International Inequalities Institute

WORKING PAPER 161

Historical social tables: advantages, methodology, and problems

Philipp Erfurth
University of Bologna

María Gómez León
Departament d'Anàlisi Econòmica, Universitat de València

Giacomo Gabbuti
Institute of Economics Scuola Superiore Sant'Anna Pisa

Branko Milanovic
Graduate Center City University of New York and International Inequalities Institute, LSE

FEBRUARY 2026

Philipp Erfurth
University of Bologna

María Gómez León
Departament d'Anàlisi Econòmica,
Universitat de València

Giacomo Gabbuti
Institute of Economics, Scuola Superiore
Sant'Anna Pisa

Branko Milanovic
Graduate Center City University of New York
and International Inequalities Institute, LSE

In addition to our working papers series all these publications are available to download free from our website: www.lse.ac.uk/III

International Inequalities Institute
The London School of Economics and
Political Science, Houghton Street,
London WC2A 2AE

E Inequalities.institute@lse.ac.uk
W www.lse.ac.uk/III
𝕏 @LSEInequalities

Historical social tables: advantages, methodology, and problems

Philipp Erfurth, María Gómez León, Giacomo Gabbuti and Branko Milanovic*

Abstract

This paper provides a methodological contribution to the study of historical income inequality by examining the construction and use of social tables for the nineteenth century. In a period when modern household surveys were absent, social tables represent one of the only feasible approaches for providing distributional evidence for the entire population. At the same time, existing studies rely on a wide range of assumptions, classifications, and data treatments, which makes comparisons across countries and over time difficult.

The paper reviews the main methodological challenges involved in constructing social tables, including class definitions, within-group inequality, units of analysis, and the external validation of income levels and subsistence benchmarks. Using simulations and historical examples, it shows how alternative methodological choices can generate substantial differences in inequality estimates. It finally proposes a set of guiding principles and template structures aimed at improving comparability, while still preserving the country-specific nature of historical evidence.

* University of Bologna; Departament d'Anàlisi Econòmica, Universitat de València; Institute of Economics Scuola Superiore Sant'Anna Pisa; Graduate Center City University of New York and International Inequalities Institute, LSE. The authors thank participants to the Inequality & History Workshop at Bocconi University for insightful discussion around these themes, which motivated and inspired parts of this work.

1 Introduction

Just like a map of Europe in the mid-19th century (we are looking at you – future Italy, Romania and Germany!), the study of inequality in the last two centuries is marked by one key feature: fragmentation.[1] While studies of historical inequality have proliferated in recent years[2], there is no comprehensive comparative research on income inequality for the period between the Napoleonic Wars and WWI. Most notably, this holds true for the 19th century: in a stark contrast with the relatively abundant, and growing, evidence on wealth inequality accumulated for the previous centuries on the basis of land cadaster and similar fiscal sources across many European regions. From the turn of the 19th century these old sources become no longer available or reliable, while modern income taxation and household surveys had to wait often for more than a century.[3]

Accordingly, many aspects of the distributional effects of seismic societal shifts in the 19th century, such as the industrial revolution, the rise of socialism and nationalism, the spread of European Imperialism and a large number of economic and political crises, such as the economic crises of 1845-1847 and subsequent upheavals of 1848, the impact of the Great War and related political developments, remain understudied.

[1] In this study, references to the 19th century are employed in a broad analytical sense and refer to the period between 1815 and 1914.
[2] See e.g. Alfani (2021), Alfani et al. (2025), Erfurth (2025), De Haas (2022), Milanovic (2024b), Moraes and Challú (2025).
[3] As a result, the first continuous series on Top 1% income shares based on actual data available from the World Inequality Database are those of few German states from the 1870s, Australia from 1910, the US from 1913, and France from 1915.

Even for the later periods, available evidence is mostly confined to fiscal data covering a few top percentiles of the distribution. In addition to the fact that these often problematic sources –resulting from the very first experimentation in personal income taxation and its uneven application– cannot be really checked against reliable external sources, they do not allow us to explore broader trends within the working class: the alleged disappearance of the middle classes, and the impact of economic and institutional changes within the labour markets, from the emergence of trade unions to the evolution of collective bargaining, as well as the evolution of gender divides in income and labour force participation. Apart for a very limited set of countries, this holds true until the 1950s and 1960s, when the pioneering household surveys became available, often conveying very unstable, and in some cases even paradoxical, results.[4]

This paper will seek to address this gap by proposing a broadly common methodology to study inequality in this (and earlier) time periods on the basis of social tables. Social tables as a tool to estimate inequality have a long history.[5] Social tables divide societies into different social groups across the entire distribution of income and calculate the inequality between these groups. They thus rely on two major sources of data – population data and income data – both of which are more widely available than alternatives, such as tax data. Social tables come with a number of additional advantages. First, they enable the estimation of inequality in contexts in which household survey or tax data are unavailable as social tables only require two main variables as inputs: income data and – at least

[4] The surviving evidence has been collected by Milanovic (2019) in the "All the Ginis" dataset.
[5] See William Petty (1690)

somewhat disaggregated - population data. Second, social tables are intuitive and "historical" as they are based on a class structure that reflects historical realities. They are thus "flexible". This approach is particularly well suited to historical periods in which within-group differences are generally smaller than in present-day societies. At the same time, even when top income shares are available, social tables make possible to investigate the impact of structural change across all the classes and occupational group (which top shares obviously cannot do), as well as within-labour differences, including gender gaps. Yet, despite these benefits, social tables also come with a number of caveats.

This study will explore these caveats with the aim of proposing consistent solutions to address them. While we will focus on the timeframe between 1815-1914, i.e. the era of the concert of Europe that ended with WW1, such a common methodology could be meaningfully applied to adjacent time periods as well to bridge large geographical and temporal gaps in the historical literature on economic inequality. As such, this study can provide an important contribution to the literature, which suffers from a high degree of methodological fragmentation, as noted at the outset. Indeed within this fragmented area of study, it may seem to the observer that the number of approaches to social tables is as large as the number of social tables themselves. This paper should not only provide a greater understanding of existing studies but also provide greater methodological clarity on future work in the area that would ideally lead to a significant advancement of research on income inequality in the last two centuries. Enabling such research through a common set of methodological approaches is a key aim of this research.

While this research seeks to arrive at a more universal approach to the methodology of social tables, the proposals should not be taken dogmatically. The circumstances and the data that allow the creation of tables are so varied that no single rule can be laid down. Nonetheless, we hope to provide a discussion of common problems that researchers encounter and propose solutions to address them. The next section will provide a (very) short overview of existing research and methodologies of 19th century social tables. Section 3 will explore methodological issues of social tables and propose solutions to mitigate them. Section 4 will explore the practical implications of constructing a social table using concrete examples. Section 5 will provide some common templates for 19th century social tables and offer concluding remarks.

2 Existing Studies

There is a body of peer-reviewed work using social tables to estimate inequality in the 19th century. Table 1 shows the geographical coverage of such tables. As is evident, this coverage is highly uneven; however, continents, such as Latin America are relatively well represented, particularly where other types of sources are not available or reliable for most of the period (Rodríguez Weber, 2023). By contrast Africa shows only a very limited number of social tables. At the time of writing, Ghana is the only country with a long-run coverage; there is only one social table available for the Maghreb region in 1880, while countries such as Botswana, Ivory Coast, Kenya, Senegal and Uganda become covered only later in the

20th century.[6] Still, surveys such as the one by Galli et al. (2023) highlight how social tables are the most common type of evidence in the continent, while other works on the area are based on sources such as censuses which could also be used to produce social tables. For European countries, recent research based on social tables revealed important trends which had been overlooked by focusing on top incomes only, as in the case of the diverging inequality dynamics in interwar Germany and UK (Gómez León and de Jong, 2019). Nikolic et al. (2024) showed the possibility to expand similar approaches to Eastern Europe as well and the possibility to link 19th century and modern evidence by means of census-based social tables seem appealing also for countries such as France. Indeed, even for relatively well-studied countries like France, Germany or Japan, we are not able to exactly date when the post-WW2 "Great leveling" began to affect the broad majority of workers outside the top percentile or decile. As a result, historical discussion on inequality focused on policies, such as high marginal taxes, that mattered to top incomes, but often failed to discuss the impact of transformative changes such as the welfare state, nationalization, and the like.

[6] See Aboagye and Bolt, 2021; Bigston (1986, 1987); Bolt and Hillbom (2016); de Haas, (2021), Tadei and Alfani (2019).

Table 1: ***Overview of currently available social tables (1815-1914)***

Year	Country/Region	Author, Year
1831	France	Morrison and Snyder, 2000
1841	England and Wales	Allen, 2018
1847	Bavaria	Erfurth, 2025
1850, 1860, 1870	USA	Lindert and Williamson, 2016
1860-1914	Chile	Rodriguez Weber, 2017
1863	Prussia	Erfurth, 2025
1866	France	Morrison and Snyder, 2000
1867	Great Britain	Lindert and Williamson, 1983
1870	Colombia	Nieto Ramos, 2025
1872	Brazil	Bértola et al., 2008; Gómez León, 2021
1875	Austria*	Erfurth, 2025
1876	Peru	Berry, 1990
1880	Indonesia	Booth, 1988; van Zanden, 2003
1880	Maghreb	Amin, 1966
1880	China	Chang,1962
1882	Bavaria	Erfurth, 2025
1886	Japan	Moriguchi and Saez, 2005
1889	Uruguay*	Marmissolle and Willebald, 2025
1891,1911	Ghana	Aboagye and Bolt, 2021
1895,1910	Mexico	Castañeda Garza and Bengtsson 2020
1895, 1914	Argentina*	Arroyo Abad and Maurer, 2025
1900-1914	Spain*	Gómez León, 2025
1901-1914	Italy	Gómez León and Gabbuti, 2025
1901-1914	Great Britain	Gómez León and de Jong, 2019
1904	Russia	Korchmina and Malinowski, 2024
1907-1914	Germany	Gómez León and de Jong, 2019
1910	Bulgaria and Czech Lands	Nikolic et al. 2024
1914	Kenya	Bigston, 1986 and 1987

* Unpublished

There have also been limited attempts to estimate historical inequality at the global level. The research by Bourguignon and Morrisson (2002), which covers the period 1820-1980, suffers from a lack of available country-level distributional data for the 19th century. In many cases they are no more than educated guesses. Indeed, rather than treating global inequality as inequality among countries only using Maddison per capita GDPs,

Bourguignon and Morrisson try to be more precise by creating 33 country groupings and treat global inequality as the sum of inequalities between and among these groupings. The groupings are assumed to contain countries that have similar internal income distributions that (moreover) evolve in the same way over time. While such an approach does not impact significantly global inequality levels which are in the 19th century driven principally by the divergence in mean country incomes, we clearly lack the full picture and have to use "heroic" or even hardly tenable assumptions about the evolution of within-country income distributions.

There have also been a number of additional papers that have built on the Bourguignon-Morrisson approach, including the work by Milanovic (2024b, 2011), van Zanden, Baten, Foldvari and van Leeuwen (2014), and Chancel and Piketty (2021). They have tried to remedy the lack of within-country income distribution by using limited number of social tables (Milanovic 2011), distribution of heights or unskilled wage/GDP as proxies for income distribution (van Zanden et al. 2014) or very limited fiscal data (Chancel and Piketty 2021). None of them can overcome the problem of fragmented and incomplete national income distribution data for the 19th century.

Findings in Milanovic (2024b) suggest that inequality increased during the course of the 19th century up until 1950, yet in the absence of reliable data for the 19th century, such findings suffer from heterogeneity in data quality between the two centuries. Future attempts to estimate inequality in the 19th century at the global level could build on social tables, particularly if elaborated using consistent methodologies as this study proposes.

3 Towards a consistent methodology

There are a number of studies exploring methodological issues relating to social tables, particularly the work by von Fintel, Links and Green (2023), which explores methodological challenges of using social tables in historical contexts. While underlining the opportunities presented by social tables, the authors highlight a number of caveats, including issues regarding within-group inequality, overlapping (by income) classes, the number of classes as well as the challenges relating to estimating the size of the top and bottom class. They issue a call for "greater cohesion and agreement in producing comparable estimates that can give a global view of historical inequality" (p.15) using social tables.

Achieving this objective is by no means easy, particularly in the context of data scarcity. The examples of the creation and work on social tables, as discussed in Section 4, show how that the process is complex and, very importantly, socially contingent, i.e. depending on the society, type of data that exist or can be collected, and importantly on the compiler (creator) of the social table. The prior historical and economic knowledge that should help "order" the table (e.g. what are the salient social classes) is normally based on some earlier research and familiarity with the economy that is being studied. While the inherent struggle between harmonization and context specificity will not be overcome, we believe that there are a number of guardrails and guiding principles that can ensure a higher degree of consistency among approaches, while preserving context specificity.

3.1 External validation

The first guiding principle is to seek external validation, where possible.

There are evident problems with estimations where the Gini, the mean of the distribution, and the subsistence minimum (*s*) are derived from the same source. While social tables provide both the mean and the Gini, they do not provide the subsistence minimum. The latter is, in turn, needed for the calculation of Inequality Extraction Ratio (IER), one of the main variables used in historical studies to estimate the degree of "extractiveness" of past societies. A researcher can, of course, assume that the poorest class's income is equal to physiological subsistence. But it is an arbitrary assumption: is (say) the French poorest class in 1831 living at the subsistence level, or twice the subsistence, or more? This simply means that we *cannot* "internally", that is with the data from most social tables alone, produce a credible estimate of subsistence, and consequently calculate the IER. The temptation is then to assume a subsistence that generates the "desirable" IER, rather than the reverse, to have IER be a derived variable.

An external validation of the mean, and thus of the mean/subsistence (m/s) ratio is indispensable. To explain: If the social table for Utopia in year X yields the mean of 10,000 pesos, and from an independent database (such as Maddison's) we have an estimate of Utopia's GDP per capita in year X of $PPP 800 (and decide to use as the subsistence $PPP 400 per capita per year), we immediately know that Utopia's m/s ratio is 2 (=800/400) and that Utopia's 10,000 pesos from year X are worth $PPP 800. Likewise, we also know that Utopia's subsistence expressed in pesos should be around 5,000 which provides a check on the reasonableness of the numbers available in the social table. To conclude: the use of an externally and independently produced mean in PPP terms does not only serve the

verification purposes. It is indispensable to produce IERs for most social tables which cannot generate this internally because they lack explicit estimate of the subsistence.

Both the mean and subsistence can be externally validated. As already mentioned, the mean can be validated from a source such as the Maddison Database. The subsistence minimum can be estimated either from a separate calculation based on Bob Allen's methodology (2001), i.e. through the calculation of the cost of a bare-bones basket of products needed for subsistence, or, where available, directly from historical sources, such as for the case of Prussia (discussed in Section 4) for which subsistence levels in local currency are available from Prussian county statistics. While it may not be always possible to draw on the benefits of external validation, it should be considered a key part of the construction of a social table, or at the very least as one form of robustness checks.

In the Italian case, for which other income inequality estimates were available, Gómez León and Gabbuti (2025) relied on these estimates to decide their baseline series. At the same time, they checked whether – despite the obvious conceptual differences – the labour incomes reconstructed in their social tables were meaningfully compared, when expressed as a share of value added, with the labour shares estimated by means of independent sources by Gabbuti (2021).

3.2 Within-group inequality

One challenge that exists by definition when using a social table in (almost all) cases, is that we do not have the necessary information or data to explore the extent of within-group inequality (which is usually the very reason why a social table approach is considered).

Note that if social tables consisted of all households or individuals existent in a country, it would approach administrative data; or if it consisted of all households from an income survey, it would approach surveys. So, the limit cases of social tables are household surveys (or even full country-wide administrative data). This is not achievable with the available data for the 19th (and indeed any previous) century.

Thus, inequality estimates based on social tables are, by construction, biased downward, as not all dimensions of inequality can be captured with this approach. The key question is therefore how big this omission is. In total, such within-group inequality cannot or should not be greater than the between-group inequality that we retrieve from the table—because if it were, the compiler of the social table will have likely not chosen the salient classes for the given country and time period. In studies based on surveys, we generally observe the highest variability (within income percentiles) in the very low and very high percentiles (i.e. by the very poor and very rich), while the percentiles in the middle of the distribution usually have trivially low within-Ginis (generally, under 0.05). Fintel, Links and Green (2023), use bivariate and multivariate statistics in a set of over hundred social tables, but do not find that introducing more classes leads to an automatic reduction in the downward bias related to the lack of consideration of within-class inequality.[7] (In other words, greater number of social classes does not always reduce the bias.) It could also be argued that in a setting where the focus of research is on differences between social classes, as might be the case for many studies for the 19th century and the early 20th century (i.e. studies where

[7] P. 15, von Fintel, Links and Green

differences in incomes or wages *within* a social group is not a main focus of the study), within-class inequality may not add a lot to the interpretation of inequality from a historical perspective and authors may wish to consider between-class inequality in its own right. In any case, it should be considered that any ad hoc adjustment to keep track of within-group inequality which might be possible in a given case – as for instance, by means of fiscal information on the top incomes – while reducing the bias of the social table, is likely to introduce inconsistencies with other estimates based on social tables.

3.3 The size of the top and the bottom classes

Another concern connected to the within-group bias discussed above is the number of classes used in social tables. As noted by von Fintel, Links and Green (2023, p.6), while the number of classes is not, in itself, a direct source of bias (see also Milanovic 2018, Modalsli, 2015), classifications with too few categories can distort the relative size of the resulting groups. The most common problems are an excessively broad top category or an overly large bottom one. Indeed, Fintel, Links and Green (2003) observe potential downward bias to calculated overall inequality from a lack of disaggregation of bottom classes. Below, we use simulations based on social tables to test the sensitivity of the Gini to alternative levels of aggregation and to assess the potential effects of blending together classes at both ends of the distribution.

The simulations are based on constructed social tables for the cases of Germany (1907), Great Britain (1901), Italy (1901), and Spain (1900).[8] In these social tables occupational

[8] Social tables for Germany and Great Britain are from Gómez León and de Jong (2019); for Italy from Gómez León and Gabbuti (2025); and for Spain from Gómez León (2024).

data were compiled from population censuses which report the number of individuals according to their declared main occupation. For the sake of comparability, occupations have been harmonized into 22 common categories. Yet, because the original censuses differed in the extent to which they distinguish occupational status (e.g. skilled, unskilled, apprentices) and gender status (male, female) within these categories, the final number of classes in each social table varied between 60 (Italy), 78 (Germany and Great Britain), and 107 (Spain). Moreover, we have further collapsed these 22 occupational classes into 10 broad categories as presented in Table 2, so to assess how the degree of aggregation affects the observed Gini and, importantly, the comparability of cross-country results.

Table 2: 22- and 10- occupational classifications derived from population censuses

22-class scheme		10 -class scheme
1. Industrial and commercial proprietors	*12. Non-metallic building materials*	*1. Elite proprietors*
2. Agricultural proprietors	*13. Textile Fabrics, Skins, Leather*	*2. High Professions / State / Church*
3. Administration, military	*14. Woodworking, furniture.*	*3. Urban Middle-Class Professions*
4. Liberal Professions	*15. Construction*	*4. Transport and Commerce*
5. Finance, Insurance, banking	*16. Food and beverage industries*	*5. Heavy Industrial Workers*
6. Transport and logistics	*17. Clothing industries*	*6. Light Manufacturing Workers*
7. Commerce and trade	*18. Paper, printing, and graphic arts*	*7. Construction Workers*
8. Mining	*19. Textile industries*	*8. Agriculture and Fisheries Workers*
9. Quarrying and energy production	*20. Agriculture and fisheries*	*9. Personal / Domestic Services*
10. Metalworking and machinery	*21. Personal /Domestic services*	*10. Unoccupied*
11. Chemical industries	*22. Without occupation*	

Table 3 shows Gini coefficients obtained using three alternative levels of occupational disaggregation. The table also presents the share of the population located in low-income occupations (i.e. wage workers in agriculture, domestic services, and individuals without an occupation) and in high-income occupations (i.e. elite proprietors).

Table 3. Gini results under alternative levels of occupational disaggregation for four European countries, circa 1900

Country	Gini (60-107 classes)	Gini (22 classes)	Gini (10 classes A)	Gini (10 classes B)	Share of low-income occupations	Share of high-income occupations
Germany (1907)	0.42	0.39	0.38	0.39	0.39	0.15
Great Britain (1901)	0.43	0.41	0.40	0.40	0.39	0.08
Italy (1901)	0.50	0.44	0.40	0.42	0.51	0.18
Spain (1900)	0.47	0.42	0.35	0.41	0.50	0.23

Sources: Data for Germany and Great Britain are from Gómez León and de Jong (2019); for Italy from Gómez León and Gabbuti (2025); and for Spain from Gómez León (2024).
Note: Gini (10 classes A) and Gini (10 classes B) differ in the segment of the distribution that is compressed during aggregation. Version A collapses classes at the top of the distribution (i.e. proprietors), whereas Version B collapses classes at the bottom (i.e. waged workers in agriculture, domestic services, and individuals without an occupation).

We observe that when the number of groups is reduced from the maximum level of disaggregation to 22, the loss of information remains limited: the observed Gini falls by 3–6 points, and the inequality ranking is largely preserved. Compressing the distribution into only 10 groups, however, generates substantially larger declines (up to 12 Gini points) with the magnitude of this bias varying across countries. The distortions are most pronounced in countries whose occupational structures are more polarized, that is, where large segments of the population are concentrated both in low- and high-mean-income groups. This is the case of Italy and Spain, where about 50 per cent of the population belongs to low-income occupations and around 20 per cent to high-income occupations. In such settings, aggregation eliminates many of the income contrasts that drive the between-component of inequality, producing the largest reductions in the observed Gini. By contrast, in Germany and Great Britain, whose social structures contain a higher proportion of middle-income groups, the same degree of compression entails much smaller losses.

The impact of agglomeration is sufficiently large to alter international rankings. Spain, for instance, moves from being the second most unequal country (with 107 groups) to appearing as the most equal when only 10 groups are used; likewise, Great Britain appears as unequal as Italy, despite the opposite (i.e., Italy the most unequal) being true when using the most detailed table. In this sense, reducing the number of groups systematically biases comparisons in favor of countries with more polarized occupational structures, as aggregation artificially compresses internal income differences and makes them appear less unequal than they actually are.

The main source of compression in the above cases lies at the top of the distribution, where the aggregation process collapses all proprietors into a single class (10 classes A). This effectively merges agricultural proprietors with industrial and commercial ones, despite the former having substantially lower mean incomes and much larger population shares in Italy and Spain (17 per cent and 16 per cent, respectively), thereby biasing downward the resulting Gini.

But what happens when compression is instead forced at the bottom of the distribution? Gini (10 classes B) provides exactly this counterfactual, by collapsing the three lowest-income groups (agricultural workers, domestic servants and the unoccupied). In Germany and Great Britain, the resulting Gini remains very close to the value obtained when aggregating at the top—showing only a modest departure from the fully disaggregated benchmark. In Italy and Spain, the impact is still sizeable, though the decline in inequality is smaller than under top-end aggregation (10 classes A).

Overall, these results reinforce and nuance recent studies addressing the potential bias of inequality measures derived from social tables based on a small number of classes. As argued by von Fintel, Links and Green (2023), the downward bias is strong where many individuals cluster in low-income occupations; yet we find that aggregation at the top is not trivial either or may even have a greater "inequality-squeezing" effect, echoing concerns about the treatment of elite groups (Alvaredo, 2011). These findings indicates that both top- and bottom-end aggregation can introduce substantial bias, highlighting the importance of detailed class structure for cross-country comparability. In this sense, combining population censuses with other sources (such as agricultural censuses or fiscal data)

could help improve the level of detail. The challenge remains, however, as discussed in the next section, that these sources rely on different units of analysis.

3.4 Units of analysis or type of social table

Unlike household surveys, which compile information at the individual or household level, or fiscal records, where the unit of analysis is the individual or, in some cases, married couples, social tables are constructed around social groups or classes, which constitute the unit of analysis. Different criteria can be used to define these groups. One such criterion is individual occupation (or employment), in which case the social table is built from information on the occupational structure and associated incomes. Here, we ignore how households are formed (who is partnering with whom?), which is an important omission, and we also exclude most of capital income which underestimates the top of the income distribution. For this reason, social tables are in some cases complemented with information from fiscal records, which are particularly useful for capturing top income earners, such as top property owners, officials, civil servants and related categories.

Another possible criterion for defining groups is geographical. Such an approach may be used when mean incomes for many settlements (villages, towns, larger towns) are available. Gini is by definition zero within the geographical unit, but if units are many and diverse, the overall inequality may not be severely underestimated. In practice, here the place of social class is taken by a geographical unit. In both these cases, although more so in the case of occupational tables, income from capital is likely to be underestimated which imparts a downward bias to inequality measures.

3.5 Common considerations when using occupational social tables

A natural starting point is the use of population censuses as the primary source whenever possible. Using information on the occupational structure extracted from population censuses typically allows us to classify between 50 and 60 per cent of the total population, including both proprietors at the top and the unoccupied at the bottom, thus avoiding the representativeness issues that arise when relying solely on other sources, such as tax records.

Notably, compared with sector-specific industrial censuses, which might be more accurate in the representation of a subset of workers, population censuses offer much broader coverage, as they enumerate individuals across all sectors of the economy and often distinguish between active and inactive people. This is particularly important because proprietors are sometimes reported among the inactive as “persons living from their rents.” A further advantage is that population censuses record respondents’ main activity over the preceding year, regardless of their employment status at the moment of enumeration. Industrial censuses, by contrast, provide a snapshot tied to a specific survey week or day, which tends to undercount employment in activities characterized by high turnover or seasonality.

At the same time, as noted in Section 3.3, combining population censuses with complementary sources such as fiscal data or sector-specific censuses is particularly useful for achieving a higher level of group disaggregation. Tax records can supplement information on the upper and middle segments of the distribution; agricultural censuses

help distinguish between different forms of landholding and their associated incomes; and industrial censuses are highly valuable for reconstructing wage differentials across occupational categories and by gender in years close to the population census.

Second, it is essential to address compositional biases that may arise in occupational data. Although occupational censuses can be broadly comparable across countries in their sectoral structure, they often differ in ways that require adjustment to achieve cross-country consistency. Some censuses systematically under-record paid work performed by women, frequently categorizing them as housewives, resulting in an inflated share of individuals listed as "without profession."[9] Similar issues arise with day laborers, whose seasonal or multiple occupations often lead enumerators to classify them as having no occupation.[10] Conversely, other censuses tend to overstate the active population, when family helpers assisting in farms, shops, or cafés are recorded as workers even in the absence of a labor contract or regular wages.[11] To ensure comparability with other countries, these cases should be documented and, if possible, adjusted. This may involve reallocating day laborers to low-income occupational groups or subtracting individuals who may rely primarily on a family wage (such as married housewives, students, or unpaid family assistants), thereby minimizing the risk of double counting and avoiding artificial clustering at the bottom of the distribution.

[9] A notable example is Great Britain in the early 20th century.
[10] This pattern can be seen in the treatment of *jornaleros* in Spanish censuses.
[11] Germany in the early 20th century provides an illustrative case.

Finally, country-specific socioeconomic conditions must be carefully taken into account when identifying salient occupational groups. National censuses often provide initial indications of where such attention is needed. When a given occupational category represents a large share of the labor force, this requires obtaining the most accurate income estimates available for that group and, where appropriate, introducing further subdivisions. Such refinements help prevent the aggregation of individuals with heterogeneous income levels into a single category, thereby limiting the types of biases discussed in Section 3.3. An example of this can be found in the case of Italy (1901), where, as discussed in Section 4.2, the prominence of the self-employed and small proprietors made it necessary to assemble information at a more disaggregated level and to draw on fiscal data to estimate their incomes accurately and to distinguish them from wage earners and large property owners.

3.6 Sub-national data

While social tables are generally easier to make for single cities, for which the data are often more easily accessible, it needs to be acknowledged that such an approach often comes with two caveats. First, the chosen city's inequality may be unrepresentative of the country as a whole. In most countries in the 19th century, there was a large difference between urban settings and rural areas. In such cases, a social table constructed on the basis of a city may not be a representative table for the country as a whole, which is also a challenge of using other local data, such as fiscal data, to estimate inequality at the national level. Second, the objective of studying inequality is inextricably linked with a study of alternative political or social organizations. However, in principle, the power of

cities (unless they are independent or self-governing) to make their own political and economic decisions is limited. Since our objective is to study how inequality and politics interact, the proper unit of analysis is state, I.e., the unit which has the monopoly of coercion, taxation, money creation, and most importantly, policymaking. Efforts should therefore be made to use city data as a complement to nationally available data to make a table for the country as a whole. Moreover subnational data could be "re-aggregated" as outlined in greater detail in the subsequent section for the example of Prussia.

On the other hand, as recently discussed by Gabbuti and Rappa (2025) for the case of Italy, where regional divides are arguably an important component of overall income distribution, census-based social tables could also be broken down at sub-national level, as these sources often come with regional and provincial disaggregation, and the wage data adopted to estimate regional divides could be usefully employed for the estimation of subnational social tables.

4 Constructing a social table in the 19th century: Practical implications

As noted at the outset, to construct a social table for any time period, context- and period-specific knowledge of the structure of the society of interest is critical. For instance, to construct a social table for 19th century France, the creator of a social table needs to consider the class structure of 19th century France and explore key questions related to its society. First, what broad societal groups existed within this society? This may include categories such as artisans, urban workers and subsistence farmers. For different time periods and societies, the definition of these groups, their size as well as the number of

groups within society might differ dramatically. For a society in the 19th century with a more "sophisticated" and differentiated economic system, more social groups are likely to be identified than for Ancient Greece or the Aztec empire. Despite these difficulties, it is nonetheless advisable, especially when working on several social tables for the same time period across similar societies, say Italy and Germany, to focus on a set of broadly comparable social classes across both countries. We attempt to do so in the succeeding section, by identifying some broad social classes that could be observed across similar countries. As we will argue in greater detail below, such an exercise must not be seen dogmatically, but as a guide that can enhance comparability.

The second question is whether the identified classes can be sufficiently distinguished from each other. Where does one class begin and where does the previous one end? Particularly in more professionally diversified societies such as in the 19th century this question is much harder to answer. Given the assumption that within class inequality is ignored, it is critical to ensure that the broad societal classes represent groups of individuals and households with comparable incomes so as to minimize underestimation of inequality. This might be particularly difficult in settings with high inequality between regions. For instance in Prussia, a day laborer in the West of the Kingdom in Bielefeld would earn around 120 Taler[12], while a comparable wage for a day's unskilled work in the district of Habelschwerdt in Silesia in the East of the Kingdom would only be around 65 Taler[13]. Such geographical differences are effectively eliminated when putting together a social

[12] Statistische Darstellung des Kreises Bielefeld.
[13] Statistische Darstellung des Kreises Habelschwerdt.

table for the Kingdom of Prussia as a whole. Where such data is available, regional differences could be considered by distinguishing between the same classes in different regions. While in practice such regionally disaggregated data is often unavailable, authors should consider that geographical inequalities are going to be masked by "national" social tables. Where data availability is limited to one region or city, as stressed above, it should be carefully considered whether the given region or city is representative of the country as a whole.

The third question to address at the outset is by no means less important than the two preceding ones: What data is available to construct a social table and is it representative of the country as a whole? We will discuss the type of data that is required in greater detail in the next subsection, but it is critical to consider that the answers to the above two questions, namely, what social groups existed within the society and how they can be distinguished, will be heavily conditioned by the available data. Luckily, for the 19th century such data is more abundantly available than for earlier time periods such as antiquity, where in practice the answers to questions 1 and 2 are often almost entirely driven by data availability considerations.

4.1 Constructing a social table for Prussia, 1863

Once the author has pondered the three above questions, in practice, data for two variables need to be compiled to construct a social table. The first is the population share of the respective social class within a society. For instance, the share of factory workers in Prussia in 1863 was approximately 9.8 per cent and the share of teachers in Bavaria in 1847 was 0.7 per cent. For the 19th century, this exercise is facilitated in many countries and

regions by the availability of professional censuses that. by the mid to late 19th century, many countries instituted at around a once a decade frequency. Such censuses, for instance the one by the Royal Prussian Statistical Office, would often, in addition to data on the number of inhabitants of regions and cities, also include a professional census that lists the number of individuals working in specific professions. These professional censuses are often aligned with the social groups that one may have identified previously, although they may in some cases be overly granular and may need to be consolidated, such as tailors and shoemakers that could be consolidated within broader categories of artisans.

While these professional censuses are extremely helpful, they also come with some caveats and limits in coverage. First, they may underestimate the number of people at the bottom of the distribution, such as day laborers and the urban poor that may not follow well-established professions on a consistent basis. They may also systematically underreport on the work by women. Professional censuses are also prone to undercounting those at the top of the distribution, such as those earning their income from non-labor sources, including so-called rentiers and large parts of the nobility. Professional censuses may also ignore "white-collar" professions with small numbers of individuals and large capital incomes, such as industrialists and bankers. Friedrich Alfred Krupp, Prussia's most eminent steel magnate, for instance, which the *New York Times* reported on in 1887 as the richest man in Prussia[14], falls through the cracks if using professional censuses.

[14] New York Times of February 7, 1887.

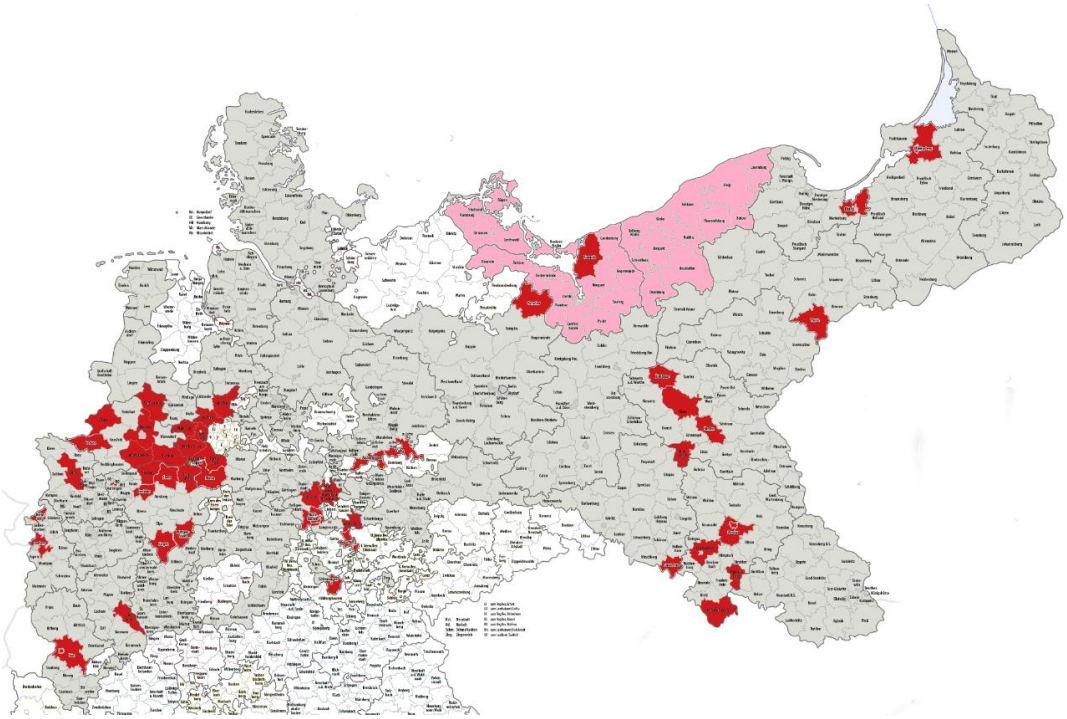
Professional censuses might also exclude categories of civil servants, which however are often available from other sources. The same is true for military hierarchies and the share of individuals of varying ranks within the military, which often can be sourced from military sources.

The second variable for which data needs to be identified is income. Unfortunately, compared to population share data, such data was in the large majority of cases not gathered through "official" channels such as was the case for the professional censuses. In most cases, income data must thus be collected from a variety of sources. In practice, the availability of data for incomes of different classes will therefore condition the choices made around the number and size of social classes. Given that income data were often not collected at the national level, it may also require the use of data from other geographical units, such as subnational units.

One such case is the Kingdom of Prussia (see Erfurth, 2024). For the Kingdom of Prussia in the 1860s no comprehensive data is available for incomes of social classes at a national level. There were, however, efforts at the subnational, specifically at the county level to gather data on incomes of the "working classes". By ministerial mandate from Berlin, all counties in the Kingdom were asked to produce comprehensive statistical compendia of their county, called the Statistical Representations, including issues such as physiographic conditions, inhabitants, migratory flows, marriage and birth statistics, real estate and property statistics as well as income statistics of working individuals. Figure 1 below shows the coverage of available Statistical Representations for Prussia. Using population weighted averages, mean incomes of several classes can be calculated for the Kingdom as

a whole. Other income data is drawn from additional sources, including for the military hierarchy and civil service (see Erfurth, 2024).

Figure 1: Coverage of available Statistical Representations for Prussia around 1860



Source: Erfurth (2024), Explanation: red indicates available counties, light red indicates available region.

On the basis of these data for population and incomes, a social table can be constructed. The table for Prussia in 1865 presented in Erfurth (2024) includes 65 distinct social classes ranging from the unemployed to the General Field Marshall, with the latter earning around 170 times mean income. As external validation of the presented results, the calculation of the inequality extraction ratio for Prussia, which Erfurth (2024) finds to be 44.2 per cent, is based on a value of *s* that is derived directly from historical sources. Indeed, a measure of

the subsistence minimum for a household of 4 or 5 individuals[15], including two adults and two to three children, is available for a large number of counties.

4.2 Constructing a social table for Italy 1901

The case of Italy in 1901, presented in Gómez León and Gabbuti (2025), relies primarily on population census data, complemented with sector-specific studies and fiscal evidence.

The occupational structure is derived from the 1901 population census, which enumerated the resident population of the Kingdom of Italy.[16] The census provides detailed information on the number of individuals employed in each sector, classified by occupation, employment status, and gender. The level of occupational detail in a social table is constrained by the availability of income data. While the census distinguishes between fine subcategories within given sectors, limitations in the wage evidence require some of these to be grouped together. Accordingly, data collection begins at the maximum feasible level of disaggregation and proceeds by aggregation only insofar as permitted by the available income sources.[17]

Secondly, known limitations of the census data are mitigated using adjustments suggested in the Italian historical literature. For instance, in line with Vitali (1968) and subsequent studies documenting the underestimation of women's agricultural employment (e.g.

[15] Erfurth (2024) uses equivalence scales to adjust for the extra child included in the measure for a small subset of counties.
[16] *Censimento della popolazione del Regno d'Italia al 10 febbraio 1901*.
[17] For instance, within *Metallurgical industries*, the Italian census distinguishes between different branches, providing details on how many workers are employed in *mechanical metallurgy*, *vehicle construction*, and *precision and luxury production*. However, limitations in the availability of wage data required these categories to be treated jointly.

Patriarca 1988; Mancini 2018), the original census figures are corrected by equating the number of women employed in family-run farms to that of men in comparable positions. Moreover, to avoid double counting individuals effectively living on a family income, housewives and students are excluded from the sample.

Finally, to enhance temporal and cross-case comparability, occupations are harmonized into 18 categories.[18] Work categories are then reclassified by sector: three categories in agriculture (owners, self-employed, and wage earners); one for owners in industry, commerce, and transport; three in industry (self-employed, salaried employees, and wage earners); three in commerce and transport (self-employed, salaried employees, and wage earners); two in public administration and services (salaried employees and wage earners); one for liberal professions; and one for the unoccupied. All categories were disaggregated by gender, resulting in a total of 60 social classes.

Once the occupational structure is established, nominal annual incomes are assigned to each occupation, employment status, and gender category. Income data for wage and salary earners rely on a combination of sector-specific studies, most notably Zamagni (1984, 1995), Rey and Vitali (1991), and official statistics (Istat, 1953). In the absence of equally detailed evidence on female wages, women's earnings are estimated using gender wage ratios derived from both secondary sources (Bettio, 1988; Felice, 2005) and primary sources, particularly the *Annuario Statistico Italiano*. These sources allow the construction

[18] Gómez León and Gabbuti (2025) present social tables for Italy covering 1901–1950, using a classification broadly comparable to that applied by Gómez León and de Jong (2019) to Germany and Britain for the same period.

of sector-specific gender ratios for agriculture, industry (distinguishing between heavy and light industries), transport, commerce, public administration, liberal professions, and services.

Information on the incomes of the self-employed is considerably scarcer and requires the use of heterogeneous sources and simplifying assumptions. In agriculture, incomes for self-employed farmers are reconstructed using daily wage data from the *Italy's Statistical Abstract*, combined with assumptions on annual working days from Giordano and Zollino (2015). For industry, services, and transport, we rely on fiscal data from the *Imposta di ricchezza mobile (MEF, 1901-1904)*, specifically Schedule B ("mixed incomes"), which mainly captures the earnings of self-employed individuals and family businesses.[19] Although imperfect, these fiscal sources provide a plausible proxy for self-employed incomes once adjusted for tax exemptions and evasion.[20]

Finally, in the absence of direct evidence on owners' incomes, these are obtained as the residual value added (VA), obtained after subtracting all labor and self-employment incomes, divided by the number of owners.[21] Residuals are computed separately for agriculture and for industry and private services. To avoid mechanically overstating

[19] This was the main direct income tax of the period and, although not formally personal, most private declarations referred to single individuals, making them a reliable proxy for incomes in these categories.
[20] As discussed in Gómez León and Gabbuti (2025, Appendix, p. 5), a conservative option is to impute self-employed incomes slightly above those of wage earners in the same sectors, as is common in labor-share estimates (Bengtsson and Waldenström, 2018). This approach, however, does not allow for different trends between dependent and self-employed workers – and by construction, renders the resulting social table unable to account for some of the most interesting historical political economy aspect of 20th century Italy; by suppressing heterogeneity within the self-employed, this might also lead to a downward bias in inequality levels.
[21] On this approach, see also Arroyo Abad and Astorga Junquera (2017).

inequality, only 80 per cent of this residual was attributed to owners. Fiscal tabulations from the *Imposta di ricchezza mobile* (MEF, 1901-1904) provide an external validation, yielding income levels broadly consistent with the residual-based estimates. While residual methods are inherently imperfect, particularly in periods affected by shocks, this approach aligned owners' incomes with observed capital shares (Gabbuti, 2021), and offered a coherent upper-bound estimate of inequality given the current state of historical evidence.

5 Concluding remarks: Towards a common "template"

As we have stressed throughout this study, considering the century from 1815 to 1914, it is difficult to establish a single social-table template, since population census classification systems evolved and did so at different paces across countries. In general, early nineteenth-century censuses typically grouped individuals by profession (e.g., notary, carpenter, barber, artisan) and did not clearly distinguish labor status, gender, or other relevant dimensions. As the century advanced (with industrialization, urbanization, and growing labor specialization) the systems gradually shifted toward grouping individuals by occupational sector and activity, and toward differentiating workers by employment status (e.g. employer, self-employed, salaried employee, wage earner, family assistant). The specific socio-economic features also meant that national censuses placed greater emphasis on some sectors than on others. Below, we present two templates as illustrative examples for Europe (c. 1900) and Latin America (c. 1870). These templates are organized according to their respective census structures.

Finally, we include a template that seeks to harmonize different censuses into categories ordered by income level, which are, in turn, based on the skill-based classification of HISCLASS.[22] While these templates should not be considered dogmatically, they could provide a guide to future studies using social tables in the 19th century, especially those aimed at providing comparative evidence for a wider discussions of inequality that transcends any one country.

[22] It should be noted that the income ordering of certain categories may vary across country-specific contexts. This is especially true for rural self-employed workers in settings where self-employment served as a subsistence strategy in response to unemployment.

Template 1. European social table structure based on 1900-1907 censuses

Sector/Occupation	Work Category	Gender
1. Agriculture and Fishing	Owners/Employers	Male/female
	Self-employed	Male/female
	Waged workers	Male/female
2. Industry	Owners/Employers	Male/female
	Self-employed	Male/female
2.1 Extractive Industries (Mining, quarrying, extraction of raw materials)	Salaried/white collar workers	Male/female
	Waged/blue collar workers	Male/female
2.2 Chemicals (Basic chemicals, oils, related chemical processing)	Salaried/white collar workers	Male/female
	Waged/blue collar workers	Male/female
2.3 Other Basic Materials (cement, glass, ceramics, other non-chemical processing)	Salaried/white collar workers	Male/female
	Waged/blue collar workers	Male/female
2.4 Metalworking and Machinery (Metals, engineering, mechanical equipment, vehicles)	Salaried/white collar workers	Male/female
	Waged/blue collar workers	Male/female
2.5 Wood, Furniture (Wood products, furniture)	Salaried/white collar workers	Male/female
	Waged/blue collar workers	Male/female
2.6 Paper and Printing (Paper production, publishing, printing)	Salaried/white collar workers	Male/female
	Waged/blue collar workers	Male/female
2.7 Construction	Salaried/white collar workers	Male/female
	Waged/blue collar workers	Male/female
2.8 Textiles and Apparel (textiles, clothing, leather)	Salaried/white collar workers	Male/female
	Waged/blue collar workers	Male/female
2.9 Food and Beverage Manufacturing (Food processing, beverages, tobacco)	Salaried/white collar workers	Male/female
	Waged/blue collar workers	Male/female
3. Commerce and Services	Owners/Employers	Male/female
	Self-employed/Artisans	Male/female
3.1 Transport	Salaried/white collar workers	Male/female
	Waged/blue collar workers	Male/female
3.3 Finance, Insurance, Banking	Salaried/white collar workers	Male/female
	Waged/blue collar workers	Male/female
3. 4. Public Administration	Salaried/white collar workers	Male/female
	Waged/blue collar workers	Male/female
3.5. Liberal Professions	Salaried/white collar workers	Male/female
3.6. Personal Services	Waged/blue collar workers	Male/female
3.7. Domestic Services	Waged/blue collar workers	Male/female
4. Without profession		Male/female
4.1 Rentiers	Owners	Male/female
4.2 Unoccupied (other than students, housewives, and institutionalized people)	Waged/blue collar workers	Male/female

Note: classification based on the population censuses of Great Britain 1901, Germany 1907, Italy 1901 and Spain 1900.

Template 2. Latin America social table structure based on 1870-1895 censuses

Sector/Occupation	Professions
1. Capitalists and Proprietors	Landowners; Industrialists
2. Liberal Professions	Doctors; Pharmacists; Professors; Lawyers
3. Public Administration	Judges; Notaries; Procurators; Justice officials; Public employees
4. Defense	Military
5. Religious	Secular and religious personnel
6. Mining	Miners
7. Artisans, Crafts, and Manufacturing	Masons; Metalworkers; Textile workers; Leather workers; Dyers; Clothing makers; Hat makers; Shoemakers; Seamstresses
8. Transport	Transport laborers
9. Commerce	Merchants; Bookkeepers; Shop assistants
10. Arts and Culture	Artists; Men of letters
11. Urban Services and Dependent Labor	Domestic service; Servants and day laborers
12. Small Rural Producers	Tenants; Sharecroppers
13. Peasantry and Small Rural Producers	Farmers; Peasants
14. No Profession	Without profession
15. Enslaved / Forced Labor	Enslaved persons / Forced labor

Note: classification based on the population censuses of Argentina 1895, Brazil 1872 and Colombia 1870

Template 3. Social table structure ordered by income level

Class	Occupation	Income level
1	Large landowners	HIGH
2	Capitalists (industrial, commercial, financial)	HIGH
3	Bureaucrats and liberal professionals (lawyers, doctors, etc.)	UPPER MIDDLE
4	Merchants and urban administrative personnel	UPPER MIDDLE
5	Self-employed urban workers (artisans, shopkeepers)	LOWER MIDDLE
6	Self-employed rural workers (small landowners, sharecroppers)	LOWER MIDDLE
7	White-collar industry employees	LOWER MIDDLE
8	Blue-collar industrial workers	LOWER MIDDLE
9	Low-skilled urban workers (porters, street vendors, urban servants)	LOW
10	Low-skilled rural workers (day laborers, rural servants)	LOW
11	Unemployed / No declared occupation	LOW

Note: own elaboration based on the classification table in Van Leeuwen, M., & Maas, I. (2011). HISCLASS: A Historical International Social Class Scheme. Leuven: Leuven University Press.

References

Aboagye, P. Y., Bolt, J., 2021. Long-Term Trends in Income Inequality: Winners and Losers of Economic Change in Ghana, 1891–1960. *Explorations in Economic History* 82.

Acemoglu, D and Robinson, J.A., 2017. The economic impact of colonialism, in Michalopoulos, S. and Papaioannou, E. (eds.), 2017. *The long economic and political shadow of history*, CEPR Press.

Alfani, G., 2021. Economic inequality in pre-industrial times: Europe and beyond. *J. Econ. Lit.* 59 (1), 3–44.

Allen, R. C., 2001, The Great Divergence in European Wages and Prices from the Middle Ages to the First World War, *Explorations in Economic History*, vol. 38, pp. 411–447.

Allen, R.C., 2009. Engels' pause: technical change, capital accumulation, and inequality in the British industrial revolution. *Explor. Econ. Hist*. 46, 418–435.

Allen, R.C., 2019. Class structure and inequality during the industrial revolution: lessons from England's social tables, 1688–1867. *Econ. Hist. Rev*. 72 (1), 88–125.

Alvaredo, F., 2011. A note on the relationship between top income shares and the Gini coefficient. *Econ. Lett*. 110 (3), 274–277.

Alvaredo, F., Assouad, L., Piketty, T., 2017. Measuring inequality in the middle east 1990-2016: the world's most unequal region? *Rev. Income Wealth* 665–711.

Arroyo Abad, L., Astorga Junquera, P., 2017. Latin American earnings inequality in the long run. *Cliometrica* 11, 349–374.

Arroyo-Abad, L., Maurer, N., 2025. Income inequality in Argentina during the Belle Époque. *Mimeo*.

Bengtsson, E., Waldenström, D., 2018. Capital shares and income inequality: evidence from the long run. *J. Econ. Hist*. 78, 712–743

Bertola, L., Castelnuevo, C., Weber, J.R., Willebald, H., 2008. Income distribution in the Latin American Southern cone during the first globalization boom: ca 1870-1920. In: Paper presented at the Midterm conference of the international sociological association. University of Neuchatel. June 26-28, 2008.

Berry, A., 1990. International trade, government and income distribution in Peru since 1870. *Lat. Am. Res. Rev.* 25 (2), 31–59.

Bigston, A., 1986. Welfare and economic growth in Kenya, 1914-76. *World Dev.* 14 (6), 1151–1160.

Bigston, A., 1987. Income distribution and growth in a dual economy: Kenya 1914-1976. Gothenburg University, Department of Economics. Memorandum No. 101.

Bolt, J., Hillbom, E., 2016. Long-Term Trends in Economic Inequality: Lessons from Colonial Botswana, 1921–74. *The Economic History Review* 69 (4): 1255–1284.

Booth, A., 1988. Living standards and the distribution of income in colonial Indonesia. *J. Southeast Asia Stud.* 19 (2), 310–334. September 1988.

Bourguignon, F. and Morrisson, C. (2002), Inequality among world citizens: 1820– 1992, *American Economic Review,* vol. 92 (4), September 2002, pp. 727–44.

Broadberry, S., Campbell, B., Klein, A., Overton, M., van Leeuwen, B., 2015. *British Economic growth, 1270-1870.* Cambridge University Press.Burnard, T. G., Panza, L., Williamson, J., 2019. Living Costs, Real Incomes and Inequality in Colonial Jamaica. *Explorations in Economic History* 71 (January): 55–71.

Castañeda Garza, D., Bengtsson, E., 2020. Income inequality in Mexico 1895-1940: industrialization, revolution, institutions. In: *Lund Papers in Economic History*, No. 2020:212.

Chang, C., 1962. *The Income of the Chinese Gentry*. University of Washington Press, Seattle.

de Haas, M., 2022. Reconstructing income inequality in a colonial cash crop economy: five social tables for Uganda, 1925–1965. *Eur. Rev. Econ. Hist*. 26.

De Rosa, M., Flores I. and Morgan M. (2022), More unequal or not as rich? Revisiting the Latin American exception, *Stone Center on Socio-economic Inequality Working Paper,* No. 53, August.

Erfurth, P., (2022), Unequal Unification?, *Int. Rev. Law and Econ.*

Erfurth, P., (2025), ?

Fourie, J., von Fintel, D., 2010. A history with evidence: income inequality in the Dutch Cape Colony. In: Working Paper No. 184. Universiteit Stellenbosch and Universiteit Utrecht.

Gabbuti, G., (2021), Labor Shares and Inequality: Insights from Italian Economic History, 1895-1970. *European Review of Economic History*, 25 (2), 355-378.

Gabbuti, G., and Rappa, E., (2025), Income Inequality in Italy and Its Regions (1871-1991), Paper presentated at the "Inequality & History" Workshop, Bocconi University, Milan, October 2025.

Galli, S., Theodoridis, D., & Rönnbäck, K., 2023. Economic inequality in Latin America and Africa, 1650 to 1950: Can a comparison of historical trajectories help to understand underdevelopment? *Economic History of Developing Regions*, *38*(1), 41–64.

Gómez León, M., 2021. The Kuznets Curve in Brazil, 1850–2010. *Rev. Hist. Econ./J. Iber. Lat. Am. Econ. Hist*. 39 (1), 37–61.

Gómez León, M., 2024. Social tables for Spain, 1900–1954. *Mimeo*.

Gómez León, M., de Jong, H., 2019. Inequality in turbulent times: income distribution in Germany and Britain, 1900–50. *Econ. Hist. Rev. 72* (3), 1073–1098.

Gómez León, M., Gabbuti, G., 2025. Wars, Depression, and Fascism: income inequality in Italy, 1901–1950. *Explor. Econ. Hist.*, 101715.

Jackson, T., 2023. The new history of old inequality. *Past. Present.* 259 (1), 262–289. May.

Korchmina, E. and Malinowski M. (2022), Inequality under serfdom: income and its distribution in early 19th century Russia, mimeo, 21 December 2021.

Lindert, P., Nafziger, S., 2014. Russian inequality on the eve of the revolution. *J. Econ. Hist.* 74 (3), 767–798. September.

Lindert, P., Williamson, J., 1982. Revising England's social tables 1688–1812. *Explor. Econ. Hist.* 19 (4), 385–408. October.

Lindert, P., Williamson, J., 1983. Reinterpreting Britain's social tables 1688-1911. *Explor. Econ. Hist.* 20, 94–109.

Lindert, P., Williamson, J., 2016. *Unequal gains: American growth and Inequality Since 1700*. Princeton University Press.

Maddison, A., 2001. *The World economy: a Millennial Perspective*. OECD, Paris.

Malanima, P., 2006. Pre-modern equality: income distribution in the Kingdom of Naples (1811). In: Paper presented at XIV International Congress of Economic History (Helsinki 2006). Available at www.helsinki.fi/iehc2006/papers3/Malanima

Malinowski, M. and van Zanden J.L. (2015)National income and its distribution in pre-industrial Poland in a global perspective", *European Historical Economics Society Working paper* No. 76, May 2015, p. 17.

Marmissolle, P., Willebald, H., 2025. A century of income inequality in Uruguay, 1860–1970: A social table approach. *Mimeo*

Martin, R.M., 1838. The history, antiquities, topography, and Statistics of Eastern India. Surveyed under the Orders of the Supreme government, and Collated from the Original Documents At the E.I. House. W. H. Allen and Co, London.

Merette, S. (2013), "Preliminary analysis of inequality in colonial Tonkin and Cochinchina", mimeo, June.

Milanovic, B., 2006. An estimate of average income and inequality in Byzantium around year 1000. *Rev. Income Wealth 52* (3), 449–470.

Milanovic, B., 2016. *Global inequality*. Harvard University Press.

Milanovic, B., 2018. Towards an explanation of inequality in pre-modern societies: the role of colonies and high population density. *Econ. Hist. Rev. 71* (4), 1029–1047, 2018.

Milanovic, B., 2019. All the Ginis (ALG) Dataset (version October 2019). Available from: https://stonecenter.gc.cuny.edu/publications/all-the-ginis-alg-dataset-version-february-2019/

Milanovic, B., 2024a. How rich were the rich? An empirically-based taxonomy of pre-industrial bases of wealth. *Explor. Econ. Hist*, 93.

Milanovic, B., 2024b. The three eras of global inequality, 1820–2020 with the focus on the past thirty years. *World Dev*. 177.

Milanovic, B., Lindert, P., Williamson, J., 2011. Pre-industrial inequality. *Econ. J.* 255–272.

Morrisson, C., Snyder, W., 2000. The income inequality of France in historical perspective. *Eur. Rev. Econ. Hist.* 4, 59–83.

Modalsli, J., 2015. Inequality in the very long run: inferring inequality from data on social groups. *J. Econ. Inequal.* 13 (2), 225–247.

Moraes, M.I., Challú, A.E., 2025. *Inequality in Preindustrial Latin America: Economic and Social Patterns*. Routledge.

Nieto Ramos, A., 2024. Distribución de ingreso en Colombia durante la Primera Globalización (1870–1918). Doctoral dissertation, Universitat de València.

Nikolić, S., Novokmet, F., Larysz, P.P., 2024. Income inequality in Eastern Europe: Bulgaria and Czechoslovakia in the twentieth century. *Explor. Econ. Hist*. 94, 101594.

Petty, William (1690). Political Arithmetick, or, A Discourse Concerning the Extent and Value of Lands…, London: Clavel and Mortlock.

Piketty, T., 2014. *Capital in the Twenty-First Century*. Harvard University Press.

Piketty, T., Saez, E., 2003. Income inequality in the United States, 1913-1998. *Q. J. Econ.* 118 (1), 1–39.

Piketty, T., Yang, Li, Zucman, G., 2019. Capital accumulation, private property and rising inequality in China, 1978-2015. *Am. Econ. Rev.* 109 (7), 2469–2496. July.

Rodríguez Weber, J. E., 2017. Desarollo y Desigualdad En Chile 1850-2009: Historia De Su Economic Politica. Centro de Investigaciones Diego Barros Arana (Dibam).

Rodríguez Weber, J. E., 2023. Top incomes and the ruling class in Latin American history. Some theoretical and methodological challenges. *Economic History of Developing Regions*, 38(3), 335–352.

Soltow, L., 1968. Long-run changes in British income inequality. *Econ. Hist. Rev.* 21, 17–29.

Soltow, L., van Zanden, J.L., 1998. Income and Wealth Inequality in the Netherlands, 16th-20th Century. Het Spinhuis.

Tadei, F., Alfani, G., 2019. Income Inequality in French West Africa: Building Social Tables for Pre-Independence Senegal and Ivory Coast. *SSRN Scholarly Paper ID 3494245*. Rochester, NY: Social Science Research Network.

van Zanden, J.L., 2003. Rich and poor before the industrial revolution: a comparison between Java and the Netherlands at the Beginning of the 19th Century. *Explor.Econ. Hist.* 40 (1), 1–23. January.

von Fintel, D.C.L. and Green E. (2023), "Estimating historical inequality from social tables: towards methodological consistency", mimeo.

Wronski, M., 2021. Income inequality in the Duchy of Warsaw (1810/11). *Scand. Econ. Hist. Rev.*

Appendix

A note on Gini estimation for social tables

The Gini decomposition by recipient (social class) can be written

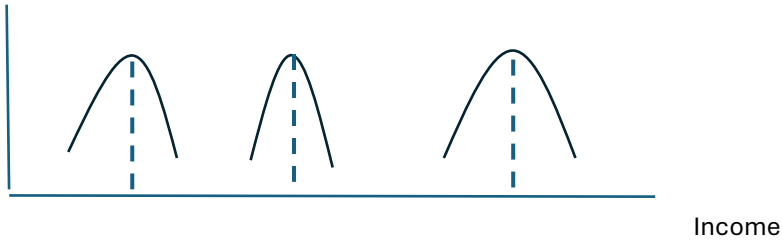
$$\frac{1}{\mu}\sum_{i}^{n}\sum_{j>i}^{n}(y_j - y_i)p_i p_j + \sum_{i}^{n} G_i\, p_i \pi_i + L$$

where y_i=mean income of group *i*, p_i=population share of group *i*, π_i=income share of group *i*. The first term gives the between-component, the second, the within, and the third term is the overlap term which is non-zero only in the cases where some individuals from a mean-poorer group have a higher income than some individuals from a mean-richer group.

Consider for simplicity the situation with three classes available in a social table and with mean incomes such that yk>yj>yi. Suppose further that, as in Figure 1 below, there is no (or almost no) variability around the mean of each class. In other words, suppose that every member of a given class has the same (or very close to same) income. In that case, only the first term (the between term) in decomposition (1) matters. Inequality within classes is by assumption zero or negligible, and the third term (the overlap) is zero, because no members of two classes overlap in terms of their incomes.[23]

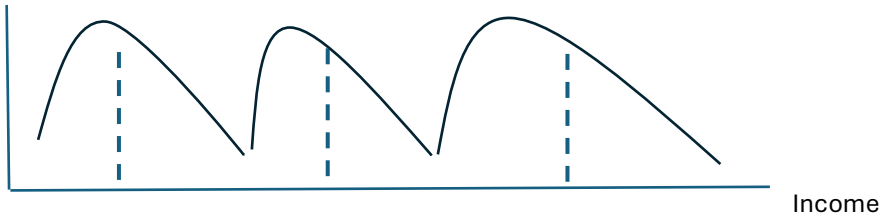
Figure 1. Social table with no, or almost no, inequality within classes

[23] Note that the Kakwani approximation that is based on maximum and minimum Ginis within tabulated income statistics does not in a general case apply to the social tables because in the Kakwani approximation all individuals within a low-end class have (by definition) a lower income than all individuals within the upper class. That is, the Kakwani approximation works only if there is no overlap between classes (such an overlap is by definition excluded in a tabulated income distribution). If we believe that classes are sufficiently distinct and their incomes' distributions around the mean rather tight, we can use the Kakwani approximation which is equal to $\frac{1}{3}$ Gini (min) +$\frac{2}{3}$ Gini max. Gini (min) is equal to the between-component alone. The maximum Gini is equal to Gini(min) + $\frac{1}{\mu}\sum_i^n p_i^2\, \beta_i(1-\beta_i)\Delta_i$, where Δ_i=distance between top and bottom income within a given tabular range, and β_i = percentage of recipients in a given tabular group that receive the lower bound income of that group. (Obviously, 1- β_i recipients will, in order to maximize inequality within the group, be all located at the upper bound income of the group. To satisfy a given mean income, we must have $\beta_i y_{lower\ bound} + (1-\beta_i)y_{upper\ bound} = group\ mean$).



Next, suppose that within each class, there is some variability of incomes but that there is no overlap between classes. The graph now becomes as in Figure 2.

Figure 2. Social table with high within-class inequality but no overlap between classes



That obviously means that the within-term can no longer be taken as equal to zero. But we do not know individual class Ginis. Suppose however that we believe that the distribution of income within each class (while classes are still non-overlapping) follows a lognormal distribution. The Gini coefficient of lognormal distribution is

$$G_i = 2N\left(\frac{s_i}{\sqrt{2}}\right) - 1$$

where s=standard deviation of logs of income and N(.)=normal (Gaussian) distribution function. We can then replace in (1) for each within-class Gini

$$\frac{1}{\mu}\sum_{i}^{n}\sum_{j>i}^{n}(y_j - y_i)p_i p_j + \sum_{i}^{n}(2N\left(\frac{s_i}{\sqrt{2}}\right) - 1)\, p_i \pi_i + L$$

It then becomes clear that the overall Gini will be equal to the between Gini that we can readily calculate, plus will depend on standard deviations of logs of income within each social class. To give an idea of the value of s. If within-class Gini is, say, 0.5 (which would be a fairly high inequality for an individual class), then N(s/$\sqrt{2}$)=0.75 and N(0.68)=0.75 and thus s=0.96. If Gini is rather low, say 0.2, then N(s/$\sqrt{2}$)=0.6 and s=0.84. Because of the double weighting (by both p and π), the within-term is likely to be small. To show that suppose that for all three classes, the Gini is always at a middling level of 0.3, and that population shares are 0.5, 0.3 and 0.2, with average incomes of 1, 2 and 3 (smaller

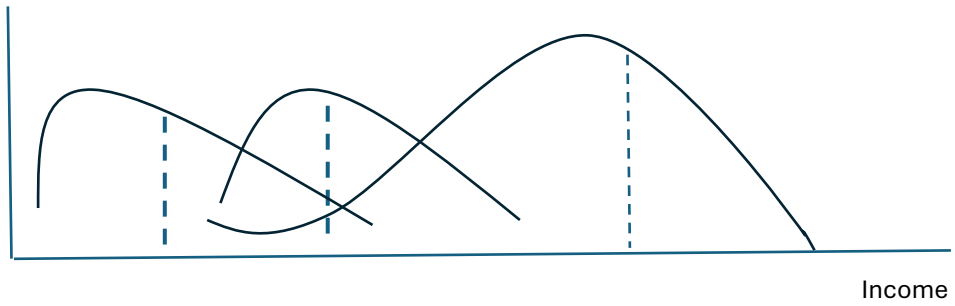
class has a higher average income). Then the income shares are 0.3, 0.4 and 0.4. The within-class Gini will be 0.1, that is within-class Gini will add 10 Gini points to whatever is the value of the between class Gini.

With greater number of classes, however, the within-Gini will under *ceteris paribus* conditions tend to go down because of double weighting. In the example below, the within-Gini is only 0.007, that it, it does not reach even 1 Gini point. The essential thing to retain is that a large number of classes not only tends to make between-Gini high but also to reduce the importance of within-Gini because with many classes, each class's either income or population share will be small and the addition of such small quantities that are double-weighted (by both population and income shares) is in turn small too.

Example 1. Calculation of total within-class Gini with ten classes

Within-class Gini	Mean class income	Population share	Income share	Contribution to within-class Gini
0.3	1	0.25	0.003	0.0002
0.3	2	0.22	0.005	0.0003
0.3	3	0.15	0.005	0.0002
0.3	4	0.13	0.012	0.0005
0.3	5	0.1	0.022	0.0007
0.3	6	0.05	0.039	0.0006
0.3	7	0.04	0.074	0.0009
0.3	8	0.03	0.135	0.0012
0.3	9	0.02	0.248	0.0015
0.3	10	0.01	0.457	0.0014
Total/mean	*5.5*	*1*	*1*	*0.0074*

Figure 3. Social table with high within-class inequality and overlap



The situation gets more complicated as in Figure 3, where income overlap becomes important. This can happen either because mean incomes are closely bunched, that is, a society has many classes but their mean incomes are not very different from each other and consequently any variability of incomes within a class spills over into the overlap component, or alternatively, because the social structure is fairly polarized (so the means are very much apart) but variability within each class (that is the standard deviation of incomes) is very high.

We should thus worry about the overlap component if the classes included in the social table are very close to each other (i.e., the means are similar) or if we have grounds to believe that the distribution of income within several classes is very widespread, namely that the within-class Ginis are high. This can be the case with traders who, despite a relatively moderate mean income, might include very prosperous traders, possibly richer than capitalists and landlords, and very poor traders, possibly poorer than many workers. Such class would then add significantly to inequality, but it is unlikely that there would be many such classes if the selected social groups are salient.

In conclusion, if mean incomes of social classes are fairly different, it is unlikely that the overlap component could be substantial. Even if each individual class Gini is large, the double weighting would reduce the sum of such within-class Ginis and in that case neither the within-class component nor the overlap would be very important. Thus, the between-class component proxies well for overall inequality when the classes selected are far apart (in terms of their mean incomes), when there are many such classes, and when the within-class inequality is moderate.

Now, consider the following possibility. Let, as before, all within-class distributions be lognormal but the top class follow the Pareto distribution. We then get

$$\frac{1}{\mu}\sum_{i}^{n}\sum_{j>i}^{n}(y_j - y_i)p_i p_j + \sum_{i}^{n-1}\left(2N\left(\frac{s_i}{\sqrt{2}}\right) - 1\right)p_i\pi_i + \frac{1}{2\alpha - 1}p_n\pi_n + L$$

Suppose, conventionally, that α=1.5, and that the top class is small (1 percent) but receives 8 percent of total income. Then the top class's contribution to within-Gini will be 0.5*0.01*0.08=0.0004 or 0.04 Gini points. In other words, very small. If that class's income is sufficiently far apart from the rest of the population, it is obvious that even the rich class's contribution to total inequality (other than through the between component) will be small.

In lieu of conclusion. Every income distribution is, by definition, heterogeneous in income. The *art* of creation of a social table is to replace this heterogeneity by homogeneity that nevertheless preserves the essential features of income distribution and its inequality. This means finding social classes that are sufficiently different from each other in terms of mean incomes (hence heterogeneity) while sufficiently homogeneous in incomes within themselves. In practical terms it means that (i) income deviations within the class should be small, while (ii) the distances in mean incomes between classes are sufficiently great. The two of them together ensure that the within component of the Gini as well as the overlap term are small, and that consequently the between component provides a plausible proxy of overall inequality. In even more practical terms, it means that if there are social groups where our belief, prior to the creation of the table, is that they exhibit strong income variability, most of our effort should be directed towards breaking them into several components. The importance of that break-up is greater the greater the size of the group(s).