# The regional dispersion of income inequality in nineteenth-century Norway

Jørgen Modalsli



The final authenticated version is available at:

Modalsli, J. (2018). The regional dispersion of income inequality in nineteenth-century Norway. Explorations in Economic History, 67, 62-79. doi: https://doi.org/10.1016/j.eeh.2017.09.001

# The regional dispersion of income inequality in nineteenth-century Norway

Jørgen Modalsli*

September 8, 2017

**Abstract**

This paper documents, for the first time, municipality- and occupation-level estimates of income inequality between individuals in a European country in the nineteenth century, using a combination of several detailed data sets for Norway in the late 1860s. Urban incomes were on average 4.5 times as high as rural incomes, and the average city Gini coefficient was twice the average rural municipality Gini. All high- or medium-income occupation groups exhibited substantial within-occupation income inequality.

Across municipalities, income inequality is higher in high-income municipalities, and lower in muncipalities with high levels of fisheries and pastoral agriculture. While manufacturing activity is positively correlated with income inequality, the association is not apparent when other economic factors such as the mode of food production is accounted for.

The income Gini for Norway as a whole is found to have been 0.546, slightly higher than estimates for the UK and US in the same period.

**JEL codes**: N33, D31, O15

**Keywords:** Income inequality, economic development, rural-urban differences, economic history, Kuznets curve

*Statistics Norway, Research Department. E-mail: mod@ssb.no. I thank Rolf Aaberge, Kjetil Telle, Pål Thonstad Sandvik, two referees, the Editor, and participants at seminars and conferences for helpful comments, and Jeanette Strøm Fjære for excellent research assistance. Support from the Research Council of Norway is acknowledged.

# 1 Introduction

Most theories of modern growth have implications for the change of income inequality across stages of development. In the canonical "simple model" by Kuznets (1955), income inequality increases with early modernization, as the population gradually shifts from low-income, agriculture to the higher-income (and relatively higher-inequality) non-agricultural economy. However, there is still little evidence of how and to what extent this process actually operated during the period of industrialization of Europe in the nineteenth and early twentieth century. In most cases, this is due to limited data availability. Nationally harmonized income taxation was rare until the turn of the twentieth century, and while there was increasing concern about the conditions of the very poor, few countries kept any records of the overall distribution of income. Where such information does exist, it is typically tabulated for countries as a whole and does not allow for a decomposition across geographic regions or occupations.

Better data on regional development in nineteenth-century Western Europe dramatically increases the number of observations useful for evaluating typical theories of income inequality and growth. Follow-up studies of the Kuznets curve, originally proposed based on only five data points from 1850 onward, have mainly examined the relationship between inequality and growth directly (see for example Van Zanden (1995)); the proposed cross-section characteristics remain underexplored. For any given country, there are events separate from secular economic development (such as wars) that have major influences on inequality trajectories. For this reason, it is of interest to see how the classical theories of economic structure and income inequality are reflected in the cross-section within a country, in this way holding the institutional setting and technological environment constant. The study of such a cross-section of inequality and economic development is one aim of this paper, which calculates complete income distributions for 491 municipalities and 19 occupation groups in Norway in 1868.

Besides structural change, other explanations have been put forward to explain observed positive relationships between economic development and income inequality. Milanovic *et al.* (2011) argue that when mean incomes are low, there is little surplus left for concentration among the rich if the poor population is to receive a subsistence income. In a sample of pre-industrial economies, they find that such subsistence income limits on feasible inequality rates can explain much historical cross-country variation in income inequality.

Climate in general, and agricultural production in particular, can also be seen as drivers of inequality. A geographical area where large agricultural units have an advantage is more likely to experience high inequality than an area where smaller units are more beneficial. An extended version of this argument was applied to the Americas by Engerman & Sokoloff (2002), who also emphasize the role geography-induced inequality can play in the development of institutions. Related institutional mechanisms are proposed by Acemoglu *et al.* (2005) and Galor *et al.* (2009). Easterly (2007) finds that countries with higher suitability for growing wheat rather than sugar experience lower income inequality.

In the present paper, a combination of several unique sources of data on economic conditions in Norway are utilized to reconstruct the income distribution as accurately as possible. A comprehensive survey of income distributions, conducted by the central government, is combined with archival data on wage distributions as well as a digitized version of the 1865 census to provide an estimate of income inequality. While the main purpose of the original survey was to gauge the impact of proposed electoral reforms, the other data sources facilitate extending the estimate to the full population of men aged 25 or above. In contrast to previous studies, the measure of income inequality is calculated directly from (reported) incomes rather than measures of status (incomes, earnings) imputed to social groups. Excellent census data makes it possible to have a precise definition of the population under study.

Three implications of the prevailing theories will be examined. First, as mentioned above, the proposed Kuznetsian relationship between income inequality and economic development. Second, the limitation of low mean incomes on feasible income inequality. Third, whether certain agricultural modes are more conducive to inequality than others.

Including data on income distributions for subnational regions makes it possible to differentiate the impact of governance of countries as a whole (in a broad sense, institutions) and factors that vary within countries. Norway in particular has a diverse geography, ranging over fourteen degrees of latitude, and substantial variation in average rainfall, altitude and type of traditional agriculture. Detailed data on regional inequality thus provides some information on the content of the "black box" of how geographic conditions influence economic development.

The paper contributes to the empirical literature on income distributions in the nineteenth century or earlier. Estimates are available only for a very limited number of countries and often have to rely on other economic characteristics as proxies for income. Typical examples are Lindert (2000) for the United Kingdom (several years), Lindert & Williamson (2012) for the United States 1774-1860, and Nafziger & Lindert (2012) for Russia 1904. For all these papers the authors use some combination of social tables (giving mean incomes by social class or occupation, but not dispersion within these groups) and occupational wages (also averages). The focus on labor income often makes it necessary to add income from property; in some cases, micro data for this is available. The snapshots give valuable information about the inequality in a given country in a given time period, but are hard to compare across countries. In many cases, the unit of measurement (household or individual; age and/or gender restriction) is not specified. Other sources of inequality estimation are based mainly on one data source, such as property registers (Alfani, 2013; Nicolini & Ramos Palencia, 2016), house rent distributions (Van Zanden, 1995) and wages (Clark, 2005). There are also papers aiming to provide long-term estimates of inequality based on tax data running into the twentieth century. Initially, these were based on top incomes only (Atkinson & Piketty, 2007); estimates of full distributions include Kopczuk *et al.* (2010) and Aaberge *et al.* (2016).

In addition to estimates of national between-individual (or household) inequality, there is a literature seeking to estimate regional differences within countries, typically only with the mean incomes of each region are taken into account. For example, Enflo & Roses (2015) describe interregional inequality for Sweden between 1860 and 2010 and Martines-Galarraga *et al.* (2015) do likewise for Spain. There is also a substantial body of work examining regional income inequalities and convergence in recent decades (for example Ezcurra *et al.*, 2005; Gennaioli *et al.*, 2013).

The papers that are closest to the present one in terms of approach consider within-country variation in intra-regional inequality, though with larger regions and less precise measures of income inequality. Tapia & Martines-Galarraga (2013) use the proposed relationship between mean and subsistence to infer inequality from wage data, and and find substantial differences in the evolution of inequality in Spanish regions between 1860 and 1913. Nafziger & Lindert (2012), using social tables, find that in Russia in 1904, inequality was higher in provinces with a higher mean income, and highest in Moscow and Saint Petersburg. Clark & Gray (2014) find some support of the first stage of the Engerman-Sokoloff hypothesis in Britain in the mid-nineteenth century. They find that equality, measured as the ratio of farmers to all men employed in farming, is negatively correlated with warmer temperatures and longer growing seasons.

The present paper is structured as follows. First, Section 2 gives a brief overview of 1860s Norway. Section 3 goes into great detail on how the income distribution for Norway in 1868 used in this paper is constructed. It differs from previous studies in the high level of regional disaggregation as well as the use of detailed tabulations of income ranges by municipality and occupation. The resulting data set, with imputed incomes for all 373,517 individuals, is available as an Online Appendix.

The new national estimates are presented in Section 4. Overall income inequality among men aged above 25 is found to be high, with a Gini coefficient of 0.546. There are substantial differences between rural and urban areas, with urban mean incomes 4.5 times higher than rural incomes on average, and an overall urban Gini coefficient nearly twice the rural Gini coefficient. This directly supports Kuznets' assertion that inequality in rural areas was lower than in urban areas at the time of industrialization. There are also large differences between regions of the country and (as expected) between occupation groups.

Section 5 presents Gini coefficient estimates at the municipal level, and describes associations between income inequality and various economic characteristics of the municipalities. Evidence is found in support of the "subsistence income" relationship, as municipalities with higher mean income exhibit higher income inequality. Manufacturing, crop value and closeness to cities are found to be positively associated with income inequality, while the extent of pastoralism and fisheries are negatively associated with income inequality. However, only the food production modes show up as significant coefficient estimates in a joint framework. There is evidence that land inequality in 1838 (when land tax records were updated) is strongly associated with income inequality thirty years later.

Section 6 compares the results to existing estimates from other countries and discusses some possible robustness checks regarding the assumptions that need to be made to arrive at an estimate of income inequality. In general, the results presented here are robust to alternative assumptions or to a tentative conversion of the men-aged-above-25 basis to a household basis.

# 2 Background: Norway in the nineteenth century

## 2.1 Political, demographic and economic context

In the 1860s, when the data used in this paper was collected, Norway was still a predominantly rural and relatively poor economy. Estimates of national accounts put Norwegian GDP per capita at around 44 percent of the United Kingdom, though above several Mediterranean economies (Bolt & van Zanden, 2013). A majority of the farmland was privately owned, and farms were on average smaller than the European average (Hodne & Grytten, 2000, p. 60). The population at the 1865 census was 1.7 million, with a median age of 23. After several years of high population growth (a net birth surplus of 1.2 per cent after 1850) there was a substantial population pressure, which was manifesting itself through rapidly increasing emigration rates. In the 1860s, Norway was still predominantly a rural society, with only 15.6 percent of the population residing in cities in 1865. While both cities and rural municipalities had some measure of local government, they were treated as qualitatively different by the central authorities, with separate legislation on issues such as trade rights and education systems.

After the temporary end of the Napoleonic Wars in 1814, sovereignty of Norway was transferred to Sweden from Denmark. In the process, Norway was able to obtain substantial internal self-rule, with a separate parliament. The following decades saw the emergence of a "civil servant state" (Seip, 1997), with a small group of educated families controlling much of civil society. The independent farmers gradually gained a strong political voice, culminating in the establishment of a parliamentary system (whereby the Cabinet answered to the elected representatives of the parliament) in 1884. Among the key political changes during the century was the gradual dissolution of trade privileges from 1854 to 1866 (Seip, 1997, p. 131), leading to a more market-oriented economy. With the farmers in power, an emphasis was placed on low public expenditure, with no state income tax being collected between 1836 and 1892 (Gerdrup, 1998). Most of the income of the central government induring this period was derived from import and export duties. Tariffs, however, were gradually decreased after 1860 (Seip, 1997, 1: 137).

The school system in Norway was relatively comprehensive for its time. Examination in Bible studies, organized by the state church, were mandatory from 1736, and more comprehensive education laws had been introduced already in 1827 (Hodne & Grytten, 2000, p. 71). Public hospitals were established in the 1850s, a law on public health in 1860, and a poverty law in 1863 (Seip, 1997, 1: 141). The poverty law was widely debated, with Sundt (1855, chap. 1) describing the common sentiment at the time that generous poverty laws would increase fertility among the poor and merely exacerbate the problem of poverty.

In the 1860s, Norway was still on the eve of industrial development. Large cultural and economic differences prevailed between rural and urban areas (Try, 1979). Most of the population had historically belonged to to one of three social classes: farmers, cottagers or servants. There was a strong element of occupational change over the life-cycle, with most individuals spending some time as servants or in similar occupations before moving on to other work; in 1865, two-thirds of all servants were younger than 25. The cottagers lived on land belonging to larger farms. They had an obligation to work for the farmer or to pay rent in kind or money, and in many cases children did not inherit the plot. As there was not much room to establish new farms, much of the population growth translated into growth in the cottager population, with the population reaching its largest point in 1855. In the 1860, a fourth, large occupation group was emerging: the working class. Together with emigration to North America, industrialization relieved the population pressure in the agricultural sector and facilitated a decrease in the cottager population.

Norway's industrial development started in the 1840s with textiles and mechanical industries (Hodne & Grytten, 2000, p. 191), though the first steam engines were already in use in 1831. After further industrialization in the 1850s, a total of 235,000 individuals (15 percent of the labor force) was listed in the 1865 census as being connected to industries (Norwegian Department of the Interior, 1868, p. 128-129). The textile industry was largest, with slightly above 50,000 employees, followed by lumber. A new wave of industrialization followed in the 1870s.

The Norwegian economy was tightly integrated with other countries. In 1868, grain and other foodstuffs accounted for more than half of total imports. The main exports were fish and lumber. Measured in the traditional way, Norway ran a large trade deficit; this was, however, more than compensated for by a large merchant fleet. The total gross income from this activity was nearly as large as all traditional exports combined. Following the repeal of the Navigation Act in Great Britain in 1849, a large share of this shipping occurred between foreign ports; in 1868 this constituted more than two-thirds of the total shipping surplus (Norwegian Department of the Interior, 1870).[1]

## 2.2 Incomes and income inequality

The long-run development of income inequality in Norway as a whole since 1875 has been estimated by Aaberge *et al.* (2016). The Gini coefficient is found to have been high in the late nineteenth and early twentieth century, with some fluctuation over time. Inequality started to fall in the late 1930s and remained low for several decades, before an increase in the late twentieth century. Estimates based on aggregate tax data can be constructed from most years from 1892 onward, but cannot be decomposed at the regional or occupation level. Moreover, Gini coefficients estimated from aggregate tax data refer to households rather than to individuals as in the present paper.

There also exist estimates of top income shares for Norway based on tax tabulations (Aaberge & Atkinson, 2010) and city Gini coefficients compiled by Soltow (1965). Soltow went through the tax archives in eight Norwegian cities to create a series of city Ginis ranging from the mid-nineteenth century to 1960. He found high inequality in the beginning of the period, and attributes the fall in inequality over time to increased economic liberalization, improved edu-

[1] A further description of the demographic and economic development in Norway in the nineteenth century is given in the Appendix, section B.1.

cation, unionization and reduction in seasonal unemployment. Morrisson (2000) discusses the long-run evolution of inequality in Norway (and several other European countries), and largely agrees with Soltow.

In general, however, little detailed inequality data exists from Norway in the nineteenth century. No estimates from before 1875 have been provided, and there is no account of within-occupation or within-region inequality. The gross domestic product has been estimated back to 1865 (Statistics Norway, 1965), and there is some long-run wage data available (Grytten, 2007), but as mentioned above, this is hard to connect to contemporary welfare measures or to other countries in the same time period.

# 3 Constructing an income distribution from contemporary sources

The data used in this paper comes from records collected by Norwegian official agencies. Nine official censuses were conducted in the nineteenth century, but, with the exception of 1801, the census in 1865 was the first to record individual characteristics rather than only aggregate counts of the population. This information is supplemented by data collected (but not always published) by ministries and other official agencies. The late 1860s is the first period with sufficient information to produce an inequality data set with an acceptable spatial resolution at the rural level. Moreover, the unique source of income distribution data used in this paper was a one-off report commissioned in 1868. The unit of observation used here is the 496 municipalities of Norway, which had populations ranging from 311 to 53652 in 1865. Because of limitations in the sources used, the population studied is men aged above 25.

The next paragraphs outline the construction of the inequality and income indices for Norway in two steps. First, using a parliamentary report on incomes from 1868 as well as the 1865 census, the population is grouped into a set of income and occupational cells. Second, within-cell income distributions are constructed using a different set of sources.

## 3.1 First step: Constructing income cells and some median incomes

The first main source allowing for regional decomposition of inequality is the *Tables informing about the voting rights, income and tax status in Norway in the year of 1868* (Norwegian Department of Justice, 1871). At the time, the Norwegian Parliament considered extending the franchise, which was restricted to men with property (including owner-occupier farmers) and a narrow set of occupations. The proposal was to set an income threshold and let all men above that threshold gain the vote. The report was commissioned to assess how many, and what social classes, would gain the vote for different proposals on the income thresholds. The investigation was conducted by asking all municipalities to collect the income data, “by a cooperation of the leaders of the municipality, the tax commission, the holder of the population records, as well as the sheriff in the countryside".[2] For all municipalities, men aged above 25 were grouped into 26 occupations times five income classes, and report how many in each group currently had the vote. Non-franchised men with incomes below 100 Spd were not included.[3] Four of the intervals are narrow, giving little uncertainty about the incomes of those in the interval, while the uppermost interval is open at the top.

An important asset of this data source is that it aims to cover all sources of income for an individual. Occupation-imputed income, frequently used for estimating historical inequality, takes into account neither the dispersion of income within occupation groups nor the extra income earned from subsidiary occupations. In the present case, the documentation of the income tabulations explicitly states that imputed home production on farms is to be included, addressing some of the challenges of income measurement in a society that was only partly monetized.

The second source is the 1865 census of Norway. The aggregate results of the census are reported in Norwegian Department of the Interior (1868), but the analysis in this paper is based on records for individuals. These have been digitized by the University of Tromsø and the Norwegian National Archives. The files made available through the North Atlantic Population Project (MPC et al., 2008) contain, among other things, information on age, sex and occupation for all individuals in Norway in 1865.[4]

The male population above 25 was selected from the census data. Then, the 1210 different occupations in MPC et al. and the 26 occupation groups in Norwegian Department of Justice (1871) were harmonized into 19 occupation groups to obtain the total number of individuals in each occupation and municipality.[5] The number of individuals with incomes of 100 Spd and above described in Norwegian Department of Justice (1871) was then subtracted from this number, resulting in six income groups per occupation and municipality, with the lowest one containing all individuals with incomes below 100 Spd.

This procedure yields a total of 15,791 cells for the 373,517 individuals in Norway in 1865-1868. Table 1 shows the number of people in each occupation class and income group for the country as a whole. The grouping of individuals into cells immediately allows for some analysis of the income distribution. For example, as the majority of people had incomes below 100 Spd,

[2] Norwegian Department of Justice (1871), "Forklaringer", page XXXIII. All citations from Norwegian sources are translated by the author unless otherwise stated.

[3] By the consumer price index of Grytten (2004), 100 speciedaler (Spd) in 1868 is equivalent to 24,116 Norwegian Krone (NOK) in 2015. The speciedaler was replaced by the krone at a rate of Spd 1=NOK 4 in 1875 when Norway entered the Scandinavian Monetary Union. Sources from the late nineteenth century frequently report amounts from before 1875 in NOK using the 1:4 ratio.

[4] The original census for five municipalities, with a total population of 11,929, is now lost. This leaves us with a sample of 491 municipalities, covering 99.3 percent of the Norwegian population at the time.

[5] The structure of occupational information in the census differs from that in the income data. For example, the census data distinguishes between owner-proprietor farmers and those who own land, whereas the income data does not; the income data distinguishes between workers on daily contracts and workers on permanent contracts, whereas the census does not. This is the reason for the reduction to 19 groups, two of which by definition have no individuals in the income source ("Servants" and "Poor"). The full correspondence between the classifications is shown in Table A14. There were some (relatively rare) cases where the number of people in Norwegian Department of Justice (1871) was larger than the census data; in such cases, people were transferred from another occupation according to a set of rules detailed in the appendix.

| Occupation group | Income group | | | | | |
|---|---|---|---|---|---|---|
| | **1**: >250 | **2**: 200-250 | **3**: 150-200 | **4**: 100-150 | **5**: 100 | **6**: <100 |
| Civil servants | 5137 | 666 | 1403 | 1454 | 788 | 1974 |
| Farmers | 11566 | 6477 | 12190 | 18444 | 12410 | 51712 |
| Merchants and shopkeepers | 4302 | 292 | 536 | 511 | 309 | 736 |
| Craftsmen and artisans | 2632 | 742 | 2146 | 5579 | 2741 | 28032 |
| Owners | 202 | 19 | 54 | 88 | 98 | 469 |
| Engineers | 107 | 9 | 16 | 9 | 1 | 30 |
| Clerks | 843 | 280 | 476 | 403 | 180 | 764 |
| Students and graduates | 256 | 32 | 31 | 20 | 4 | 215 |
| Ship owners | 650 | 44 | 63 | 58 | 54 | 126 |
| Fishermen and other seamen | 312 | 347 | 1174 | 4875 | 3056 | 12916 |
| Cottagers | 51 | 58 | 307 | 1708 | 4760 | 55081 |
| Retirees | 187 | 99 | 265 | 722 | 1270 | 22079 |
| Laborers and workers | 557 | 567 | 2437 | 9297 | 5704 | 25407 |
| Coachmen | 75 | 32 | 88 | 207 | 25 | 952 |
| Managers | 166 | 34 | 56 | 54 | 23 | 1950 |
| Nomads | 67 | 12 | 22 | 26 | 33 | 133 |
| Servants | | | | | | 13578 |
| Paupers | | | | | | 6721 |
| Others | 228 | 42 | 102 | 117 | 44 | 9961 |

Table 1: Number of people (total for all cells) per occupation class and income group.

we can conclude that the median income of Norway was below this amount. Furthermore, we see that the median income for public servants was in the 200-250 interval, and for farmers around 100. We can also see the interval of the median incomes for the 491 municipalities for which we have data.

However, our ability to study mean income or inequality based on these intervals is hampered by the uncertainty about income levels and distributions at the top and bottom. For example, an assumption that the poorest group was concentrated on 90 Spd while the richest group on 300 Spd would yield a Gini coefficient of 0.181 and a mean income of 117, while a decrease of the lowest group estimate to 50 Spd and an an increase of the upper group to 1000 Spd would give a Gini coefficient of 0.565 and a mean income of 152, still not accounting for the inequality effect of dispersion within income groups.

Fortunately, several other sources give more information on the incomes within each group, particularly for the bottom and top income cells. The next section shows how this information is used to generate a full income distribution.

### 3.2 Second step: Within-cell distributions, mean incomes and Gini coefficients

Four additional sources are used to interpolate incomes within groups: agricultural information from the census, wage averages for working-class groups, wages for public servants, and aggregate tax receipts by income groups.

### 3.2.1 Interpolations using agricultural wealth

The 1865 census also contains a set of questions about agricultural conditions, specifying the crops planted and animals owned for each farm. The individual records are digitized and kept by the Norwegian National Archives. Information is entered for each individual regarding the number of animals owned (sheep, goats, horses, pigs, cattle and reindeer) and crops planted (barley, oat, wheat, rye, mixed grain, potato and pea). The crop planting data is converted to expected yield using national averages from 1866-1870 as reported in Statistics Norway (1880). Thereafter, the crop yields are valuated at 1865 prices (the year of the census). Only 1875 prices are available for animals. The price growth for crops is used to construct an indicator of price change from 1865 to 1875. This indicator is then used to deflate the 1875 animal prices to 1865 prices. Thereafter, the total value of animals and crops per individual is calculated. This aggregated information is taken as an indicator of each individual's agricultural wealth. For all men aged above 25, the characteristic is then merged onto the main NAPP data set at the individual record level, and used together with the information on occupation and municipality of residence.

The key assumption used on the within-group distributions is that the ranking of individuals with respect to agricultural property is equal to the ranking of individuals by income, within each occupation group and municipality (no assumptions are made on this relationship across occupations or municipalities). 52 percent of the total population has positive agricultural wealth; with the assumed sorting, this means that 5,590 of the 15,791 cells have agricultural information on all individuals.

For Groups 2 to 4, which contain closed income intervals, the highest- and lowest-wealth individuals, with wealth $a_h$ and $a_l$ are assumed to have incomes at the group borders $y_h$ and $y_l$ (for example, for Group 2, these are 250 and 200). This creates a within-cell relationship between income and wealth $y = y_l + \frac{y_h - y_l}{a_h - a_l}(a - a_l)$. For groups for which no agricultural information is available, a uniform distribution is used. All imputations are done within occupations and income groups.

For Group 1, the richest group, the income-wealth relationship from the second-richest group within each occupation and municipality is assumed to also hold for the richest group. This also pins down the mean income of most of these groups; 1,285 out of 2,200 top-income groups have agricultural information. Information on the rest of the individuals in the richest groups is taken from tax data, as detailed below.

For Group 5, all incomes are set at 100 Spd. The reduction of this group to just one amount in the original source likely reflects the common practice of rounding off income values before assessing them for tax purposes. Introducing a small within-group dispersion here would conflict with the source. Moreover, it would not significantly affect the reported inequality.

### 3.2.2 Wages and the working class

The second source of within-group incomes, and the most important one for the poorer groups, is the set of wage levels in the municipalities. Handwritten lists were found in the Norwegian National Archives. These lists formed the basis of a report on wages from 1865-1885 (Statistics Norway, 1888). The published report gives only the regional aggregates of these numbers.

There are wage observations for up to three different occupations in each municipality. These are daily wages (by season, and by whether food is provided) for cottagers and workers, and annual wages for servants (with food provided). The procedure used to convert these into annual incomes is detailed in the appendix. The main issue of contention is the number of work days $\alpha$ used when converting daily wages to annual wages. The literature uses a range from 280 to 313 days (Grytten, 2007; Lindert & Williamson, 2012); this paper will use 300 as the baseline.

These wages are then taken to be mean incomes for the population of the lowest income group (Group 6). As it is unlikely that all individuals in each municipality earned exactly the same wage, a lognormal distribution is imposed on each occupation-municipality cell. The lognormal distribution is frequently used when modeling incomes, is only defined over positive outcomes, and has low density at the extreme lower end of the distribution. For all groups, the distribution is truncated at 100 Spd. The transformation from a theoretical continuous distribution to discrete individual outcomes is described in the appendix. All lognormal distributions have the same theoretical standard deviation $\zeta$, while the mean is taken from data.

Cottager and servant mean income is taken directly from the wage data. Other working-class groups receive the "worker" wage. Individuals in high-skill occupations (a relatively small number in this income category) are given a markup $\xi$ on the low-skill wage, set at 1.2.

Occupations that are mainly based on capital income (Owners, Students, Ship owners and Retirees) are all given a mean income of 90, as these individuals (few in number) are assumed to be among the richer in this poorest group. All individuals in the "pauper" category are given mean incomes of half the lowest municipality wage observation. Finally, the "other" category, for which no information is available save that the income is less than 100 Spd, is given the mean of the three wage observations.

### 3.2.3 Aggregate taxes and the high-income earners

The highest income group (Group 1) is only bounded at the bottom, at 250 Spd. As no common income tax for the entire country was introduced until 1892, there is no information on total income for this category.

However, there is information for most municipalities on the total amount of municipal taxes paid by men aged above 25 in each income group.[6] Within each municipality, a simple linear regression is run on income groups (2)-(4) with tax receipts as the dependent variable and the number of taxpayers times the midpoints of the income intervals on the right-hand side. Where

[6] These are listed in Norwegian Department of Justice (1871), "Tabel IV".

this yields a consistent tax rate (an increasing slope for the tax-income relationship and an $R^2$ value of above .5), this tax rate is then used to back out the total income of the individuals in the municipality with incomes above 250 Spd. For the forty percent of top-income occupation-municipality cells where there is no agricultural information, this income is then used as mean income in a Pareto distribution with a dispersion parameter of $\alpha = 1.7$ (see Appendix).

For municipalities where neither imputation by taxes nor imputation by agriculture could be used, a data set on the wages of high-ranking public servants was utilized (Norwegian Department of Finance, 1871, p. 92-111). Using linear regression across 230 municipalities that had information on both counts, with a dummy variable to capture rural-urban differences, the mean income is predicted and used as mean incomes for the Pareto distribution.

A total of six out of 491 municipalities lacked sufficient information from any of these sources. For these, the mean income of the richest group was set at the average of the other municipalities for the purpose of estimating the national mean income and Gini coefficient. These municipalities are excluded from all cross-municipality regressions in the later sections.

### 3.3 The final data set

The result of these imputations is a data set with imputed incomes for all men above age 25 in Norway in 1868, with each imputed income begin associated with a municipality and an occupation. The imputation takes into account variation within municipalities and variation within occupations. This data set is available as an Online Appendix. Appendix Table A15 shows the number of municipalities that has populations in any given cell, as well as the methods used for within-group interpolation. The within-group interpolation forms the basis for the calculation of means and Gini coefficients within municipalities and for the country as a whole.

The reliability of the data set hinges on the reliability of the original sources. In particular, we are constrained by the municipalities' evaluations of the incomes in 1868, and their inclusion of incomes in kind. When going back to the nineteenth century, there is no obvious way to get external validation of incomes in kind. Hence, the remaining analysis assumes that incomes in kind are correctly evaluated. An alternative would be to accept an income concept that is defined by the interpretation of the data collectors' notion of incomes. Any bias deriving from omitted incomes in kind are likely to underestimate the incomes of farmers and cottagers relative to other groups, and to underestimate rural relative to urban incomes.

## 4 Regional and national estimates of inequality

### 4.1 The national income distribution

Using this combination of tabular, census and wage data to impute distributions across and within municipalities and occupation groups, we obtain the income distribution plotted in Figure 1. The median income of the country as a whole is 91 Spd, with a 10th percentile of 59 and

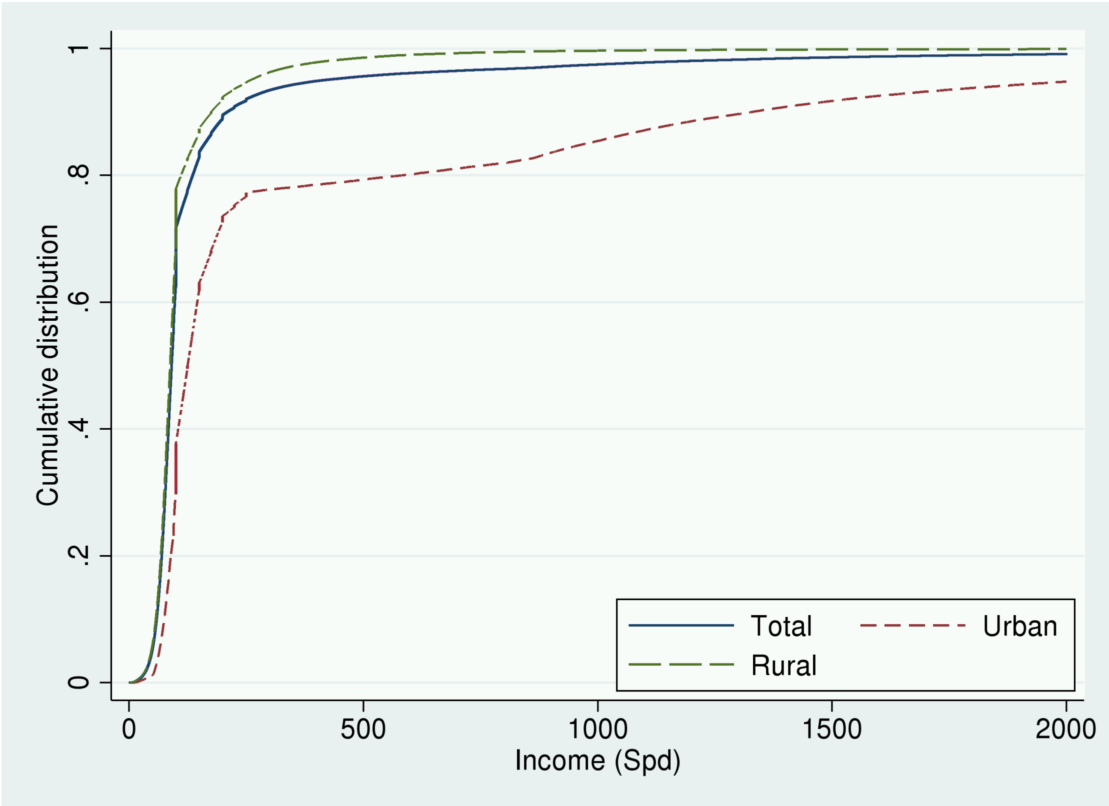


Figure 1: Income cumulative density plot, men aged 25 or above. National, rural and urban. Plot is truncated at 2,000 Spd.

a 90th percentile of 210. For the rural areas the numbers are slightly lower, at 87, 57 and 175, respectively, while the urban areas have much higher incomes, with median 125, 10th percentile 75 and 90th percentile 1,330 Spd. Key statistics for the country as a whole as well as for geographical regions and occupation groups are given in Table 2.[7]

| | Gini coeff. | Mean inc. | Pop. | Rural Gini | Urban Gini | Urb./Rur. mean inc. | Urb. pop. share |
|---|---|---|---|---|---|---|---|
| Total for Norway 1868 | 0.546 | 179 | 373,517 | 0.346 | 0.715 | 4.5 | 15% |
| By region (*Stift*): | | | | | | | |
| Christiania (East) | 0.643 | 246 | 96,375 | 0.418 | 0.725 | 4.3 | 26% |
| Hamar (Central inland) | 0.349 | 105 | 53,416 | 0.331 | 0.576 | 2.8 | 2% |
| Christiansand (South) | 0.537 | 189 | 71,682 | 0.277 | 0.707 | 4.5 | 20% |
| Bergen (West) | 0.487 | 165 | 58,577 | 0.306 | 0.696 | 4.2 | 13% |
| Throndhjem (Central coast) | 0.549 | 166 | 58,991 | 0.388 | 0.737 | 4.7 | 11% |
| Tromsø (North) | 0.338 | 130 | 34,476 | 0.269 | 0.563 | 3.0 | 7% |
| By occupation group: | | | | | | | |
| Civil servants | 0.665 | 719 | 11,422 | 0.500 | 0.593 | 5.0 | 44% |
| Farmers | 0.369 | 150 | 112,799 | 0.366 | 0.672 | 2.1 | 0.4% |
| Merchants and shopkeepers | 0.622 | 1133 | 6,686 | 0.498 | 0.560 | 5.1 | 68% |
| Craftsmen and artisans | 0.568 | 193 | 41,872 | 0.172 | 0.671 | 3.4 | 47% |
| Owners | 0.668 | 346 | 930 | 0.537 | 0.708 | 2.3 | 48% |
| Engineers | 0.602 | 1102 | 172 | 0.539 | 0.553 | 3.2 | 62% |
| Clerks | 0.692 | 617 | 2,946 | 0.434 | 0.670 | 4.9 | 65% |
| Students and graduates | 0.673 | 851 | 558 | 0.560 | 0.619 | 4.2 | 57% |
| Shipowners | 0.596 | 691 | 995 | 0.469 | 0.557 | 3.5 | 33% |
| Officers, merchant marine | 0.710 | 724 | 7,143 | 0.669 | 0.676 | 2.5 | 55% |
| Fishermen and other seamen | 0.229 | 112 | 22,680 | 0.175 | 0.305 | 1.5 | 25% |
| Cottagers | 0.164 | 75 | 61,965 | 0.164 | 0.125 | 1.2 | 0.2% |
| Retirees | 0.152 | 89 | 24,622 | 0.152 | 0.161 | 1.0 | 0.4% |
| Laborers and workers | 0.264 | 112 | 43,969 | 0.174 | 0.366 | 1.8 | 23% |
| Coachmen | 0.402 | 146 | 1,379 | 0.161 | 0.490 | 2.1 | 51% |
| Managers | 0.490 | 174 | 2,283 | 0.355 | 0.647 | 2.1 | 28% |
| Nomads | 0.389 | 168 | 293 | 0.390 | . | . | 2% |
| Servants | 0.156 | 67 | 13,578 | 0.159 | 0.105 | 1.1 | 8% |
| Paupers | 0.295 | 35 | 6,731 | 0.299 | 0.254 | 1.1 | 10% |
| Others / Occupation unknown | 0.417 | 112 | 10,494 | 0.179 | 0.735 | 5.6 | 10% |

Table 2: Income inequality in Norway, 1868, for men aged above 25, by region and occupation.

The first line of the table shows mean income and Gini coefficient for the country as a whole. Mean income for men aged above 25 in Norway in 1868 is 179 Spd, and the Gini coefficient is 0.546. The following columns show the urban and rural Gini coefficients separately, as well as the ratio of urban to rural mean incomes. The difference between the cities and the countryside

[7]For Theil and Atkinson indices, see Table A3. The data also allows for calculation of top income shares; the top 0.1 per cent obtain 9 per cent of total incomes; the top 1 per cent obtain 25 per cent, and the top 10 per cent obtain 54 per cent. These top income shares are, however, less reliable than the countrywide Gini coefficients as imputation is used at the top; for this reason, estimates for years for which detailed tax data is available (such as 1875) should be preferred to these estimates. This is discussed in further detail in the Appendix, Section F.2.

is striking: urban mean income is more than four times that in the rural areas; inequality in the cities, as measured by the Gini coefficient, is more than twice that in the countryside.[8]

## 4.2 Inequality across regions and occupations

The second panel of Table 2 shows mean income and inequality in the six main regions of Norway, as reflected in the dioceses (*Stift*) as they existed in the 1860s. As is the case today, the income in the capital region (then called Christiania) is highest, with a mean of 246 Spd. It is followed by the diocese of Christiansand in the south with 189 Spd, with Bergen (west) and Trondhjem (central) both having a mean income of around 165 Spd. In all these four dioceses the mean urban income is more than four times the mean rural income, and at least 10% of the population live in cities. The two remaining dioceses, Hamar in the central inland region and Tromsø in the far north, have lower mean incomes, smaller urban populations and smaller rural/urban income differences.

The diocese Gini coefficients follow the same ordering as the mean incomes, as regions with high mean incomes also have higher inequality. The exception is Trondhjem, which ranks third in income but second in terms of income inequality. Income inequality within the rural areas is moderate in all six regions. Here Christiansand is an exception to the ordering, with low rural inequality despite a high overall mean income. Gini coefficients within the urban areas broadly follow the ranking of the overall Gini coefficients; these incorporate both between- and within-city differences within each diocese. We return to individual municipal and city Ginis below.

One might be puzzled that the geographical differences reported in Table 2 do not reflect the conventional view that western Norway did in general have lower inequality than eastern Norway. However, these coefficients incorporate both differences between areas as well as differences within municipalities. If instead we take the average of rural municipal Gini coefficients within the dioceses, we obtain high values in the east (Christiania: 0.347, Hamar: 0.307), intermediate for Trondhjem (0.286) and lower values in the other coastal districts (Christiansand: 0.241, Bergen: 0.264 and Tromsø 0.249).

Mean incomes and inequality for each occupation group is reported in the third panel of Table 2. Traditionally, the uppermost positions in the occupation hierarchy were those held by the civil servants. The 11,422 individuals here occupy both elite and more modestly-paid civil servant positions; overall, they have a mean income of 719 Spd, with the mean urban income being five times the mean rural income. A few occupation groups with few members have higher mean income, including merchants, engineers and merchant marine officers. Farmers, who constitute around 30% of the population by the definitions used here, have a mean income of only 150 Spd. It is possible that farmer incomes are slightly under-reported. However, it is also the case that individuals with higher status would often hold an additional occupation and choose to report

[8] Rural and urban Gini here refers to the Gini coefficient of the entire rural and urban subpopulations rather than to the average city or urban municipality Gini; municipality averages are reported in the next section.

this instead of the farmer occupation. For this reason, some well-off farmers are reported in other occupation categories here. The remaining large groups are craftsmen and artisans, with a mean income of 193 Spd., fishermen/seamen and laborers with a mean income of 112 Spd., retirees (presumably at farms) at 89 Spd and cottagers at 75 Spd.

In most of these occupations there is substantial income variation. For example, the Gini coefficient among civil servants is 0.665. There is lower dispersion in moderate-income occupations. The farmer Gini is 0.369, the laborer Gini 0.264 and the cottager Gini only 0.164. This depression can be partly understood in terms of a subsistence income floor, where low mean incomes simply do not leave room for substantial variation.

The urban-rural income gap is also highest for the highest-income occupations. However, even among laborers, urban incomes are twice as high as rural incomes.

## 4.3 Inequality decomposition and the rural/urban gap

Using the underlying data we can decompose the country Gini coefficient into the conventional between- and within-group components (see, for example, Lambert & Aronson (1993)) as

$$G = G^B + a_r G_r^W + a_u G_u^W + R \tag{1}$$

where $a_i$ is the group weight (product of group income and group size), $G_i^W$ is the within-group Gini, $G^B$ is the between-group Gini and $R$ is an overlap term. We obtain the decomposition

$$0.546 = \underbrace{0.296}_{\text{Between-group}} + \underbrace{(0.847 \cdot 0.551) \cdot 0.346 + (0.153 \cdot 0.449) \cdot 0.715}_{\text{Within-group}} + \underbrace{0.040}_{\text{Residual/overlap}} \tag{2}$$

The between-group component 0.296 is the inequality that would be obtained if all individuals in cities earned the urban mean income while all individuals in the countryside earned the rural mean income. This is slightly more than half of total inequality. The scaled within-group Gini coefficients total 0.21, or most of the remaining pairwise income differences. As there is relatively little overlap between rural and urban income distributions, the residual (overlap) term is small at only 0.04.

Kuznets (1955) discusses both the role of differential income growth across sectors (rural and urban) and the role of dispersion within sectors for aggregate income inequality. We can consider how such changes would influence inequality as measured here. First, let us double the observed urbanization rate of 15.3%. We then obtain an urban population share of 30.6%, while keeping the rural and urban income distributions constant. Such an increase in the urban population, with its higher average incomes and higher dispersion, increases the Gini coefficient from 0.546 to 0.633. Second, we can eliminate the rural/urban income gap while preserving the within-sector income dispersion. This lowers the nationwide Gini to 0.429. Third, the distribution of the rural and urban sector can be equalized, maintaining the income gap between rural and urban areas. If this same distribution was the rural, low-inequality distribution, the overall Gini coefficient would be reduced to 0.498. On the other hand, if rural areas had an urban dispersion, we would have a Gini coefficient of 0.763, both through an increase in the rural within-group component (larger difference between high-income and low-income rural individuals) and an incrase in the overlap component (larger incrase between high-income rural individuals and low-income urban individuals).

## 4.4 The role of aggregation level

There are few historical cases where a large set of sources is available to estimate income inequality. For that reason, it is of interest to see how estimated Gini coefficients depend on the level of aggregation available in sources. The estimated Gini of 0.546 is based on imputed incomes for all men above 25 in Norway in 1868. What if, instead, we only based our calculation on tabulations of aggregated information on these incomes?

| Regional aggregation | Occupational differentiation | Gini | Number of cells |
|---|---|---|---|
| Entire country | No differentiation | 0.000 | 1 |
| Urban/Rural | No differentiation | 0.296 | 2 |
| Region | No differentiation | 0.147 | 6 |
| Region + Urban/Rural | No differentiation | 0.320 | 12 |
| Municipality | No differentiation | 0.384 | 491 |
| Entire country | 20 occupation groups | 0.395 | 20 |
| Urban/Rural | 20 occupation groups | 0.433 | 40 |
| Region | 20 occupation groups | 0.414 | 119 |
| Region + Urban/Rural | 20 occupation groups | 0.445 | 228 |
| Municipality | 20 occupation groups | 0.477 | 6746 |
| Imputed individual incomes | | 0.546 | 373,517 |

Table 3: Calculated Gini coefficients for different levels of aggregation of information. In each line, the Gini coefficient is calculated (using appropriate weights) based on a grouping of the population into a set of regions and/or occupations, and calculating mean incomes for these groups.

The results of such calculations are reported in Table 3. In each case, incomes are reduced to a limited set of cells with mean incomes of subgroups based on geographical region and occupation. Trivially, if the country was collapsed to just one cell with one mean income the Gini coefficient would be zero (the first line).

In the first panel, occupational income differences are disregarded and we calculate income inequality as if we only had mean incomes and population sizes for each geographical region. In the case of only two regions, namely urban and rural regions, the Gini coefficient would be 0.296 (the between-group Gini coefficient in Equation (2)). Regional differences are less salient than urban-rural differences, so a table with mean incomes and population sizes for each of six regions would give a Gini coefficient of only 0.147. Increasing the regional variation by combining region and urban/rural status and finally using one cell for each of the 491 municipalities further increases the calculated Gini coefficient, but only to a maximum of 0.384 if we do not include occupation information.

The first row in the second panel of Table 3 has 20 cells, one for each occupation. We then proceed to add regional disaggregation while maintaining occupational average incomes within each region. There is still a relationship between regional resolution and calculated Gini coefficient, but the dependence on regions is lower when occupational mean incomes are used. Using one cell per municipality-occupation combination gives 6,746 cells and a Gini coefficient of 0.477.

The difference between the highest regional-variation Gini of 0.384 and the "best estimate" of 0.546 is similar to the present-day gap between Norway and Argentina.[9] Surprisingly, the highest Gini using only between-occupation variation (at 0.395) is not much higher than the between-region Gini, giving a similar difference. Even when we include both occupation and regional variation, we end up at an estimate of seven Gini points lower than the best estimate. This highlights the importance of using actual income tabulations to estimate historical income inequality, and to carefully consider differences between sources when comparing across countries. We return to this in Section 6.

# 5 What determines income inequality at the local level?

While the previous section considered income inequality in Norway as a whole, an advantage of the data is that one can also study income inequality at the local municipality level. This allows for a test of whether common hypotheses of historical income growth and inequality also hold in the cross-section within a single country. The advantage of using cross-sectional variation in inequality is that the legal and cultural environment will, to some extent, be held constant across municipalities, while economic conditions vary. We will focus here on the relationship between rural and urban areas and on the structure of the rural economy.

From the data set presented in the previous sections, we get the Gini coefficient and mean income of men aged above 25. The average city Gini coefficient is 0.580 and the average rural municipality Gini coefficient is 0.280. The municipal Gini coefficients are shown on a map in Figure 2. A visual inspection clearly shows that more municipalities with higher inequality are located in eastern Norway, though there are also some pockets of high inequality in Trøndelag, in the western interior and in the far north. High inequality within cities is prevalent across the country.

The following analysis will proceed in three steps. First, the relationship between inequality and economic development in a traditional sense will be examined. Then we more to the role of food production. Finally, these explanations are tested in a joint regression framework.

[9]The 2012 Gini coefficients of Norway and Argentina, respectively, were .259 and .527, according to the World Bank Development Indicators (January 2017).

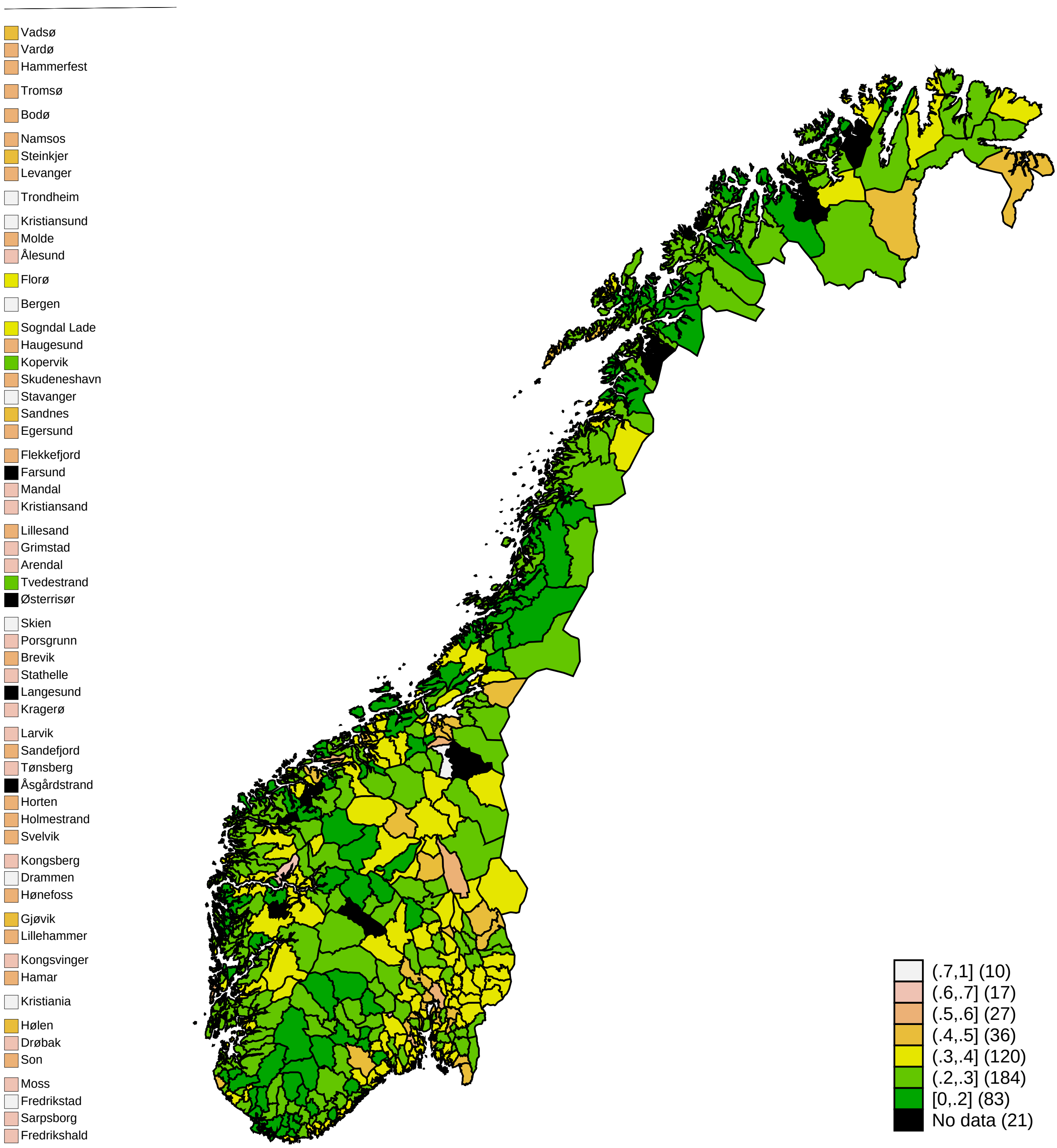


Figure 2: Within-municipality income Gini coefficients for men aged 25 or older in Norway, 1868. City municipalities are not visible in the map; refer to boxes at left (shown from north to south)

## 5.1 Sources of municipality characteristics

Information on the characteristics of municipalities are obtained from various official sources. Information about industrialization in the municipalities in 1870 is taken from a report by the Norwegian Ministry of the Interior (1876). The source lists the number of workers in manufacturing ("industri") in all municipalities, as well as the composition across various industries and the type of fuel used. Many municipalities do not have any manufacturing sector at all, and many of the industries are concentrated in only a few municipalities. For this reason, only data on the total number of industrial employees (relative to the population of the municipality) will be used.

From data on administrative borders we can obtain some geographical characteristics of the municipalities. A key geographical variable of interest is the distance from a given municipality to the nearest city. This is calculated as the shortest direct line from the geographical midpoint of a municipality to the nearest midpoint of a city municipality. The data was obtained from a map (shapefile) of Norwegian municipalities in 1868 obtained from the Norwegian Social Science Data Services (NSD). Mean "remoteness" for the rural municipalities—the distance between the geographic midpoint of the municipality and the closest city—is 47 km, while the most remote municipality is 177 km from the nearest city.

Data on food production modes is obtained directly from the 1865 census and its associated census of agriculture. The calculation of crop and animal values are documented in Section 3; as values of one goat and one sheep are similar, these are simply added together. There is no separate census of fishing; the share of fishermen is calculated as the share of individuals stating "fisherman" as one of their occupations, as recorded in the original census forms.

Descriptive statistics for all municipalities and subsamples are given in the Appendix, Table A4. For twelve of the municipalities, national averages were used to construct within-group inequality. These are removed from this sample, giving a total of 479 observations.

Pairwise correlations between income inequality and the covariates, with associated 95% confidence intervals are shown in Figure 3. The sign after the variable name denotes whether the variable of interest is positively or negatively correlated with income inequality.[10]

## 5.2 Income inequality and economic development

Kuznets (1955) hypothesized that income inequality first increased, then decreased, with economic development. Kuznets proposed that inequality began to increase around around 1850, followed by a reversal in the early twentieth century. This development was linked to the movement of individuals from an agricultural sector with low mean income and low inequality to a modern sector with high inequality and high mean income.

Moreover, Kuznets acknowledged the direct, mechanical link between development and income inequality: societies with higher mean incomes can sustain higher inequality, as the distance

[10] This information is also presented in Appendix Tables A5- A6.

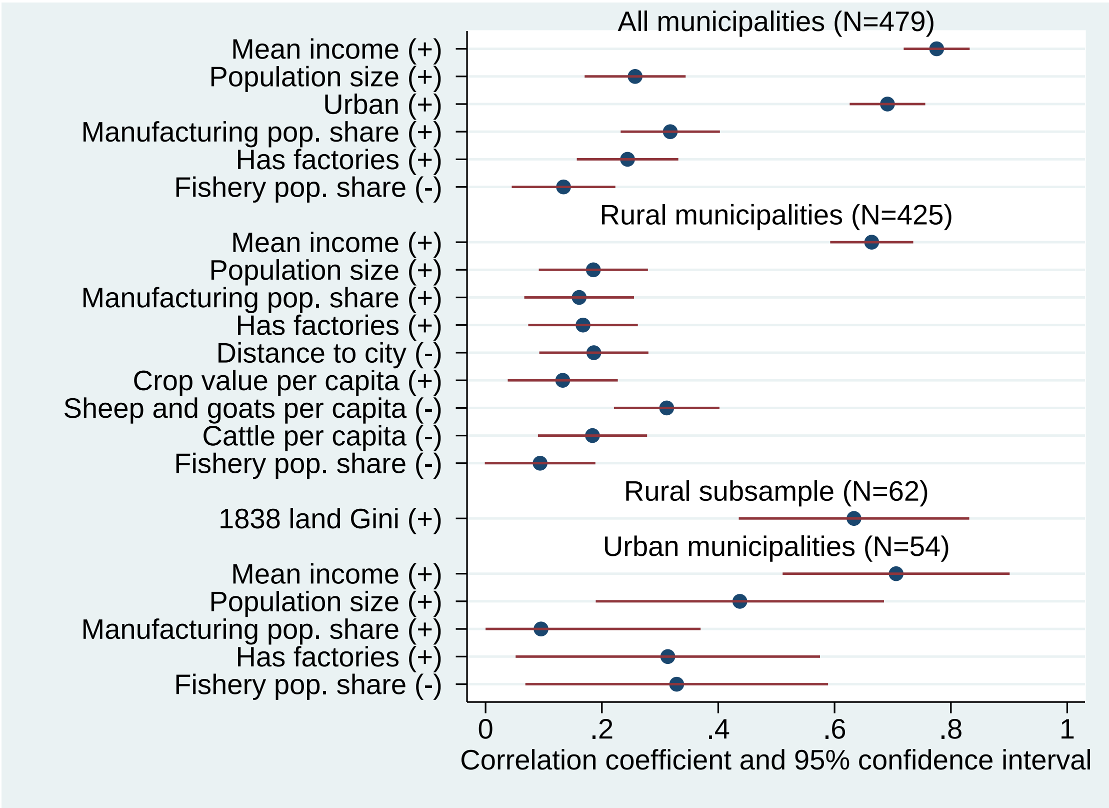


Figure 3: Correlation coefficients between within-municipality income inequality and other characteristics of municipalities. 95% confidence intervals shown.

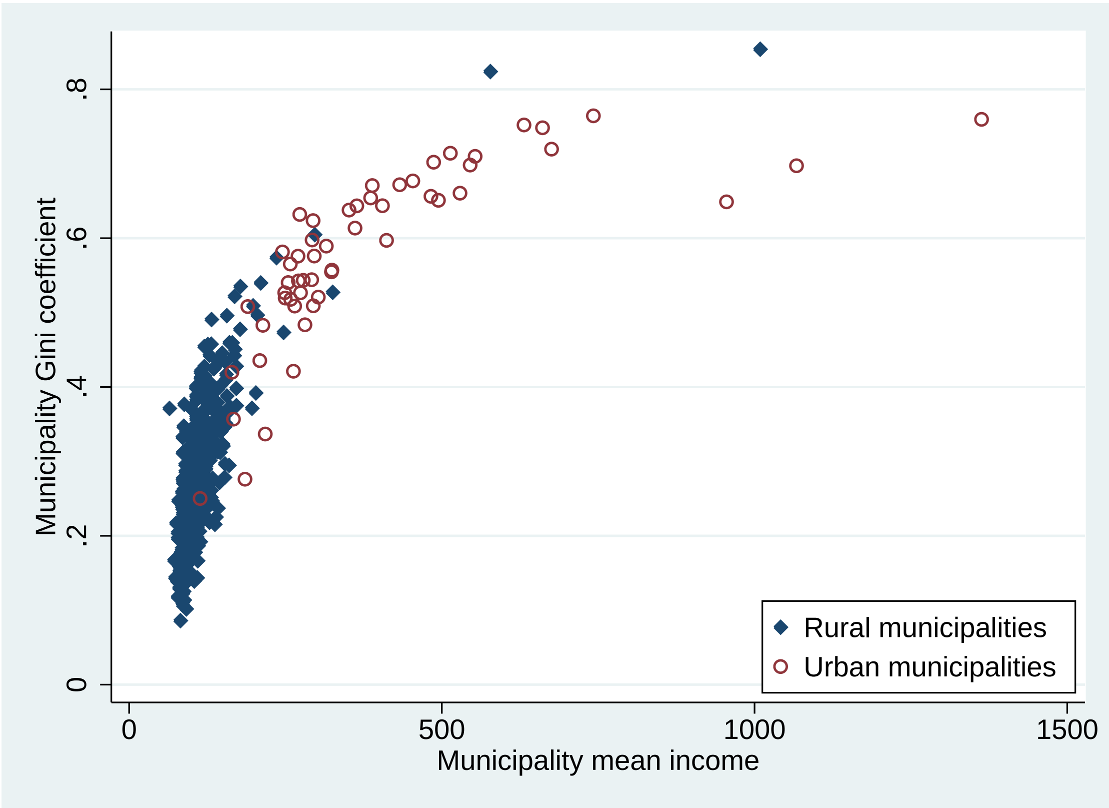


Figure 4: Municipality Gini coefficients and mean incomes

widens between the incomes of the upper classes and the absolute income level needed for the sustenance of the lower classes. This idea was further formalized by Milanovic *et al.* (2011), who document the positive relationship with a cross-country data set of countries covering a large time period. Van Zanden (1995) find a similar trend for inequality within cities (and to some extent within rural areas) for Europe before 1800. A cross-section plot of income inequality against mean income for all municipalities is provided in Figure 4, and also confirms the concave curve in the present case. The correlation between income level and income inequality is $\rho = 0.78$, and is shown with associated confidence interval in Figure 3. Replacing income with a linear transformation (such as log income) gives a similar value for the correlation.

In the country as a whole, the correlation between the population share in manufacturing and income inequality is positive ($\rho = 0.35$), and the same relationship holds if we look only at rural municipalities ($\rho = 0.18$). The positive relationship remains if we compare only municipalities that have factories with those that do not ($\rho = 0.39$) or if we compare only the share of workers in municipalities with manufacturing ($\rho = 0.32$, not shown). Within urban areas the correlation between share of workers and income inequality is lower, and the confidence interval overlaps with zero; however, the extensive margin (whether there is manufacturing or not) is still statistically significant.

The results indicate a positive relationship between income inequality and a measure of modernity and economic growth, namely manufacturing, that also holds within rural areas. However, manufacturing is not an exclusively urban phenomenon, and there are many other ways in which rural areas differ from cities. Figure 3 shows a positive correlation between urban status and income inequality. We can examine the rural-urban aspect of Kuznets' theories in more detail by differentiating between municipalities that are close to cities and municipalities in more remote areas. This captures differences such as access to markets for agricultural products. As shown in Figure 3, distance from city is negatively correlated with income inequality ($\rho = -0.22$); inequality is lower further away from cities.

These relationships hint at a possible way of disentangling the Kuznetsian hypothesis of an U-shaped relationship between inequality and development from a more mechanical view of the "inequality possibility frontier" proposed by Milanovic *et al.* (2011). In the Kuznetsian framework, one would expect some high-income municipalities to have lower income inequality than intermediate-income municipalities. Based on Figure 4, this does not appear to be the case, though this could result from the general development not having reached the level where inequality starts to decrease.

The share of the population in manufacturing, and the proximity to urban areas, can both be seen as indicators of economic development that capture different aspect than mean income does. Hence, the correlations found for these variables suggests support for the Kuznets hypothesis. It is interesting to test whether the associations remain when we also control for income levels. Before we test this more formally, however, we turn to the relationships between inequality and the variables indicating modes of agriculture and food production.

## 5.3 Income inequality and food production

There are several ways in which the predominant way food is produced can affect the distribution of income in a society. For the purpose of this analysis, it is convenient to distinguish three distinctly different ways of producing food. First, planting and harvesting of crops. This can have varying capital intensity, but needs land, which is usually unequally distributed. Second, animal husbandry. A large number of animals per capita can be an indicator either of high wealth (capital density) or of marginal land requiring animals to be fully utilized. Cattle production was frequently market-oriented. Butter and cheese could be transported over long distances, and in the 1860s improved communications also made the sale of milk more feasible, at least close to the cities (Try, 1979, p. 238). In summer months, animals grazed on common pastures, and sheep and goats in particular were able to graze on land that was otherwise unsuitable for agriculture.

Third, and radically different, is fishing. While some capital (at the very least, a share in a boat) is needed for offshore fishing, we would expect a lower persistence of inequality because of the non-use of land in production. Ocean fishing was open to all, and was an important source of food all along the coast.

Correlations between modes of food production and income inequality are shown in Figure 3 for the rural municipalities. Municipalities with higher average crop values exhibit higher inequality. Conversely, municipalities oriented towards (pasture-intensive) sheep and goats are associated with lower inequality. Cattle production is also associated with lower inequality, though with a lower correlation coefficient. Similarly, municipalities where fishermen constitute a large share of the population have on average lower income inequality.

## 5.4 Comparing two theories: Development versus agriculture

The correlations in Figure 3 show that several economic factors are associated with inequality. However, these factors are likely to interact and some correlations could simply reflect associations in the underlying data. For example, municipalities with high crop values are likely to exhibit higher incomes on average, and this could be the channel of influence. For this reason, we now move to a joint test of the various factors that are associated with inequality. Based on the municipal covariates, we can use a simple regression framework to examine in more detail the drivers of inequality.

Using ordinary least squares, we estimate the following regression, where $i$ indexes municipality, $G_i$ is the municipality Gini coefficient and $X_i, Y_i, ...$ are covariates at the municipal level:

$$G_i = \alpha + \beta X_i + \gamma Y_i + ... + \epsilon_i \tag{3}$$

Coefficient estimates for $\beta, \gamma...$ for rural municipalities are given in Table 4. The first column confirms the positive association between income levels per capita and income inequality; an increase in the income level of one per cent on average corresponds to an increase in the Gini coefficient of 0.30 percentage points. We maintain the control for this association throughout the analysis, but note that the results remain similar if this variable is removed from the analysis (see Appendix, table A9).

| | (1) Gini | (2) Gini | (3) Gini | (4) Gini | (5) Gini | (6) Gini |
|---|---|---|---|---|---|---|
| Log income | 29.64*** | 31.83*** | 31.02*** | 31.05*** | 29.33*** | 31.05*** |
| | (25.83) | (30.75) | (29.23) | (29.51) | (25.48) | (29.41) |
| Crop value pc | | 0.0446*** | 0.0403*** | 0.0350*** | | 0.0350*** |
| | | (10.87) | (9.31) | (7.49) | | (7.48) |
| Sheep/goats pc | | | -0.236*** | -0.226*** | | -0.226*** |
| | | | (-2.72) | (-2.63) | | (-2.59) |
| Cattle pc | | | 0.209 | 0.0614 | | 0.0611 |
| | | | (0.66) | (0.19) | | (0.19) |
| Fishery pop. share | | | | -5.018*** | | -5.020*** |
| | | | | (-2.88) | | (-2.85) |
| Manufacturing pop. share | | | | | 31.75** | -0.0851 |
| | | | | | (2.18) | (-0.01) |
| Constant | -111.7*** | -127.9*** | -122.6*** | -121.3*** | -110.5*** | -121.3*** |
| | (-20.62) | (-25.49) | (-22.99) | (-22.86) | (-20.40) | (-22.83) |
| Observations | 424 | 424 | 424 | 424 | 424 | 424 |
| $R^2$ | 0.612 | 0.697 | 0.704 | 0.710 | 0.617 | 0.710 |

*t* statistics in parentheses

* p<0.1, ** p<0.05, *** p<0.01

Table 4: Results from OLS regressions at the municipal level; dependent variable is the municipality income Gini coefficient (×100). Rural municipalities only

In the model shown in the second column, a variable for the average crop value per capita in the municipality is added. The positive coefficient estimate indicates that more affluent agricultural areas have higher income inequality, even when the average income level of the municipality is taken into account. The third column supplements this analysis by also accounting for the number of sheep and goats per capita, as well as cattle per capita. A higher number of sheep and goats, representative of more pastoral agriculture on lower-yield land, are negatively associated with income inequality. When these agricultural controls are included, there is no statistically significant coefficient on the number of cattle.

The joint associations of income levels, crop values and pastoral/extensive agriculture with income inequality do show a substantial influence of agricultural structure. The fourth column confirms this picture by also including a control for the share of the population involved in fisheries, which is negatively associated with income inequality, without substantially altering the other coefficient estimates.

The fifth column of Table 4 shows an alternative model, where the share of the population involved in manufacturing (together with the average income level) represents economic development in a tradititional sense. There is a positive association with income inequality if the agricultural characteristics are not considered. However, when these explanations are put together in column 6, the coefficient on manufacturing is close to zero (and not statistically significant), while there is little change in the agricultural variables.[11]

These results suggest that factors influencing agricultural modes of production (as in Engerman & Sokoloff (2002)) are more important drivers of inequality than Kuznetsian mechanisms related to modernizing factors. This becomes even more evident if we include the control for distance to the nearest city (shown in the Appendix, table A8). The coefficient on city distance decreases substantially when agricultural variables are included and is not significant in model 4 when the share of population in fisheries is controlled for.

To the knowledge of this author, the competing explanations of industrialization and agricultural structure have not previously been analyzed in a joint framework in a within-country setting from the nineteenth century. The results, while not disproving the Kuznets hypothesis as such, suggest that at in some settings, agricultural production modes, presumed to derive from geographical and climate differences, can be as important for societal structure as the degree of industrialization.

### 5.5 Relationship to historical land inequality

Persistence in inequality is frequently put forward as an explanation for present-day differences between countries. While there is no available source on the local dispersion of income in Norway before 1868, it is possible to obtain the dispersion of farm values from an earlier source for a subset of the municipalities.

The use of farm value registries for tax purposes have a long history in Norway. Land registries go back to at least the sixteenth century, and were occasionally updated through the seventeenth and eighteenth centuries.[12] A major revision of the land register was begun in 1818 and completed in 1838. In every *tinglag*, roughly equivalent to a municipality, a local farm was chosen as a reference farm and thoroughly examined, and other farms were then compared to the reference farm. The old system of stating tax obligations in kind was replaced with a one-dimensional system using a monetary value. The register was later criticized for its comparison between municipalities and replaced with a new calculation in 1886.

The 1838 register has been digitized (at farm level) for four counties: Hedmark, Buskerud, Telemark and Troms, and is available at the website of the University of Tromsø.[13] Adjusting municipalities to conform to our 1868 sample results in the loss of some areas, but for a total of 62 municipalities the reported farm values can be collapsed to municipality Gini coefficients. Because of the controversies surrounding the between-municipality comparisons, the mean level of tax obligations will not be used as an explanatory variable here.

[11]The results are qualitatively similar if we include the urban municipalities in the regressions (see Table A7).

[12]For a brief review of Norwegian farm registries, on which this description is based, see "Matrikkel" in the Norwegian Historical Encyclopedia (in Norwegian); also available at http://www.rhd.uit.no/matrikkel/hl.html.

[13]Download link: http://www.rhd.uit.no/matrikkel/excel.html

The right-hand column of Table A4 summarizes the municipalities for which the 1838 land Gini is available. They are slightly more populous than the average rural municipalities, have lower mean income, lower inequality, and are on average further away from the nearest city. The land Gini ranges from 29 to 65, with 48 as the mean value.

The 1838 land Gini is strongly correlated with the 1868 income Gini ($\rho = 0.63$, se=0.10). This shows that persistence in inequality in nineteenth-century Norway was high. The correlation coefficient remains positive when the exercise is repeated within each of the four counties where 1838 data is available, though in one of the counties (Telemark) the confidence interval overlaps zero. We can interpret this as a strong influence from the underlying agricultural structure—present for generations—to the income inequality observed in 1868.

The regression analysis from the previous section can be repeated with a control for historical land inequality. For the municipalities where historical land inequality is available, results are similar to those shown in the previous section (see Appendix Table A10 for a replication of Table 4 on the subsample). When the model is altered to include historical land Gini as an explanatory variable, the estimated coefficient on pastoralism (sheep and goats) as well as on fisheries become weaker (see Table A11). This is not surprising, as there are several joint determinants of agricultural mode and land inequality, as well as direct connections. It does show, however, that land inequality is likely one of the channels through which the agricultural modes of production influences income inequality, and one that persists for a long time.

## 5.6 Occupation-specific inequality dispersion

The analysis above has only considered income inequality among all individuals in a given municipalities. We can further examine the relationship between the underlying factors and 1868 income inequality by utilizing the information on inequality within occupation groups. To do this, we calculate the within-group Gini coefficient of each occupation group in each municipality. We then replace the municipality Gini coefficient in the correlations with this within-occupation Gini, and re-examine the correlation coefficients.

The results are shown in Figure 5 for all cases where there were at least 40 municipalities with at least 10 people in the relevant occupation. Each panel of the figure shows the correlations between one of the municipal characteristics and municipal within-occupation income inequality for each of the 13 occupation categories where there is sufficient data. For comparison, correlations with the between-occupation Gini as well as the overall municipal income Gini coefficient (used in the previous section) are also presented. The vertical line in each panel corresponds to the correlation with the overall Gini coefficient.

In general, the correlations are of similar magnitude and have the same signs when we consider inequality within occupations. First, correlations with the manufacturing population share are higher for income inequality among laborers and workers than for the population in general, while they are lower for inequality among servants or among civil servants. Second, inequality decreases less, relatively speaking, with distance to city for civil servants, merchants and servants

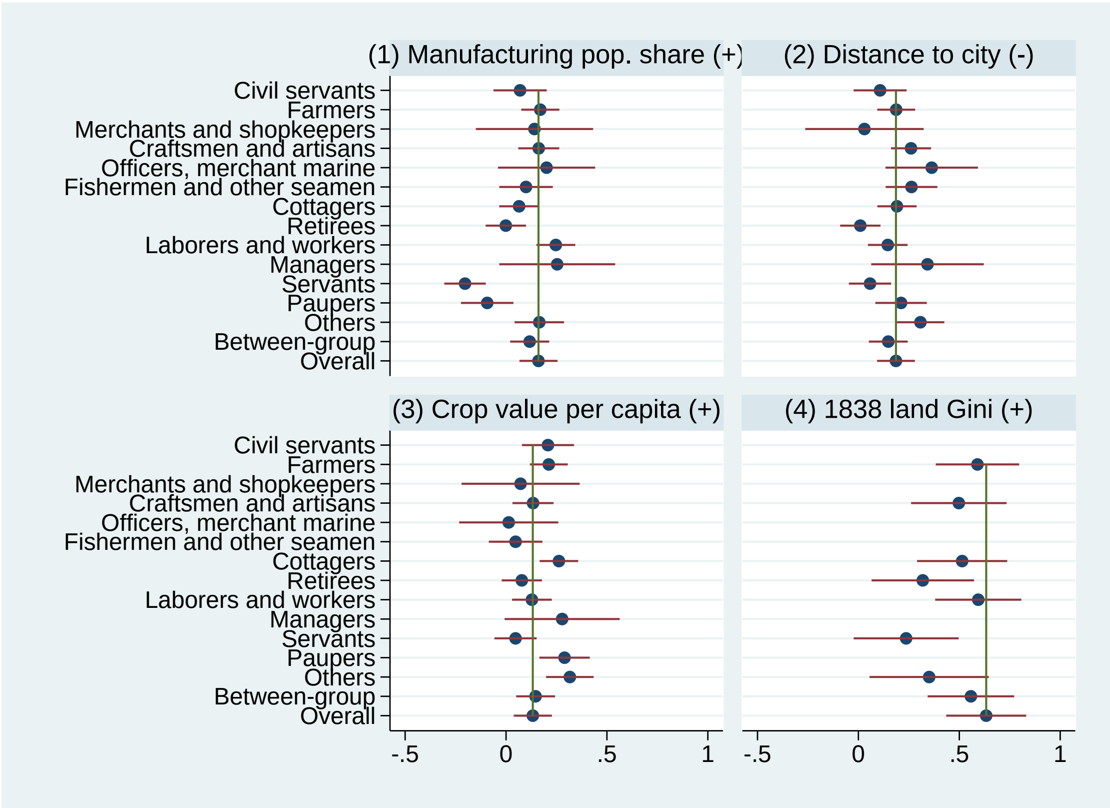


Figure 5: Correlation coefficients between within-municipality income inequality in a given occupation group and other characteristics of municipalities. 95% confidence intervals shown.

than for other occupation groups. Higher average crop values are more strongly associated with income inequality among farmers, cottagers and managers. Finally, we have fewer observations for the smaller sample of the 1838 land Gini, but see stronger associations for laborers and farmers than for other occupation groups.

Shown below the within-occupation inequality correlations in each panel is the correlation of between-occupation inequality (the income inequality that would prevail within the municipality if everyone received the mean income of their occupation) and the relevant variable. In all four cases, these correlations are similar to the correlations with income inequality in general.

A replication of the regression analysis from Section 5.4 on within-occupation inequality is shown in the Appendix (Table A12). The relationship between inequality and mean income, crop value, pastoral agriculture and fisheries hold up for most occupation groups, but there are some interesting exceptions. There is no higher within-servant income inequality in high-income municipalities, and within-servant income inequality is also higher in municipalities with pastoral agriculture. Pastoral agriculture is also associated with higher inequality among fishermen.

There is no strong association between within-occupation income inequality and manufacturing activity, except for inequality among laborers and workers. A one percentage point increase in the share of population in manufacturing is associated with a Gini coefficient that is 0.5 percentage points higher. This suggests a role for the Kuznets mechanisms in the specific case of worker inequality. The worker category includes both agricultural and industrial workers. Thus a higher dispersion in manufacturing municipalities is consistent with Kuznets' sectoral shift hypothesis where a subset of the population obtains higher incomes while a smaller share remains in lower-paid occupations.

# 6 Discussion and conclusion

## 6.1 Was income inequality in 1868 Norway high?

This paper has shown substantial variation in income both within and between occupation groups and regions of Norway in 1868. As stated in Section 4, these income differences add up to a Gini coefficient of 0.54. How does this compare to other countries?

Three relevant reconstructions of income distributions are comparable to those found here. First, Lindert (2000) gives an income Gini of Great Britain in 1867 of 0.49. Second, Lindert & Williamson (2012) give a US Gini coefficient in 1860 of 0.51. Third, Nafziger & Lindert (2012) calculate a Gini coefficient for Russia in 1904 at 0.36. These studies all define inequality at the household level rather than the individual level.

Today, Norwegian income inequality is well below that of all these three countries. We know from studies based on tax data (Aaberge & Atkinson, 2010; Aaberge *et al.*, 2016) that income inequality in Norway has fallen substantially over the past 150 years. However, it might still appear surprising that inequality in nineteenth-century Norway was so high, particularly when compared to Russia. As the present paper is defined on a specific population (men aged above

25) and with several assumptions as described in Section 3, we now consider whether, and how, adjusting these would change the estimated Norwegian income inequality.

First, we can see how the Norwegian Gini would respond if a simple back-of-the-envelope calculation is applied to translate the men-aged-above-25 Gini to a household Gini. Second, we can adjust the assumptions used in the calculation of the Norwegian Gini to better match those used in the other countries.

There are two adjustments that need to be made to convert the Gini coefficient of 0.546 for adult men in Norway to a household basis. First, we must account for the fact that some households have multiple income holders. Second, we must account for the households that are not headed by men aged above 25. From the 1865 data, a tabulation of households by the number of men aged above 25 can be obtained. Of a total of 338,795 household, 232,494 are headed by one man aged 25 or above; 60,023 households have two or more men aged above 25, and 46,278 have zero (headed by either unmarried or widowed women or by men aged below 25). A simple framework for an extension of the data set is set out in the Appendix. If one assumes that low-income households are more likely to include several adult men than high-income households, and that men who are not head-of-household have incomes toward the lower end, the estimated Gini coefficient is reduced to 0.537. This is likely to be a conservative estimate. If we take either assumption to the other extreme, a Gini coefficient of around 0.60 is obtained. Hence, there is no indication that a household-basis Gini for Norway would be any *lower* than that reported for men aged above 25 here. For this reason, in order to better understand why the Norwegian Gini estimate is so high, we move to a discussion of the assumptions used in the calculation of the Gini coefficient.

Section 3 lists several assumptions that have to be made in order to estimate the Gini coefficient from the available sources. One can directly adjust a number of these. Table A17 shows the result of four key adjustments: the dispersion parameters used in imputing top incomes; the number of days worked; the skill premium assumed for lower incomes; and the dispersion parameter in the lognormal distribution used for the lower end of the income distribution. None of these changes the Gini coefficient by more than a couple of points.

Some of the comparison studies do not use within-group imputations of income inequality for upper income groups. An alternative robustness check is therefore to remove all dispersion for the income group above 250 Spd, while maintaining the mean income of this group. This reduces the Gini coefficient from 0.546 to 0.523.

As all of these robustness checks still give a Gini coefficient of more than 0.5, we maintain that income inequality in Norway was high in this period, level with the US and Britain and substantially higher than Russia.

### 6.2 Feasibility and applicability: lessons from a detailed estimate

It is evident from the results presented in this paper that there is substantial variation of incomes both between and within occupation groups. Because of this, the level of granularity used in a given analysis matters greatly for estimates of overall inequality. As shown in Table 3, not taking into account the dispersion within occupation groups leads to a substantially lower estimate of the Gini coefficient.

An increasing body of economic history literature makes use of linked micro data, where records for the same individual in multiple data sets are linked together to compare outcomes at different times. In most cases, one or both of these sources are official censuses, but with the exception of the 1940 U.S. census, none of the large-scale public use census data sets contain any information on income at the individual level. On the other hand, the individuals' occupations are usually recorded, and various measures of occupational income scores are often in practice used in lieu of actual income information (Abramitzky *et al.*, 2012; Olivetti & Paserman, 2015). This is a reasonable solution given present-day data availability, but as the analysis here has shown, relying on average scores by occupation removes a substantial part of the actual variation in economic status in the population.

The income estimates from this paper cannot be linked directly to specific census records as the within-occupation dispersions are obtained from tabulations that do not identify individuals. However, the occupational means (nationally, separately for rural and urban areas or by municipality) can be used to impute incomes in individual mobility studies. For future research, the underlying sources at the individual level could be consulted to construct individual-level samples. Among other things, this would give insight on whether different results are found if one uses occupation-averaged instead of individual-level income data in studies of historical mobility.

### 6.3 Concluding comments

This paper has established the feasibility of combining detailed nineteenth-century income tabulations with census data in order to produce a well-founded estimate of income inequality. The assumptions needed have been clearly laid out, and it is shown that having data tabulated explicitly by incomes is a key requirement in obtaining a precise estimate.

Obtaining estimates of income inequality at the municipal level allows for an examination into whether patterns predicted by standard theories of the evolution of income inequality can be observed in the cross-section. Economic development is found to be positively associated with higher income inequality, while pastoralism and fisheries are associated with lower income inequality. However, when these characteristics are examined together in a joint framework, the regression coefficients on measures of development other than the level of income are small and not statistically significant.

One interpretation of this result is that climate and agricultural structure (as discussed by Engerman & Sokoloff, 2002; Easterly, 2007; Clark & Gray, 2014) is more relevant in explaining income inequality than economic development (Kuznets, 1955; Van Zanden, 1995). However, there are more nuanced readings of Kuznets' hypothesis than a simple relationship between development and inequality, some of which have support in the data presented here. Cities are found to have higher income inequality than rural areas, and among workers, there is higher income inequality in areas with more manufacturing activity. The generally high literacy levels throughout nineteenth-century Norway, as well as the presence of a fairly unified rural school system, suggests that the relationship between agriculture and income inequality does not operate solely through the role of education.

The "inequality possibility frontier" framework (Milanovic *et al.*, 2011) finds clear support in the data. High-income municipalities can sustain higher income inequality than low-income municipalities, possibly due to the role of subsistence income.

The estimated Gini coefficient of 0.546 is high compared to what we know about inequality in other countries in this time period. This is somewhat puzzling, as subsequent development in Norway was relatively peaceful and not marked by confrontations between social classes. However, high inequality manifested itself in ways other than social unrest, most notably in very high emigration rates, with the first substantial wave starting around 1866.

# References


Aaberge, Rolf, & Atkinson, Anthony. 2010. Top Incomes in Norway. *In:* Atkinson, A, & Piketty, T (eds), *Top incomes: A global perspective*. OUP.

Aaberge, Rolf, Atkinson, Anthony, & Modalsli, Jørgen. 2016. On the measurement of long-run income inequality: Empirical evidence from Norway, 1875-2013. *Statistics Norway Discussion Paper*, **847**.

Abramitzky, Ran, Boustan, Leah Platt, & Eriksson, Katherine. 2012. Europe's Tired, Poor, Huddled Masses: Self-Selection and Economic Outcomes in the Age of Mass Migration. *American Economic Review*, **102**(5), 1832–1856.

Acemoglu, Daron, Johnson, Simon, & Robinson, James A. 2005. Institutions as a Fundamental Cause of Long-Run Growth. *Handbook of Economic Growth*, **1A**, 385–472.

Alfani, Guido. 2013. Economic inequality in northwestern Italy: a long-term view. *mimeo, Bocconi University*.

Atkinson, Tony, & Piketty, Thomas. 2007. *Top Incomes over the Twentieth Century*. Oxford University Press.

Bolt, Jutta, & van Zanden, Jan Luiten. 2013. The First Update of the Maddison Project; Re-estimating Growth Before 1820. *Maddison-Project Working Paper*, **4**.

Clark, Gregory. 2005. The Condition of the Working Class in England, 1209-2004. *Journal of Political Economy*, **113**(6).

Clark, Gregory, & Gray, Rowena. 2014. Geography is not destiny: geography, institutions and literacy in England, 1837-63. *Oxford Economic Papers*, **66**(4), 1042–1069.

Easterly, William. 2007. Inequality does cause underdevelopment: Insights from a new instrument. *Journal of Development Economics*, **84**(2), 755–776.

Enflo, Kerstin, & Roses, Joan Ramon. 2015. Coping with regional inequality in Sweden: structural change, migrations and policy, 1860-2000. *Economic History Review*, **68**(1), 191–217.

Engerman, Stanley, & Sokoloff, Kenneth. 2002. Factor Endowments, Inequality, and Paths of Development among New World Economies. *Economia*, **3**(1), 41–109.

Ezcurra, R., Gil, C., Pascual, P., & Rapún, M. 2005. Regional inequality in the European Union: does industry mix matter? *Regional Studies*, **39**(6), 679–697.

Galor, Oded, Moav, Omer, & Vollrath, Dietrich. 2009. Inequality in Landownership, the Emergence of Human-Capital Promoting Institutions, and the Great Divergence. *Review of Economic Studies*, **76**(1), 143–179.

Gennaioli, Nicola, Porta, Rafael La, Lopez-de Silanes, Florencio, & Shleifer, Andrei. 2013. Human capital and regional development. *Quarterly Journal of Economics*, **128**(1), 105–164.

Gerdrup, Karsten. 1998. Skattesystem og skattestatistikk i et historisk perspektiv. *Rapporter fra Statistisk Sentralbyrå*, **6**.

Grytten, Ola H. 2004. A Consumer Price Index for Norway 1516-2003. *In:* Øyvind Eitrheim, Klovland, Jan T, & Qvigstad, Jan F (eds), *Historical monetary statistics for Norway.* Norges Bank.

Grytten, Ola H. 2007. Nominal wages in Norway 1726-1940 by occupation. *In:* Øyvind Eitrheim, Klovland, Jan T, & Qvigstad, Jan F (eds), *Historical monetary statistics for Norway, Part II.* Norges Bank.

Hodne, Fritz, & Grytten, Ola Honningdal. 2000. *Norsk økonomi i det 19. århundre.* Fagbokforlaget.

Kopczuk, Wojciech, Saez, Emmanuel, & Song, Jae. 2010. Earnings inequality and mobility in the United States: Evidence from Social Security data since 1937. *Quarterly Journal of Economics*, **125**(1), 91–128.

Kuznets, Simon. 1955. Economic Growth and Income Inequality. *American Economic Review*, **45**(1), 1–28.

Lambert, Peter J., & Aronson, J. Richard. 1993. Inequality Decomposition Analysis and the Gini Coefficient Revisited. *Economic Journal*, **103**(420), 1221–1227.

Lindert, Peter H. 2000. Three centuries of inequality in Britain and America. *Handbook of Income Distribution*, **1**, 167–216.

Lindert, Peter H, & Williamson, Jeffrey G. 2012. American incomes 1774-1860. *NBER Working paper*, **18369**.

Martines-Galarraga, Julio, Roses, Joan Ramon, & Tirado, Daniel A. 2015. The long-term patterns of regional inequality in Spain, 1860-2000. *Regional Studies*, **49**(4), 502–517.

Milanovic, Branko, Lindert, Peter H., & Williamson, Jeffery G. 2011. Pre-Industrial Inequality. *Economic Journal*, **121**(551), 255–272.

Morrisson, Christian. 2000. Historical perspectives on income distribution: The case of Europe. *Handbook of Income Distribution*, **1**, 217–260.

Nafziger, Steven, & Lindert, Peter H. 2012. Russian Inequality on the Eve of Revolution. *NBER Working paper*, **18383**.

Nicolini, Esteban A., & Ramos Palencia, Fernando. 2016. Decomposing income inequality in a backward pre-industrial economy: Old Castile (Spain) in the middle of the eighteenth century. *Economic History Review*, **69**(3), 747–772.

Norwegian Department of Finance. 1871. *Om civile og militære Embeders Lønningsforholde.* Ot prp 3.

Norwegian Department of Justice. 1871. *Tabeller til oplysning om stemmerets- indtægts- og skatteforholdene i Norge i aaret 1868.* Norges Officielle Statistik B no. 3.

Norwegian Department of the Interior. 1868. *Resultaterne af Folketællingen i Norge.* Norges Officielle Statistik C no. 1.

Norwegian Department of the Interior. 1870. *Tabeller vedkommende Norges handel og skibsfart i aaret 1868.* Norges Officielle Statistik C no 3.

Norwegian Department of the Interior. 1876. *Statistiske opgaver til belysning af Norges industrielle forholde i aarene 1870-1874.* Norges Officielle Statistik C no. 13.

Olivetti, Claudia, & Paserman, M Daniele. 2015. In the Name of the Son (and the Daughter): Intergenerational Mobility in the United States, 1850-1930. *American Economic Review*, **105**(8), 2695–2724.

Seip, Jens Arup. 1997. *Utsikt over Norges historie: tidsrommet 1814-ca. 1860, tidsrommet ca. 1850-1884 (originally published 1974, 1981).* Gyldendal Fakkel.

Soltow, Lee. 1965. *Toward Income Equality in Norway.* University of Wisconsin Press.

Statistics Norway. 1880. *Statistik angaaende det Norske Jordbrug i femaarsperioden 1871-1875.* Norges Officielle Statistik C no 15.

Statistics Norway. 1888. *Tabeller vedkommende arbeidslønninger i aarene 1875, 1880, 1885 samt delvis tidligere Aar*. Norges Officielle Statistik Tredie Række no. 61.

Statistics Norway. 1965. *Nasjonalregnskap / National Accounts 1865-1960*. Norges Offisielle Statistikk XII 163.

Sundt, Eilert. 1855. *Om Giftermaal i Norge*. Mallings Bogtrykkeri. http://www.rhd.uit.no/sundt/bind2/eilert_sundt_bd2b.html.

Tapia, Francisco J Beltran, & Martines-Galarraga, Julio. 2013. Inequality and early economic growth in Spain, 1860-1930. *mimeo*.

The Digital Archive (The National Archive), Norwegian Historical Data Centre (University of Tromsø) and the Minnesota Population Center. 2008. *National Sample of the 1865 Census of Norway, Version 2.0 [Data file.]*. University of Tromsø.

Try, Hans. 1979. *To kulturer - en stat 1851-1884*. Oslo: Cappelen.

Van Zanden, Jan L. 1995. Tracing the Beginning of the Kuznets Curve: Western Europe during the Early Modern Period. *Economic History Review*, **48**(4), 643–664.

**Appendix (for online publication)**

# A File appendices

This paper is accompanied by two data files in Stata format to facilitate further analysis.

- One individual-level income file for calculation of any inequality measure across geographical or occupational groups.
  - Do note that this is not individual data, and that individual observations reflect a discretization of a theoretical continuous distribution. For this reason, operators such as `max` or `min` may not be applied to the data. Some of the very high values simply reflect high dispersion in high-income groups and should only be considered as inputs to a particular income inequality metric. Metrics placing very high weight on high-income observations might not be appropriate.
- One municipality-level file with municipality Gini coefficients and a range of covariates, as described in Section 5.

# B Tables and figures

## B.1 Historical context

Figure A1 shows the development over time of some of the key economic-demographic indicators discussed in Section 2.1. The first panel shows population growth. After harsh conditions during the Napoleonic Wars, with grain imports from Denmark being blocked by the British navy, the birth surplus was fairly stable in the following decades, as seen in the upper-left panel. The dip in the 1840s can be attributed to the low level of births in the 1810s during the war (Sundt, 1855, chap. 7). The average net birth surplus of 1.2 percent after 1850 would lead to a doubling of the population every 35 years. However, from mid-century onward large numbers left Norway for the United States, reaching a peak of 1.5 percent of the population in 1882. In 1865, however, we are still at the very beginning of the first of the three waves, and population pressure was still heavy. The second panel of Figure A1 shows the development of the urban (and total) population from 1815 to 1910.

The second panel shows the level of total population, as measured at decennial censuses, as well as the urban population.

The third and fourth panels shows the development of the four major occupational gropus — farmers, cottagers, servants and industrial workers.

Section 2.1 also refers to industrial statistics and the composition of imports and exports. These are shown in Tables A1 and A2, respectively. The source for the trade statistics is Norwegian Department of the Interior (1870), page 4 (imports and exports) and 111 (shipping).

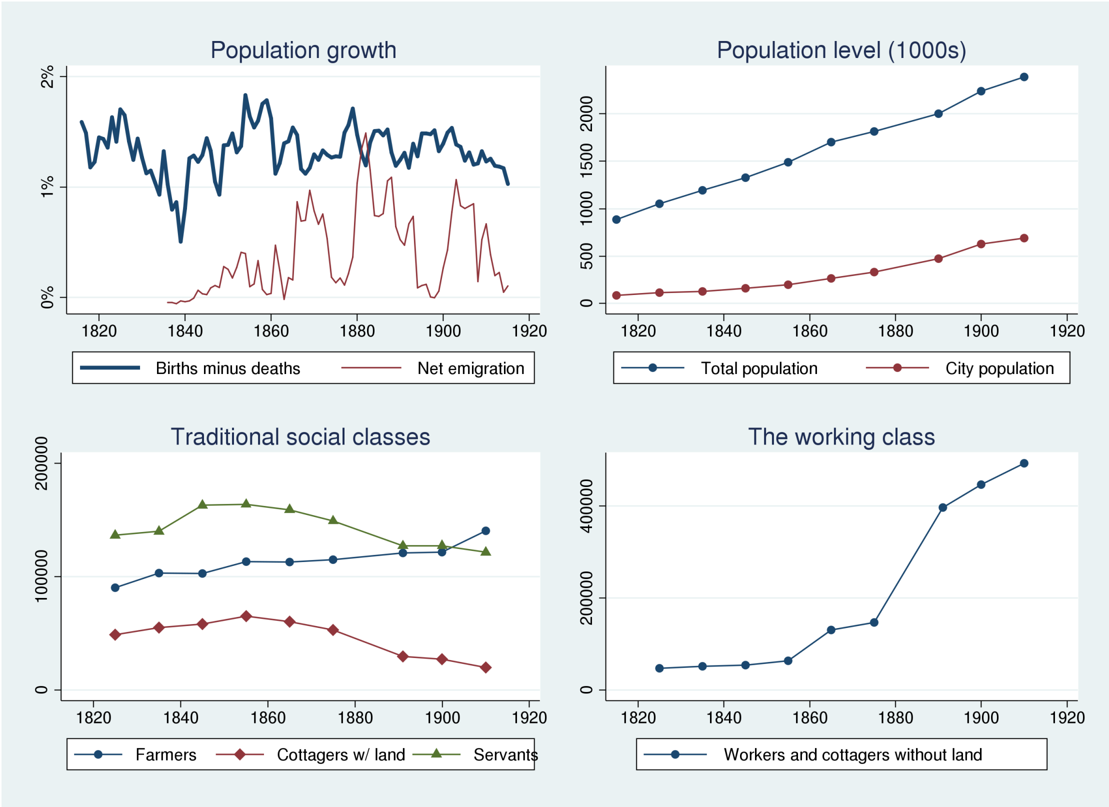


Figure A1: Economic development in Norway, 1820-1910. Source: Official statistics (censuses and birth/death statistics)

| **Industrial composition** | Rural | Urban | Total |
|---|---|---|---|
| Agriculture, forestry, fishery | 72% | 3% | 61% |
| Mining and industry | 10% | 39% | 14% |
| Trade, shipping, transport | 4% | 33% | 9% |
| Unclassified work | 5% | 9% | 6% |
| "Immaterial" work | 2% | 10% | 3% |
| Non-productive | 7% | 6% | 7% |

Table A1: Industrial composition, 1868, from official statistics

| **Imports** (mill. Spd) | | **Exports** (mill. Spd) | |
|---|---|---|---|
| Grain | 8.5 | Fish | 5.9 |
| Other food and drink | 5.3 | Fish oil | 1.1 |
| Clothing | 3.3 | Lumber | 7.7 |
| Oil, coal, metal | 3.6 | Other | 2.4 |
| Other | 5.8 | | |
| | | *Shipping (Norw. ↔ abroad)* | *4.4* |
| | | *Shipping (abroad ↔ abroad)* | *10.0* |

Table A2: Imports and exports, 1868, from official statistics

## B.2 Alternative inequality metrics

Table A3 shows the regional decomposition of inequality measures other than the Gini coefficient.

| | Gini | | | GE(1) (Theil) | | | A(1) | | |
|---|---|---|---|---|---|---|---|---|---|
| | Tot | Rur. | Urb. | Tot | Rur. | Urb. | Tot | Rur. | Urb. |
| Total for Norway 1868 | 0.546 | 0.346 | 0.715 | 1.016 | 0.380 | 1.248 | 0.418 | 0.200 | 0.632 |
| By region (*Stift*): | | | | | | | | | |
| Christiania (East) | 0.643 | 0.418 | 0.725 | 1.260 | 0.574 | 1.275 | 0.529 | 0.272 | 0.650 |
| Hamar (Central inland) | 0.349 | 0.331 | 0.576 | 0.333 | 0.291 | 0.723 | 0.200 | 0.183 | 0.433 |
| Christiansand (South) | 0.537 | 0.277 | 0.707 | 0.991 | 0.195 | 1.220 | 0.410 | 0.132 | 0.617 |
| Bergen (West) | 0.487 | 0.306 | 0.696 | 0.813 | 0.226 | 1.204 | 0.351 | 0.154 | 0.611 |
| Throndhjem (Central coast) | 0.549 | 0.388 | 0.737 | 1.095 | 0.583 | 1.330 | 0.427 | 0.250 | 0.671 |
| Tromsø (North) | 0.338 | 0.269 | 0.563 | 0.383 | 0.218 | 0.771 | 0.198 | 0.133 | 0.418 |
| By occupation group: | | | | | | | | | |
| Civil servants | 0.665 | 0.500 | 0.593 | 0.991 | 0.544 | 0.773 | 0.573 | 0.339 | 0.511 |
| Farmers | 0.369 | 0.366 | 0.672 | 0.367 | 0.358 | 1.127 | 0.206 | 0.203 | 0.566 |
| Merchants and shopkeepers | 0.622 | 0.498 | 0.560 | 0.839 | 0.498 | 0.696 | 0.547 | 0.339 | 0.474 |
| Craftsmen and artisans | 0.568 | 0.172 | 0.671 | 1.172 | 0.096 | 1.312 | 0.453 | 0.062 | 0.565 |
| Owners | 0.668 | 0.537 | 0.708 | 1.084 | 0.643 | 1.172 | 0.559 | 0.397 | 0.625 |
| Engineers | 0.602 | 0.539 | 0.553 | 0.669 | 0.528 | 0.558 | 0.532 | 0.417 | 0.489 |
| Clerks | 0.692 | 0.434 | 0.670 | 1.093 | 0.507 | 0.972 | 0.604 | 0.283 | 0.591 |
| Students and graduates | 0.673 | 0.560 | 0.619 | 0.925 | 0.632 | 0.766 | 0.627 | 0.424 | 0.604 |
| Shipowners | 0.596 | 0.469 | 0.557 | 0.720 | 0.433 | 0.591 | 0.479 | 0.320 | 0.470 |
| Officers, merchant marine | 0.710 | 0.669 | 0.676 | 1.177 | 1.391 | 0.988 | 0.623 | 0.557 | 0.608 |
| Fishermen and other seamen | 0.229 | 0.175 | 0.305 | 0.222 | 0.092 | 0.432 | 0.109 | 0.063 | 0.185 |
| Cottagers | 0.164 | 0.164 | 0.125 | 0.051 | 0.051 | 0.027 | 0.048 | 0.048 | 0.026 |
| Retirees | 0.152 | 0.152 | 0.161 | 0.107 | 0.106 | 0.133 | 0.057 | 0.057 | 0.073 |
| Laborers and workers | 0.264 | 0.174 | 0.366 | 0.321 | 0.076 | 0.644 | 0.138 | 0.056 | 0.248 |
| Coachmen | 0.402 | 0.161 | 0.490 | 0.587 | 0.123 | 0.716 | 0.269 | 0.064 | 0.346 |
| Managers | 0.490 | 0.355 | 0.647 | 0.908 | 0.486 | 1.306 | 0.381 | 0.233 | 0.566 |
| Nomads | 0.389 | 0.390 | . | 0.325 | 0.325 | . | 0.221 | 0.222 | . |
| Servants | 0.156 | 0.159 | 0.105 | 0.040 | 0.041 | 0.017 | 0.044 | 0.046 | 0.018 |
| Paupers | 0.295 | 0.299 | 0.254 | 0.138 | 0.142 | 0.102 | 0.138 | 0.142 | 0.101 |
| Others | 0.417 | 0.179 | 0.735 | 0.865 | 0.132 | 1.346 | 0.312 | 0.072 | 0.664 |

Table A3: Alternative inequality measures (cf. Table 2)

## B.3 Municipality characteristics: Descriptive statistics and correlation coefficients

Descriptive statistics are given in Table A4.

| Sample | Urban+Rural | | Rural | | Subsample: Rural with 1838 data | |
|---|---|---|---|---|---|---|
| | Mean | Std. Dev. | Mean | Std. Dev. | Mean | Std. Dev. |
| Gini coefficient (×100) | 31.37 | 13.75 | 27.98 | 9.70 | 26.57 | 7.39 |
| Mean income | 147.35 | 127.93 | 116.33 | 56.71 | 111.69 | 20.02 |
| Urban | 0.11 | 0.32 | 0.00 | 0.00 | 0.00 | 0.00 |
| Population size | 773.14 | 747.55 | 739.57 | 426.83 | 879.60 | 483.24 |
| Has factories | 0.55 | 0.50 | 0.52 | 0.50 | 0.52 | 0.50 |
| Distance to city | 41.48 | 40.40 | 46.75 | 39.91 | 54.21 | 42.18 |
| Cattle pc | 2.51 | 1.31 | 2.79 | 1.08 | 2.65 | 1.00 |
| Sheep and goats pc | 5.70 | 4.37 | 6.38 | 4.15 | 5.32 | 2.64 |
| Crop value pc | 119.20 | 72.57 | 132.94 | 64.75 | 135.67 | 57.06 |
| Fishery pop share | 0.08 | 0.16 | 0.08 | 0.17 | 0.04 | 0.10 |
| 1838 land Gini | | | | | 47.75 | 9.61 |
| Number of obs. | 479 | | 425 | | 62 | |

Table A4: Summary statistics

See Tables A5-A6 for correlation coefficients. Information on manufacturing is missing for one municipality (Askvoll).

| Variable | $\rho(\text{Gini}, \text{Variable})$ | s.e. | $N$ |
|---|---|---|---|
| **All municipalities** | | | |
| Mean income | 0.776 | ( 0.029 ) | 479 |
| Population size | 0.257 | ( 0.044 ) | 479 |
| Urban | 0.691 | ( 0.033 ) | 479 |
| Manufacturing pop. share | 0.318 | ( 0.043 ) | 478 |
| Has factories | 0.244 | ( 0.044 ) | 478 |
| Fishery pop. share | -0.134 | ( 0.045 ) | 479 |
| **Rural municipalities** | | | |
| Mean income | 0.664 | ( 0.036 ) | 425 |
| Population size | 0.185 | ( 0.048 ) | 425 |
| Manufacturing pop. share | 0.161 | ( 0.048 ) | 424 |
| Has factories | 0.168 | ( 0.048 ) | 424 |
| Distance to city | -0.186 | ( 0.048 ) | 425 |
| Crop value per capita | 0.133 | ( 0.048 ) | 425 |
| Sheep and goats per capita | -0.311 | ( 0.046 ) | 425 |
| Cattle per capita | -0.184 | ( 0.048 ) | 425 |
| Fishery pop. share | -0.094 | ( 0.048 ) | 425 |
| **Rural subsample** | | | |
| 1838 land Gini | 0.633 | ( 0.099 ) | 62 |
| **Urban municipalities** | | | |
| Mean income | 0.706 | ( 0.097 ) | 54 |
| Population size | 0.437 | ( 0.124 ) | 54 |
| Manufacturing pop. share | 0.095 | ( 0.137 ) | 54 |
| Has factories | 0.313 | ( 0.130 ) | 54 |
| Fishery pop. share | -0.329 | ( 0.130 ) | 54 |

Table A5: Correlation coefficients across municipalities

| Occupation | $\rho(\text{Gini}_{\text{Occ}}, \text{Variable})$ | s.e. | $N$ |
|---|---|---|---|
| Variable: Manufacturing population share | | | |
| Civil servants | 0.070 | ( 0.067 ) | 223 |
| Farmers | 0.170 | ( 0.048 ) | 421 |
| Merchants and shopkeepers | 0.141 | ( 0.144 ) | 48 |
| Craftsmen and artisans | 0.162 | ( 0.052 ) | 367 |
| Officers, merchant marine | 0.201 | ( 0.121 ) | 67 |
| Fishermen and other seamen | 0.099 | ( 0.067 ) | 220 |
| Cottagers | 0.065 | ( 0.050 ) | 395 |
| Retirees | -0.001 | ( 0.051 ) | 387 |
| Laborers and workers | 0.247 | ( 0.049 ) | 392 |
| Managers | 0.254 | ( 0.143 ) | 47 |
| Servants | -0.203 | ( 0.052 ) | 352 |
| Paupers | -0.093 | ( 0.066 ) | 229 |
| Others | 0.165 | ( 0.062 ) | 252 |
| Between-group | 0.117 | ( 0.049 ) | 408 |
| Overall | 0.161 | ( 0.048 ) | 424 |
| Variable: Distance to city | | | |
| Civil servants | -0.108 | ( 0.067 ) | 224 |
| Farmers | -0.187 | ( 0.048 ) | 422 |
| Merchants and shopkeepers | -0.030 | ( 0.146 ) | 48 |
| Craftsmen and artisans | -0.261 | ( 0.050 ) | 368 |
| Officers, merchant marine | -0.363 | ( 0.115 ) | 67 |
| Fishermen and other seamen | -0.263 | ( 0.065 ) | 221 |
| Cottagers | -0.191 | ( 0.049 ) | 396 |
| Retirees | -0.010 | ( 0.051 ) | 388 |
| Laborers and workers | -0.146 | ( 0.050 ) | 393 |
| Managers | -0.342 | ( 0.139 ) | 47 |
| Servants | -0.058 | ( 0.053 ) | 353 |
| Paupers | -0.212 | ( 0.065 ) | 229 |
| Others | -0.308 | ( 0.060 ) | 253 |
| Between-group | -0.148 | ( 0.049 ) | 409 |
| Overall | -0.186 | ( 0.048 ) | 425 |
| Variable: Crop value per capita | | | |
| Civil servants | 0.208 | ( 0.066 ) | 224 |
| Farmers | 0.212 | ( 0.048 ) | 422 |
| Merchants and shopkeepers | 0.072 | ( 0.145 ) | 48 |
| Craftsmen and artisans | 0.134 | ( 0.052 ) | 368 |
| Officers, merchant marine | 0.013 | ( 0.123 ) | 67 |
| Fishermen and other seamen | 0.047 | ( 0.067 ) | 221 |
| Cottagers | 0.262 | ( 0.049 ) | 396 |
| Retirees | 0.078 | ( 0.051 ) | 388 |
| Laborers and workers | 0.128 | ( 0.050 ) | 393 |
| Managers | 0.278 | ( 0.142 ) | 47 |
| Servants | 0.047 | ( 0.053 ) | 353 |
| Paupers | 0.290 | ( 0.063 ) | 229 |
| Others | 0.316 | ( 0.060 ) | 253 |
| Between-group | 0.147 | ( 0.049 ) | 409 |
| Overall | 0.133 | ( 0.048 ) | 425 |
| Variable: 1838 land Gini | | | |
| Farmers | 0.590 | ( 0.103 ) | 62 |
| Craftsmen and artisans | 0.498 | ( 0.118 ) | 55 |
| Cottagers | 0.514 | ( 0.112 ) | 60 |
| Retirees | 0.319 | ( 0.127 ) | 57 |
| Laborers and workers | 0.594 | ( 0.107 ) | 58 |
| Servants | 0.237 | ( 0.130 ) | 57 |
| Others | 0.351 | ( 0.146 ) | 42 |
| Between-group | 0.558 | ( 0.107 ) | 61 |
| Overall | 0.633 | ( 0.099 ) | 62 |

Table A6: Correlation coefficients across municipalities within occupations

## C Regressions

The analysis behind table in the main text (Table 4) is re-done for all municipalities (rural and urban) in Table A7 and for the subsample with historical land Ginis in Table A10.

The analysis for the rural sample is re-done including a variable for distance to city (Table A8) and removing mean incomes from the regression (Table A9). Table A11 shows the analysis with historical land Ginis.

In all regressions presented here, the dependent variable is multiplied by 100 (i.e. zero=no inequality and 100=full inequality) for ease of interpretation of the coefficients.

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| | Gini | Gini | Gini | Gini | Gini | Gini |
| Log income | 26.68*** | 29.83*** | 28.40*** | 27.88*** | 26.46*** | 27.94*** |
| | (43.68) | (45.74) | (37.93) | (36.68) | (40.85) | (36.41) |
| Crop value pc | | 0.0392*** | 0.0352*** | 0.0303*** | | 0.0302*** |
| | | (9.46) | (8.09) | (6.61) | | (6.59) |
| Sheep/goats pc | | | -0.293*** | -0.283*** | | -0.288*** |
| | | | (-3.24) | (-3.17) | | (-3.21) |
| Cattle pc | | | 0.0765 | -0.00773 | | -0.0275 |
| | | | (0.24) | (-0.02) | | (-0.09) |
| Fishery pop. share | | | | -5.402*** | | -5.524*** |
| | | | | (-3.15) | | (-3.20) |
| Manufacturing pop. share | | | | | 9.824 | -5.354 |
| | | | | | (1.02) | (-0.60) |
| Constant | -97.77*** | -117.7*** | -108.8*** | -105.1*** | -96.83*** | -105.2*** |
| | (-32.92) | (-34.16) | (-26.05) | (-24.45) | (-31.13) | (-24.43) |
| Observations | 478 | 478 | 478 | 478 | 478 | 478 |
| $R^2$ | 0.800 | 0.832 | 0.838 | 0.841 | 0.801 | 0.841 |

*t* statistics in parentheses

* p<0.1, ** p<0.05, *** p<0.01

Table A7: Reference table, all municipalities (urban+rural)

| | (1) Gini | (2) Gini | (3) Gini | (4) Gini | (5) Gini | (6) Gini |
|---|---|---|---|---|---|---|
| Log income | 29.35*** | 31.56*** | 30.89*** | 30.97*** | 29.22*** | 30.98*** |
| | (26.25) | (30.44) | (29.14) | (29.38) | (25.96) | (29.30) |
| Distance to city | -0.0355*** | -0.0152** | -0.0136* | -0.00887 | -0.0336*** | -0.00900 |
| | (-4.95) | (-2.23) | (-1.91) | (-1.22) | (-4.55) | (-1.22) |
| Crop value pc | | 0.0416*** | 0.0378*** | 0.0340*** | | 0.0339*** |
| | | (9.72) | (8.37) | (7.14) | | (7.13) |
| Sheep/goats pc | | | -0.236*** | -0.227*** | | -0.229*** |
| | | | (-2.73) | (-2.64) | | (-2.63) |
| Cattle pc | | | 0.353 | 0.172 | | 0.167 |
| | | | (1.08) | (0.52) | | (0.50) |
| Fishery pop. share | | | | -4.448** | | -4.477** |
| | | | | (-2.46) | | (-2.46) |
| Manufacturing pop. share | | | | | 15.66 | -2.063 |
| | | | | | (1.07) | (-0.15) |
| Constant | -108.7*** | -125.6*** | -121.5*** | -120.7*** | -108.2*** | -120.7*** |
| | (-20.48) | (-24.60) | (-22.70) | (-22.65) | (-20.36) | (-22.62) |
| Observations | 424 | 424 | 424 | 424 | 424 | 424 |
| $R^2$ | 0.634 | 0.701 | 0.707 | 0.711 | 0.635 | 0.711 |

*t* statistics in parentheses

* p<0.1, ** p<0.05, *** p<0.01

Table A8: Reference table, controls for distance to city added

| | (1) Gini | (2) Gini | (3) Gini | (4) Gini | (5) Gini | (6) Gini |
|---|---|---|---|---|---|---|
| Crop value pc | | 0.0200*** | 0.00909 | 0.00433 | | 0.00455 |
| | | (2.77) | (1.24) | (0.54) | | (0.57) |
| Sheep/goats pc | | | -0.710*** | -0.702*** | | -0.672*** |
| | | | (-4.79) | (-4.74) | | (-4.47) |
| Cattle pc | | | 0.0923 | -0.0389 | | 0.0501 |
| | | | (0.17) | (-0.07) | | (0.09) |
| Fishery pop. share | | | | -4.459 | | -3.970 |
| | | | | (-1.46) | | (-1.29) |
| Manufacturing pop. share | | | | | 77.10*** | 27.12 |
| | | | | | (3.35) | (1.14) |
| Constant | 27.98*** | 25.31*** | 31.03*** | 32.35*** | 27.29*** | 31.60*** |
| | (59.31) | (23.64) | (19.82) | (17.93) | (53.65) | (16.46) |
| Observations | 424 | 424 | 424 | 424 | 424 | 424 |
| $R^2$ | 0.000 | 0.018 | 0.101 | 0.105 | 0.026 | 0.108 |

*t* statistics in parentheses

* p<0.1, ** p<0.05, *** p<0.01

Table A9: Reference table, without income control

| | (1) Gini | (2) Gini | (3) Gini | (4) Gini | (5) Gini | (6) Gini |
|---|---|---|---|---|---|---|
| Log income | 21.73*** | 25.05*** | 22.95*** | 23.04*** | 20.39*** | 22.81*** |
| | (4.57) | (5.22) | (5.42) | (5.72) | (4.55) | (5.64) |
| Crop value pc | | 0.0341** | 0.0335** | 0.0272** | | 0.0275** |
| | | (2.34) | (2.45) | (2.07) | | (2.08) |
| Sheep/goats pc | | | -1.628*** | -1.317*** | | -1.179*** |
| | | | (-4.56) | (-3.67) | | (-3.00) |
| Cattle pc | | | 2.932*** | 2.036** | | 2.123** |
| | | | (3.01) | (2.07) | | (2.14) |
| Fishery pop. share | | | | -19.22** | | -18.02** |
| | | | | (-2.66) | | (-2.44) |
| Manufacturing pop. share | | | | | 233.9*** | 73.15 |
| | | | | | (3.02) | (0.87) |
| Constant | -75.58*** | -95.83*** | -85.00*** | -83.01*** | -70.62*** | -83.39*** |
| | (-3.38) | (-4.12) | (-4.05) | (-4.15) | (-3.35) | (-4.16) |
| Observations | 62 | 62 | 62 | 62 | 62 | 62 |
| $R^2$ | 0.258 | 0.321 | 0.503 | 0.558 | 0.358 | 0.564 |

*t* statistics in parentheses

* p<0.1, ** p<0.05, *** p<0.01

Table A10: Subsample, reference table

| | (1) Gini | (2) Gini | (3) Gini | (4) Gini | (5) Gini | (6) Gini |
|---|---|---|---|---|---|---|
| Log income | 18.39*** | 20.53*** | 20.97*** | 21.16*** | 18.36*** | 21.11*** |
| | (5.07) | (5.46) | (5.58) | (5.57) | (5.02) | (5.51) |
| 1838 land Gini | 0.442*** | 0.420*** | 0.378*** | 0.346*** | 0.432*** | 0.338*** |
| | (6.77) | (6.44) | (4.21) | (3.10) | (5.60) | (2.96) |
| Crop value pc | | 0.0203* | 0.0267** | 0.0259** | | 0.0260** |
| | | (1.78) | (2.20) | (2.11) | | (2.10) |
| Sheep/goats pc | | | -0.412 | -0.449 | | -0.412 |
| | | | (-0.96) | (-1.03) | | (-0.91) |
| Cattle pc | | | 1.495 | 1.425 | | 1.475 |
| | | | (1.62) | (1.52) | | (1.54) |
| Fishery pop. share | | | | -4.161 | | -4.003 |
| | | | | (-0.50) | | (-0.48) |
| Manufacturing pop. share | | | | | 17.57 | 30.18 |
| | | | | | (0.24) | (0.38) |
| Constant | -80.99*** | -92.77*** | -95.48*** | -94.14*** | -80.50*** | -94.04*** |
| | (-4.78) | (-5.18) | (-5.12) | (-4.97) | (-4.68) | (-4.92) |
| Observations | 62 | 62 | 62 | 62 | 62 | 62 |
| $R^2$ | 0.583 | 0.604 | 0.622 | 0.624 | 0.583 | 0.625 |

*t* statistics in parentheses

* p<0.1, ** p<0.05, *** p<0.01

Table A11: Subsample, with 1838 Gini

Results for within-occupation inequality (where the municipal Gini coefficient is replaced by the within-occupation Gini coefficient of a given occupation group) are presented in Table A12. Only municipality Ginis based on at least 10 individuals are utilized. As there is not population in all occupations in all municipalities, the samples for each within-occupation Gini are not necessarily comparable. Table A13 balances the panel by only including municipalities where all seven occupation categories have at least 10 individuals.

Within-occupation results for the smaller occupation groups are available on request.

| | (1)<br>Civ. servants | (2)<br>Farmers | (3)<br>Craftsm./Art. | (4)<br>Fisher/Seamen | (5)<br>Cottagers | (6)<br>Lab./Workers | (7)<br>Servants |
|---|---|---|---|---|---|---|---|
| Log income | 6.223** | 20.47*** | 7.860*** | 9.965*** | 4.266*** | 3.984*** | -1.124 |
| | (2.38) | (12.40) | (5.58) | (4.36) | (3.72) | (3.16) | (-1.45) |
| Crop value pc | 0.0237* | 0.0336*** | 0.0136** | 0.0113 | 0.0209*** | 0.00894 | 0.00592* |
| | (1.86) | (4.59) | (2.14) | (1.05) | (4.20) | (1.58) | (1.67) |
| Sheep/goats pc | -0.290 | -0.313** | -0.0678 | 0.385* | 0.0139 | -0.130 | 0.360*** |
| | (-0.83) | (-2.31) | (-0.53) | (1.76) | (0.15) | (-1.21) | (5.04) |
| Cattle pc | 0.307 | -0.732 | -1.212** | -2.218** | -0.177 | -0.0740 | 0.451* |
| | (0.29) | (-1.45) | (-2.44) | (-2.58) | (-0.52) | (-0.19) | (1.83) |
| Fishery pop. share | -10.04** | -8.597*** | -6.992** | -4.288 | -5.854*** | -5.567* | -6.151*** |
| | (-1.98) | (-3.06) | (-2.12) | (-1.31) | (-2.79) | (-1.93) | (-3.97) |
| Mfg. pop. share | 1.079 | 3.376 | 10.31 | 5.134 | 1.218 | 53.12*** | -13.95 |
| | (0.03) | (0.16) | (0.58) | (0.21) | (0.09) | (3.30) | (-1.41) |
| Constant | 8.288 | -69.71*** | -23.39*** | -34.65*** | -10.28* | -8.377 | 14.23*** |
| | (0.61) | (-8.43) | (-3.30) | (-2.98) | (-1.82) | (-1.32) | (3.66) |
| Observations | 223 | 421 | 367 | 220 | 395 | 392 | 352 |
| $R^2$ | 0.100 | 0.374 | 0.167 | 0.133 | 0.121 | 0.116 | 0.272 |

*t* statistics in parentheses

* p<0.1, ** p<0.05, *** p<0.01

Table A12: Regressions of income inequality, by occupation. Dependent variable: Within-occupation Gini coefficient. Rural municipalities (compare to Table 4, column 6). Covariates refer to the the municipality as a whole.

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
|---|---|---|---|---|---|---|---|
| | Civ. servants | Farmers | Craftsm./Art. | Fisher/Seamen | Cottagers | Lab./Workers | Servants |
| Log income | 12.53** | 21.42*** | 21.91*** | 19.95*** | 17.66*** | 12.62*** | -0.782 |
| | (2.44) | (5.90) | (6.17) | (4.29) | (5.94) | (3.66) | (-0.51) |
| Crop value pc | 0.000463 | 0.0515*** | 0.0168 | 0.0165 | 0.0261** | 0.0146 | 0.00847 |
| | (0.02) | (3.80) | (1.27) | (0.95) | (2.35) | (1.14) | (1.49) |
| Sheep/goats pc | -0.581 | -0.127 | 0.441 | 0.350 | 0.508 | 0.215 | 0.348** |
| | (-1.06) | (-0.33) | (1.16) | (0.70) | (1.60) | (0.58) | (2.14) |
| Cattle pc | 1.146 | -2.436* | -3.190** | -2.504 | -0.829 | -1.229 | 0.691 |
| | (0.56) | (-1.69) | (-2.26) | (-1.36) | (-0.70) | (-0.90) | (1.14) |
| Fishery pop. share | -14.16* | -4.670 | -11.35** | -10.50 | -2.056 | -5.330 | -8.800*** |
| | (-1.96) | (-0.91) | (-2.27) | (-1.61) | (-0.49) | (-1.10) | (-4.11) |
| Mfg. pop. share | -22.00 | 10.66 | -8.604 | -13.31 | -19.11 | 45.22* | -16.01 |
| | (-0.63) | (0.43) | (-0.36) | (-0.42) | (-0.95) | (1.94) | (-1.56) |
| Constant | -17.00 | -74.80*** | -87.01*** | -80.17*** | -75.36*** | -48.47*** | 11.82 |
| | (-0.66) | (-4.09) | (-4.86) | (-3.42) | (-5.03) | (-2.79) | (1.54) |
| Observations | 101 | 101 | 101 | 101 | 101 | 101 | 101 |
| $R^2$ | 0.155 | 0.509 | 0.418 | 0.250 | 0.298 | 0.254 | 0.385 |

*t* statistics in parentheses

* p<0.1, ** p<0.05, *** p<0.01

Table A13: Regressions of income inequality, by occupation. Dependent variable: Within-occupation Gini coefficient. Rural municipalities (compare to Table 4, column 6). Covariates refer to the the municipality as a whole. Balanced panel: Only municipalities with sufficiently many observations for all seven occupation categories.

# D Occupation categories

| Category (this paper) | Category in 1868 report | HISCO codes in 1865 census |
|---|---|---|
| Civil servants | Embedsmænd | 05100 |
| | Bestillingsmænd, i Kirkens og Skolens Tjeneste | 06110-06300 |
| | Bestillingsmænd, militære | 07210-07320 |
| | Bestillingsmænd, andre | 09000-13300 |
| | | 14120-14140 |
| | | 17120 |
| | | 19100-20200 |
| | | 31010-31090 |
| | | 36010-37030 |
| | | 37090-38030 |
| | | 39600 |
| | | 58100-58220 |
| | | 58420-58430 |
| Farmers | Gaardbrugere, derunder Leilændinge og Forpagtere | 61110 (A) |
| | | 61240 |
| | | 61320 (A) |
| Merchants and shopkeepers | Handelsmænd | 06400 |
| | | 41010-44320 |
| | | 45220-51090 |
| Craftsmen and artisans | Haandværkere | 06500 |
| | Haandværkssvende | 07500-07990 |
| | | 16300-16400 |
| | | 21230 |
| | | 55200-57040 |
| | | 71300-95700 |
| | | 95920-95990 |
| | | 99450 |
| Owners | Værks- og Fabrikeiere | 21120-21140 |
| | Huseiere | 21190 |
| | | (B) |
| Engineers | Ingenieurer | 02000-03210 |
| Clerks | Kontorister, derunder Handelsbetjente | 32000-33190 |

(Continued on next page)

(Continued from previous page)

| Category (this paper) | Category in 1868 report | HISCO codes in 1865 census |
|---|---|---|
| | | 39000-39500<br>45120 |
| Students and graduates | Studenter<br>Kandidater | (C) |
| Shipowners | Skibsredere | 21160 |
| Officers, merchant marine | Skippere og Styrmænd | 04220-04250 |
| Fishermen and other seamen | Matroser, Fiskere og andre Sømænd | 04260<br>64100-64950<br>98120-98200 |
| Cottagers | Husmænd og Strandsiddere | 61115 (A)<br>61330 (A) |
| Retirees | Føderaadsmænd | 99150<br>(D) |
| Laborers and workers | Faste Arbeidere<br>Dagarbeidere<br>Inderster | 21210<br>37040<br>58300<br>62110-63290<br>71120-71190<br>95910<br>96230-97490<br>98490<br>98720-98730<br>99120-99140<br>99200<br>99430 |
| Coachmen | Vognmænd | 98320-98440<br>98510-98590<br>98900 |
| Managers | Gaards-, Bruugs- og Værksbestyrere | 05300<br>21152<br>21182<br>22110-24100<br>61400 |
| Nomads | Nomader | 61260 |
| Servants | (N/A) | 52020-55100 |

(Continued on next page)

(Continued from previous page)

| Category (this paper) | Category in 1868 report | HISCO codes in 1865 census |
|---|---|---|
| | | 58500-59990 |
| Paupers | (N/A) | (E) |
| Others | Andre | 13920-13990<br>15220-16130<br>17130-17300<br>21194<br>64970<br>99440<br>99300 (F)<br>99999 (F) |

*Notes:*

*(A): if not "occupation status"=="retired"*

*(B): Those with missing occupation info (99300, 99999) and "occupation status"=="Owner" are placed here*

*(C): Anyone with "occupation status"=="Student" is placed here regardless of occupation code*

*(D): Those with occupation codes 61110, 61115, 61320 (farmers and cottagers) and occupation status "Retired" are also placed here*

*(E): Anyone with "occupation status"=="Pauper" is placed here regardless of occupation code*

*(F) if not "occupation status"=="owner"*

Table A14: Occupation categories. First column has the categories used in this paper. Second column lists the corresponding categories in the 1868 report. Third column lists the corresponding HISCO codes in the 1865 census. There is no correspondence assumed between second and third column other than through the category listed in the first column.

Occupation categories used in this paper are shown in Table A14. In a small number of cases, there are "too few" in the census data to match the numbers in the 1868 report. In such cases, the number of the category is increased and that of an adjoining category is decreased by a similar amount.

The adjustments made are as follows (numbers are total for the country as a whole; only those with 10 or more individuals listed): **Civil servants**: 498 people from Farmers; 27 people from Craftsmen and artisans; 11 people from Owners; 17 people from Officers, merchant marine; 91 people from Fishermen and other seamen; 1288 people from Farmer-fishermen; 2163 people from Others; **Farmers**: 44 people from Craftsmen and artisans; 16 people from Officers, merchant marine; 582 people from Fishermen and other seamen; 176 people from Retirees;

16 people from Nomads; 30 people from Servants; 2461 people from Farmer-fishermen; 988 people from Others; **Merchants and shopkeepers**: 11 people from Civil servants; 70 people from Farmers; 86 people from Craftsmen and artisans; 691 people from Others; **Craftsmen and artisans**: 55 people from Cottagers; **Owners**: 12 people from Farmers; 124 people from Craftsmen and artisans; 153 people from Others; **Engineers**: 62 people from Others; **Clerks**: 24 people from Merchants and shopkeepers; 153 people from Craftsmen and artisans; 170 people from Others; **Students and graduates**: 22 people from Craftsmen and artisans; 96 people from Others; **Shipowners**: 169 people from Farmers; 15 people from Merchants and shopkeepers; 95 people from Craftsmen and artisans; 306 people from Others; **Officers, merchant marine**: 470 people from Farmers; 31 people from Merchants and shopkeepers; 318 people from Craftsmen and artisans; 15 people from Retirees; 19 people from Laborers and workers; 12 people from Coachmen; 412 people from Others; **Fishermen and other seamen**: 264 people from Farmers; 120 people from Craftsmen and artisans; 174 people from Cottagers; 185 people from Laborers and workers; 69 people from Coachmen; 43 people from Servants; 518 people from Farmer-fishermen; 150 people from Cottager-fishermen; 17 people from Paupers; 129 people from Others; **Cottagers**: 13 people from Farmers; 27 people from Fishermen and other seamen; 54 people from Laborers and workers; 2451 people from Cottager-fishermen; 21 people from Others; **Retirees**: 34 people from Others; **Laborers and workers**: 279 people from Farmers; 29 people from Merchants and shopkeepers; 1411 people from Craftsmen and artisans; 11 people from Clerks; 733 people from Fishermen and other seamen; 1133 people from Cottagers; 273 people from Coachmen; 56 people from Managers; 103 people from Servants; 30 people from Paupers; 515 people from Others; **Managers**: 25 people from Farmers; 41 people from Craftsmen and artisans; 20 people from Others; **Nomads**: 16 people from Fishermen and other seamen; 12 people from Retirees; 103 people from Others; **Others**: 23 people from Fishermen and other seamen; 13 people from Managers.

(Farmer-Fishermen and Cottager-Fishermen are separate HISCO codes but are recoded to Farmers, Cottagers or Fishermen through this procedure)

# E Construction of income estimates

## E.1 Number of income cells

Table A15 shows the number of income cells per occupation and income class.

Cottager and servant mean income is taken directly from the wage data ($L_D$ and $L_C$ in Table A15). Other working-class groups receive the "worker" wage ($L_B$). Individuals in high-skill occupations (a relatively small number in this income category) are given a markup $\xi$ on the low-skill wage, set at 1.2 ($L_A$).

The six income intervals used are (1) above 250 Spd, (2) 250-200 Spd, (3) 200-150 Spd, (4) 150-100 Spd, (5) 100 Spd and (6) below 100 Spd.

| Occupation group | Income group | | | | | |
|---|---|---|---|---|---|---|
| | **1**: >250 | **2**: 200-250 | **3**: 150-200 | **4**: 100-150 | **5**: 100 | **6**: <100 |
| Civil servants | 461 $^{P}_{A_1}$ | 218 $^{U}_{A}$ | 302 $^{U}_{A}$ | 369 $^{U}_{A}$ | 281 · | 237 $L_A$ |
| Farmers | 399 $^{P}_{A_1}$ | 389 $^{U}_{A}$ | 421 $^{U}_{A}$ | 428 $^{U}_{A}$ | 398 · | 431 $U$ |
| Merchants and shopkeepers | 323 $^{P}_{A_1}$ | 125 $^{U}_{A}$ | 160 $^{U}_{A}$ | 194 $^{U}_{A}$ | 159 · | 173 $L_A$ |
| Craftsmen and artisans | 121 $^{P}_{A_1}$ | 107 $^{U}_{A}$ | 165 $^{U}_{A}$ | 251 $^{U}_{A}$ | 291 · | 466 $L_B$ |
| Owners | 69 $^{P}_{A_1}$ | 6 $^{U}_{A}$ | 13 $^{U}_{A}$ | 16 $^{U}_{A}$ | 11 · | 161 $U$ |
| Engineers | 36 $^{P}_{A_1}$ | 6 $^{U}_{A}$ | 6 $^{U}_{A}$ | 5 $^{U}_{A}$ | 1 · | 25 $L_A$ |
| Clerks | 101 $^{P}_{A_1}$ | 68 $^{U}_{A}$ | 81 $^{U}_{A}$ | 109 $^{U}_{A}$ | 85 · | 222 $L_A$ |
| Students and graduates | 78 $^{P}_{A_1}$ | 15 $^{U}_{A}$ | 21 $^{U}_{A}$ | 16 $^{U}_{A}$ | 4 · | 73 $U$ |
| Ship owners | 92 $^{P}_{A_1}$ | 24 $^{U}_{A}$ | 31 $^{U}_{A}$ | 21 $^{U}_{A}$ | 16 · | 36 $U$ |
| Fishermen and other seamen | 68 $^{P}_{A_1}$ | 67 $^{U}_{A}$ | 118 $^{U}_{A}$ | 176 $^{U}_{A}$ | 190 · | 379 $L_B$ |
| Cottagers | 34 $^{P}_{A_1}$ | 34 $^{U}_{A}$ | 91 $^{U}_{A}$ | 197 $^{U}_{A}$ | 247 · | 422 $L_C$ |
| Retirees | 80 $^{P}_{A_1}$ | 62 $^{U}_{A}$ | 110 $^{U}_{A}$ | 197 $^{U}_{A}$ | 231 · | 444 $U$ |
| Laborers and workers | 103 $^{P}_{A_1}$ | 94 $^{U}_{A}$ | 165 $^{U}_{A}$ | 274 $^{U}_{A}$ | 321 · | 432 $L_B$ |
| Coachmen | 9 $^{P}_{A_1}$ | 8 $^{U}_{A}$ | 9 $^{U}_{A}$ | 12 $^{U}_{A}$ | 8 · | 194 $L_B$ |
| Managers | 34 $^{P}_{A_1}$ | 12 $^{U}_{A}$ | 16 $^{U}_{A}$ | 13 $^{U}_{A}$ | 14 · | 339 $L_A$ |
| Nomads | 4 $^{P}_{A_1}$ | 3 $^{U}_{A}$ | 4 $^{U}_{A}$ | 4 $^{U}_{A}$ | 5 · | 28 $U$ |
| Servants | | | | | | 470 $L_D$ |
| Paupers | | | | | | 460 $L_E$ |
| Others | 42 $^{P}_{A_1}$ | 18 $^{U}_{A}$ | 36 $^{U}_{A}$ | 32 $^{U}_{A}$ | 21 · | 333 $L_F$ |

Table A15: Number of income cells per occupation and income class. Letters in cells refer to interpolation methods. Where two letters are given, the upper refers to cells where no agricultural information is available. Key: $P$: Pareto distribution, $U$: Uniform distribution; $L$: Lognormal distribution; $A$: Imputed from agricultural information. For parameter values, see text. Total number of municipalities: 491

## E.2 Wages

For **cottagers and workers**, the average of winter and summer wages is taken to obtain an annual daily wage. The difference between wage with food and wage without food is used to obtain a food premium (to use on the servant wage later on). These average daily wages (without food) then have to be converted to annual wages. This conversion is challenging for three reasons. First, we do not know exactly how many days a year a fully employed person would work. Second, for daily workers, there is likely to have been some unemployment. Third, the cottagers spent substantial time working their own farm. Did this pay more or less than the waged labor? We can take some guidance from the literature. Grytten (2007) uses 313 days per year (six days per week) when converting yearly wages to daily wages for Norway from 1850 onward. Lindert & Williamson (2012) use 313 days as a "full-time" estimate, with robustness checks at 280 and 222 days. Abramitzky *et al.* (2012) uses 297 days. Allowing for some religious celebration and unemployment, this paper will use 300 working days per year, subject to robustness checks as explained below. For brevity, the number-of-days-worked constant will be denoted $\phi$.

For **servants**, the wage is given as an annual amount, but includes room and board. To add the value of room and board to the money amount given, we add the difference between average wages with and without board for cottagers and workers in the municipality.

For **other occupations**, there is no direct wage data. Note that for many of the occupations, there are relatively few individuals in the lower income category. For several "high-skill" labor groups, a twenty-percent markup on the worker wage is used. For farmers, nomads and the other social classes that mainly derive their income from non-labor sources, as well as those with insufficient occupation information, a uniform distribution between lowest wage and 100 Spd. is used. For the poor, half the lowest wage observation in the municipality is used.

**Missing data.** Some municipalities are missing some of the wage series. This is often because those occupations are not found in that municipality, in which case this is not a problem. In other cases, it is interpolated by using average differences between the occupations, or, when there is no wage information, by the average for the county.

**Within-group distribution.** The reported wage observation is assumed to reflect the mean of the income of all individuals of a given occupation in that municipality, and is used to construct a lognormal distribution within each municipality-occupation cell, with a common standard deviation $\zeta$. The discretization algorithm takes into account that individuals with incomes above 100 Spd are already placed and accounted for in the other income categories (with the exception of servants, who are explicitly excluded from the income data).

## E.3 Top incomes

### Inference from taxes

As stated in the main text, the source on incomes does not include information on the mean incomes of those with incomes above 250 Spd. It does, however, include information on the total amount of municipal income taxes paid by income group.

The information is used to impute incomes in the following manner. We consider the tax function to be

$$t_i = \tau_{0,j} + \tau_{1,j} y_i \tag{4}$$

where $t$ is taxes paid (per capita) and $y$ is income per capita, $j$ denotes municipality and $i$ denotes individual. This amounts to a constant marginal tax on income. If $\tau_0 = 0$, the average tax is also constant; $\tau < 0$ corresponds to the average tax being increasing in income. For brevity, municipality subscript will be omitted in the following.

For the four middle-income groups $y_2$ - $y_5$, we assume the mean incomes to be 225, 175, 125 and 100, respectively. The parameters $\tau_0$ and $\tau_1$ of the tax function can then be estimates by linear regression of $t$ on $y$. The inferred mean income of the richest group $y_1$ can then be inferred from the tax function as $\hat{y}_1 = -\frac{\hat{\tau}_0}{\hat{\tau}_1} + \frac{1}{\hat{\tau}_1} t_1$.

This assumption may not be correct for all municipalities. For this reason, $y_1$ is only estimated using this method in the case where the coefficient of determination $R^2$ is higher than 0.5 and where there are at least two points from which to estimate the line from.

### Imposing a distribution

The canonical distribution function for the upper income tail is the Pareto distribution. For a dispersion parameter $\alpha$ and a location parameter $b$, the distribution function is

$$F(y) = 1 - \left(\frac{b}{y}\right)^{\alpha}$$

The relationship between the mean income of the richest group $y_1$ and the dispersion parameter $b$ is

$$y_1 = b\frac{\alpha}{\alpha - 1}$$
$$b = y_1\frac{\alpha - 1}{\alpha}$$

$b$ is also the lower bound of the incomes of the richest group. To maintain consistency with the income tables, $b$ is not allowed to be less than 250; if $y_1\frac{\alpha-1}{\alpha} < 250$, $\alpha$ is set to $\frac{y_1}{y_1 - 250}$ giving $b = 250$.

### The dispersion parameter

The only (nationwide) top income tabulations available for Norway in the relevant period are from 1859 and 1876. The data for 1859 is available for urban areas only, while the 1876 data is available for urban and rural areas separately. This gives us three top-income tabulations on which to base our estimates. Each table consists of income intervals and the number of individuals in each of the intervals. Given an assumption on total population size, this gives us points on the cumulative income distribution for these years and sectors. Denote the income levels as $z_i$ and the number of people with incomes less than this level as $x_i$. We can then identify the Pareto shape parameter by defining $X_i = \log(1 - x_i)$ and $Z_i = \log(z_i)$ and running the regression[14]

$$X_i = \alpha_0 - \alpha Z_i$$

Figure Z shows the fit of the linear regression. For the 1859 data, the Pareto distribution is not a perfect fit, through the 95% confidence interval around the regression line is still quite narrow. For the two 1876 data sets the fit is good with the exception of the uppermost point (covering only a very small share of the population).

The Pareto distribution is only assumed valid in the upper tail of the distribution. We do not want to include the points of the entire distribution in our identification of $\alpha$. Moreover, the poorest group is very large in all three data sets, containing 89%, 78% and 54% of the population, respectively. Limiting our estimation at the upper 30% of the distribution and above, we obtain the results in Table A16. "Obs" refers to the number of points on the CDF used to estimate the dispersion parameter.

| Sample | 1859, urban | | 1876, rural | | 1876, urban | |
|---|---|---|---|---|---|---|
| Cutoff percentile | $\hat{\alpha}$ | Obs. | $\hat{\alpha}$ | Obs. | $\hat{\alpha}$ | Obs. |
| 0.70 | 1.756 | 10 | 1.795 | 32 | 1.386 | 29 |
| 0.80 | 1.756 | 10 | 1.797 | 31 | 1.389 | 28 |
| 0.90 | 1.808 | 9 | 1.792 | 29 | 1.464 | 23 |
| 0.95 | 1.834 | 8 | 1.790 | 27 | 1.499 | 20 |
| 0.99 | 2.053 | 5 | 1.846 | 21 | 1.579 | 12 |

Table A16: Estimates of the top-income dispersion parameter

As is evident from the table, $\alpha$ increases as we cut off the estimate higher up, because the outlier at the top is given more weight. The urban estimate for 1859 and the rural estimate for 1876 have similar values, while the urban 1876 estimate has lower values for $\alpha$ (higher dispersion). The population of the upper group in the 1868 data is often substantial, favoring the 0.70 estimates rather than the 0.90 or 0.95. A reasonable compromise between these estimates is

[14] Note that the Pareto CDF is $F(y) = 1 - (b/y)^{\alpha}$. Looking at the survival function $S = 1 - F$ we get $S(y) = (b/y)^{\alpha}$. Taking logs of this gives $\log S(y) = \alpha \log b - \alpha \log y$.

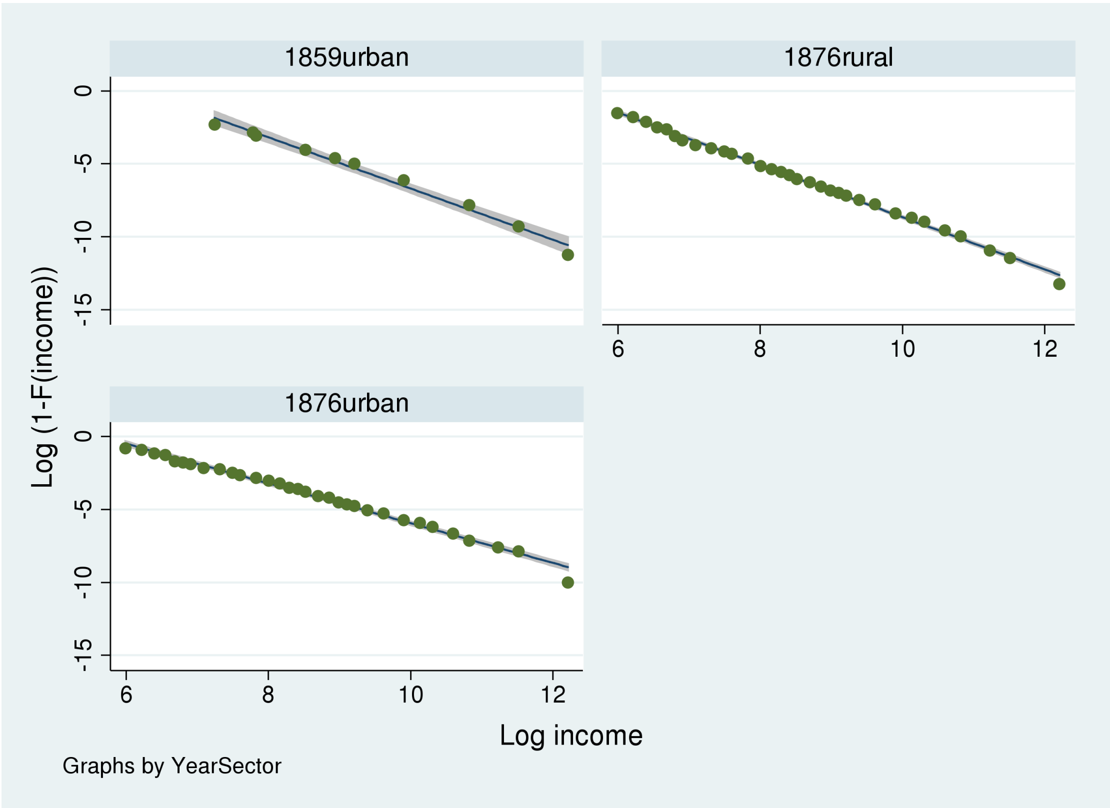


Figure A2: Top dispersions of 1859 and 1876, national data

$\alpha = 1.7$. This will be used in the following analysis.

Note that the within-group Gini coefficient among those with incomes above 250 Spd in a given municipality is

$$G_1 = \frac{1}{2\alpha - 1}$$

giving a within-group Gini of 0.4166 for those with incomes above 250 Spd.

**The estimated top income distributions**

The restrictions of $R^2 > 0.5$ and $\hat{y}_1 \geq 250$ yield estimates of the mean income of $y_1$ for 311 out of 488 municipalities with tax data. The median estimated mean income of the richest group is 528 Spd, with a mean of 748, a 10th percentile of 310 and a 90th percentile of 1387. With a shape parameter $\alpha$ of 1.7, distributions are compressed ($\alpha$ reduced) if $y_1$ is greater than 607. This case applies for 129 of the municipalities.

## E.4 Constructing discrete observations from imputed distributions

In theory, some parts the estimates constructed here are based on theoretical (imputed) distributions rather than on individual observed incomes. Gini coefficients and other metrics of inequality can be calculated directly from the distribution functions. However, for the purpose of constructing a "pseudo-cross-section" data set and calculating inequality metrics across municipalities and occupations, a discrete version of these distributions was used, where each individual was allocated a specific income. This gives only very minor departures from the continuous distributions.

The algorithm for the top incomes (Pareto distribution) is as follows:

- Set parameter $\alpha$ (dispersion) and $\mu$ (mean income) as stated in the main text
- Obtain lower bound $b = \mu \cdot (\alpha - 1)/\alpha$
  - If this gives $b < 250$, set $b = 250$ and adjust $\alpha$ to match
- Define the CDF $F(c) = b \cdot (1 - c)^{-1/\alpha}$
- For a population of size $N$, define a population vector

$$V = \left\{ \frac{1}{N} - z, \frac{2}{N} - z, \ldots, \frac{N-1}{N} - z, 1 - z \right\} \tag{5}$$

- Use bisection search to obtain a value for $z \in (0, 1/N)$ so that $mean(F(V)) = \mu$ (that is, so that the mean of the discrete distribution is the same as the mean of the continuous distribution)

- Use the incomes $F(V)$ for this particular municipality and occupation when calculating inequality.

For one observation, inequality is per definition zero. However, for populations larger than 1, the algorithm quickly yields a distribution with a Gini coefficient close to the theoretical value.

As an example, take a municipality with mean income $\mu = 800$ and a dispersion parameter $\alpha = 1.7$. The theoretical Gini coefficient is $1/(2\alpha - 1) = 0.417$. For a population of 2, the present algorithm obtains 0.227, for a population of 5, 0.347, for a population of 20, 0.401, and for a population of 100, 0.414.

# F Robustness

## F.1 Adjusting from men aged above 25 to households

This section outlines a simple conversion from men aged above 25 to household as population basis, for comparison to international estimates. These assumptions are only used in the discussion of those comparisons.

We make two assumptions on households with several men aged above 25:

- A: Men with high incomes live together; we form households starting with the highest-income men and work downward
- B: Men with low incomes live together; we form households starting with the lowest-income men and work upward

Assumption A is not applied to households with three or more men aged above 25. (These constitute around 13,000 households nationally); men in these large households are always assumed to be at the lower end of the distributions. For simplicity, incomes of other individuals at these households are not considered.

Furthermore, two assumptions are made on the households headed by individuals who are not men aged above 25:

- 1: Their incomes are similar to the upper tail of men aged above 25 in the municipality
- 2: Their incomes are similar to the lower tail of men aged above 25 in their municipality

This gives four Gini estimates:

- A1: $G = 0.657$
- A2: $G = 0.606$
- B1: $G = 0.597$
- B2: $G = 0.537$

Assumptions A and 1 are more radical than B and 2. For this reason one might consider putting more weight on B2 than on the others when comparing these estimates.

## F.2 Parameter adjustments

See Table A17. The adjustment of $\alpha$ refers to the imputed top dispersion where there is no data. The parameter $\phi$ (days worked) is used when converting daily wages to yearly wages. The parameter $\xi$ (skill premium) is used to impute high-skill labor incomes. The parameter $\zeta$ is the dispersion for the lognormal distribution used at the bottom of the income distributions.

| | $\alpha$ | $\phi$ | $\xi$ | $\zeta$ | Mean inc. | Gini |
|---|---|---|---|---|---|---|
| Reference | 1.7 | 300 | 1.2 | 20 | 179 | 0.546 |
| Higher dispersion in imputed top incomes | 1.5 | 300 | 1.2 | 20 | 179 | 0.547 |
| Lower dispersion in imputed top incomes | 2 | 300 | 1.2 | 20 | 178 | 0.544 |
| Fewer days worked | 1.7 | 280 | 1.2 | 20 | 179 | 0.553 |
| More days worked | 1.7 | 320 | 1.2 | 20 | 180 | 0.539 |
| Lower skill premium | 1.7 | 300 | 1 | 20 | 179 | 0.546 |
| Higher skill premium | 1.7 | 300 | 1.5 | 20 | 179 | 0.545 |
| Lower dispersion at bottom | 1.7 | 300 | 1.2 | 10 | 182 | 0.525 |
| Higher dispersion at bottom | 1.7 | 300 | 1.2 | 30 | 175 | 0.568 |

Table A17: Robustness checks: Alternative parameters

The top income shares described in Footnote 7 are not sensitive to the estimates of number of days worked, the skill premium or the dispersion at the bottom. However, as expected, they react substantially to the restrictions imposed towards the top. Recall that the baseline estimates of top income shares for the top 0.1, 1 and 10 per cent were 9, 25, and 54 per cent, respectively. For $\alpha = 1.5$ (high dispersion at the top), one obtains shares of 11, 27, and 54, while for $\alpha = 2$ (low dispersion at the top) one obtains 7, 23, and 54. For this reason, estimates of top income shares based on detailed tabulations of tax data at the top (such as those by Aaberge & Atkinson (2010) for 1875, with a different population definition, at 8, 18, and 40 per cent) should be preferred to the estimates from the present paper, which is mainly concerned about the full income distribution. As less than 10% of the population is in the top income group (incomes above 250 Spd), the top 10% income share estimate is not as sensitive to the top dispersion parameter $\alpha$ as the top 1% or top 0.1%.