includes the graphs provided by Figure 5, but does also include the evolution of the Gini of the bottom 99 per cent (the dashed line). While there are similarities between the development of the top 1per cent income share and the Gini coefficient for the 99 per cent times the income share of the 99 per cent, there are also important differences, notably for the changes during the mid-twentieth century.

We now turn to a similar decomposition for the top 10 per cent. In this case, the approximation proves to be less precise. Most importantly, the last term (x) of the decomposition is 0.10 whereas it was 0.01 as for the 1 per cent decomposition case, which means that the remaining components no longer add up to the overall Gini coefficient. This makes the interpretation of changes more demanding as we omit a term that is constant across time with terms that vary across time. Moreover, the two remaining terms that are omitted in the approximation show to vary somewhat over time, but these terms are small compared to the two terms that defines the approximation of the Gini coefficient.

Figure A13. Decomposition of Gini coefficient into top 10 per cent income share and Gini coefficient components



Note: The components are estimated using the same distribution as the one used for obtaining Gini coefficients in Figure 3. The graphs display the mean of the upper and lower bound estimates. For sources, methods and assumptions, see text.

Figures A12 and A13 demonstrate that the decompositions for x=1 and x=10 roughly spoken provide congruent pictures of the evolution of top income shares and the associated Gini coeffcients. However, during the mid-twentieth century the top 10 per cent income share shows to decrease more sharply than the top 1 per cent. Moreover, the decrease of the Gini of the bottom 90 per cent times the income share of the bottom 90 per cent shows to be less systematic than the decrease of of the Gini of the bottom 99 per cent times the income share of the bottom 99 per cent.