TOP EARNERS AND THE GREAT COMPRESSION:

NEW ESTIMATES BASED ON TAX RECORDS

MIGUEL ARTOLA BLANCO
VICTOR MANUEL GÓMEZ-BLANCO

WORKING PAPER N°2024/11

MARCH 2024

WORLD INEQUALITY LAB

Top Earners and the Great Compression: New Estimates Based on Tax Records

Abstract

This paper presents new estimates of wage inequality in the United States from 1918 to 1949. Building upon a new top-income methodology, we provide various definitions of top wage groups that account for the sharp fluctuations in the employed population during this period. The results confirm the decline in wage inequality during the Second World War, primarily due to the relative stagnation of the top 1% group and a sharp increase at the bottom. However, the underperformance of top wage earners was driven by a significant compositional shift that resulted from an unprecedented rise in the corporate tax. This change prompted a shift in business preferences regarding their legal status, fostering a surge in partnerships during the 1940s. Consequently, a significant number of workers transitioned from salaried positions to self-employment, which amplified the compression observed in the wage distribution.

Keywords: *wage inequality; Great Compression; proprietors' income; executive compensation*

JEL codes: J31, J82, N32

This paper has been presented in the World Inequality Conference (Paris, 2021) and in seminars in Universidad Carlos III de Madrid, CUNEF, EHES, Universitat de Valencia, and Wageningen University. We are grateful for the comments made by participants, and especially for the feedback provided by Luis Bauluz, Matthew Fisher-Post, Daniel Gallardo, Esteban Nicolini, Onur Yukcu, Fabio Braggion, and Thomas Piketty, among others. The authors are also grateful for the financial support provided by the BBVA Foundation [Beca Leonardo, Year 2020]; and the Ramon Areces Foundation [Ayudas a la Investigación en Economia, Year 2017].

1. Introduction

Wage inequality in the United States declined substantially in the years surrounding the Second World War (Goldin and Margo 1992). This phenomenon stands out as a primary catalyst in the broader process of reducing income inequality (Lindert and Williamson 2016, 194). However, important questions about the causes and tempo of the Great Compression remain. The first question involves the causes of this phenomenon, especially whether it was driven by a large rise in wages at the bottom of the distribution or by a relative (or absolute) underperformance of the top. The seminal paper by Goldin and Margo (1992) credited both forces, as it highlighted the surge in the demand for unskilled labor and a reduction in the college premium due to the expansion of education. However, more factors seem to be relevant for explaining the growing fortunes of the lowest-paid workers, as reflected by the decline in the racial wage gap (Maloney 1994; Bailey and Collins 2006) or the impact of union density (Collins and Niemesh 2019).

Scholars who highlight the declining fortunes of the highest-paid employees present a different perspective that is mostly tied to the effects of institutions and regulations. Piketty and Saez (2003), Frydman and Saks (2010) and Frydman and Molloy (2012) provide evidence of the relative decline in corporate executive pay during the Second World War. However, each set of authors has a different argument: the first highlights the control over executive pay imposed by the National War Labor Board, while the others insist on the indirect effect of unionization and the decreased compensation of executives from the largest corporations. Philippon and Reshef (2012) studied a different group—workers in the financial industry—and found a decrease in the wage premium from the 1930s onward because of the growing regulation of this sector and a reduction in risk-taking activities.

Furthermore, scholars have varying perspectives on the timing of the decline in wage inequality. Although everyone agrees that the Second World War caused a levelling in pay, not everyone necessarily attributes such a key role to the war experience. Some scholars insist that several secular forces were already present during the interwar period, pushing inequality toward lower levels. Goldin and Katz (2010) and, more recently, Autor, Goldin, and Katz (2020), argue that a potential decrease in wage inequality might have occurred between 1900 and 1920, primarily because the premium paid to white-collar workers was reduced. However, other studies that focus on specific events, such as the Great Depression, find opposite trends, with wage inequality growing due to the unequal effect of unemployment and wage rigidity (Hanes 2000, 1433). Unfortunately, the literature on education and skill premia and that related to business and monetary cycles have hardly interacted.

As a contribution to this field, this paper provides new estimates on wage inequality in the United States for this period. The main objective is to clarify whether the Great Compression led to a decline—in absolute or relative levels—among top earners (typically, defined as those within the top 1% of the wage distribution) and for which reasons. This research is carried out by focusing on a broader timeframe (1918-49) and refining the precision of previous estimates (Piketty and Saez 2003, 2007). Piketty and Saez provided a first set of estimates on the wages earned by workers at the top from 1927 onward based on income tax statistics. In this paper, we follow a similar approach, refine the estimates for the top percentiles, making them consistent with other sources and extending the results to preceding years. Additionally, we provide a better calibration for the impact of sudden employment shocks, such as those experienced during the Great Depression and Second World War, on top wage metrics. The results confirm that an

accurate perspective of wage inequality cannot be obtained without correctly capturing the top of the distribution.

The second major contribution of this paper is to have a better understanding on the factors defining the number, profile, and relative compensation of top earners. By reconciling the results obtained from tax records with the population census and other administrative data, we narrow the gap between the top incomes literature and the studies on inequality based on occupations and specific industries (Gómez León and De Jong 2019; Lindert and Williamson 1983; Milanovic, Lindert, and Williamson 2011; Modalsli 2015). This enables us to conduct an in-depth analysis of the crucial roles played by managers, professionals, and the self-employed at the very top of the distribution. Notably, a key aspect often overlooked by economic historians, despite its growing importance in present day times, is the shift in business preferences regarding their legal status. Firms could move between being a corporation or a partnership, and this transition facilitated the movement between wage and self-employment, thereby affecting the wage distribution.

The results of this paper confirm that wage inequality was high in the 1920s and 1930s, with no evidence of being in a secular decline. Later, during the Second World War, real wage growth among the top 1% stagnated, contributing significantly to the Great Compression. However, this abrupt shock is more complex than usually thought. The compensation of highly paid occupations lagged behind the overall wage growth of the economy while the legal form of businesses experienced an unprecedented shift due to the substantial rise in the corporate tax rate that began in 1938. A general preference for the creation of partnerships and a relative decline in the number of corporations altered the employment status of highly paid workers from salaried to self-employed, thereby

exacerbating the decrease in wage inequality. This paper does not contest the significance of the Great Compression but emphasizes that most of this phenomenon stemmed from a substantial upsurge in wages at the lower end of the distribution, whereas the situation at the top exhibited greater complexity. The traditional top wage metric shows that top earnings declined in real terms. However, considering the expansion of the employed population and the transition to self-employment, the top 1% experienced more substantial growth in their earnings. This has resulted in an overestimation of the decline in inequality during the 1940s by approximately 30%. Furthermore, this perspective highlights that the economic boom of the 1940s primarily favored low-wage earners, although eventually, every cohort of workers experienced a rise in real earnings owing to advancements in productivity.

The remainder of the paper is organized as follows. The methodology and data sources are presented in Section 2, which addresses the process of deriving wage estimates from income tax tabulations. Moreover, we explain how these figures can be reconciled with other sources, such as the national census and surveys. Section 3 presents the main results, followed in Section 4 by a specific discussion on top corporate managers and the effect of the upsurge in self-employed high-earners. Finally, Section 5 provides a brief conclusion.

2. Data and methods

Wage inequality has been a subject of substantial research in U.S. economic history. However, from a methodological standpoint, a noticeable disparity exists between studies conducted before and after 1940. This difference can mostly be explained by information on wages and education being recorded in the 1940 national census for the first time, which augmented the traditional variables on age, gender, occupation, and industry. In

previous periods, information was more limited, so scholars had to rely primarily on census employment figures supplemented by separate records on pay for representative occupations from various sources, such as state-level census records and business surveys.

Alternatively, this paper contends that starting from 1918 tax records offer a consistent means to compute wages in the upper part of the distribution and derive metrics on wage inequality. However, to achieve this goal, several adjustments to the information provided in personal income tax statistics are needed, as elaborated in the following section. These improvements follow the path of a growing number of studies that have revised the first estimates provided almost two decades ago (Jaworski and Niemesh 2018; Geloso and Magness 2020; Fisher-Post 2020; Geloso et al. 2022; Auten and Splinter 2023). Following these adjustments, estimates derived from tax records can be harmonized with 1940–50 census records and other nationwide administrative data. This alignment results in a more comprehensive understanding of top earners during this period.

2.1 Top wage shares

This paper takes the literature on top incomes, which was pioneered by Kuznets (1953) and further developed by other scholars (A. B. Atkinson and Harrison 1978; Piketty 2001), as a starting point. This method focuses on measuring the upper tail of the distribution, most commonly through tax records, while the bottom is estimated as the residual from predefined population and income denominators. The same approach can be applied to wage inequality. The share of the top *X* percent is obtained by dividing the cumulative wages above the corresponding threshold, as shown in Equation 1.

$$Wage\ share\ of\ top\ x\% = \frac{\sum_{i=nx/100}^{n} w_{x\,i}}{\sum_{i=1}^{n} w_i} \qquad (1)$$

where w_i is the wage of individual i, x is the selected percentile, and w_{xi} is the wage of individuals who are above the threshold. This exercise is usually performed by employing Pareto interpolation techniques; in this paper, these are carried out by incorporating the generalized Pareto curves by Blanchet, Fournier, and Piketty (2022). The employed population and income denominators are defined according to the U.S. National Income and Product Accounts (NIPA) from 1929 onward and prior to that date by using the series of Piketty and Saez (2007, 210) built from the national accounts of Kuznets (Kuznets 1954).

Table 1. Cross-tabulation of Income Tax Returns Reporting Wages, 1936

Income brackets (net income)	Number of returns (A)	Returns with wages (B)	Wage brackets (C)											
			All	1	1.000	2.000	3.000	4.000	5.000	10,000	15,000	25,000	50,000	100,000
1	277,803	Not available	Not available											
1,000	2,111,789													
2,000	1,317,752													
3,000	729,755													
4,000	299,389													
5,000	440,886	287,378	100%	5%	4%	5%	6%	9%	66%	4%	0%	0%	0%	0%
10,000	105,582	64,732	100%	6%	4%	4%	4%	4%	28%	44%	6%	0%	0%	0%
15,000	71,067	42,353	100%	7%	4%	3%	3%	3%	18%	23%	36%	3%	0%	0%
25,000	41,137	23,940	100%	8%	4%	3%	3%	2%	13%	13%	25%	28%	1%	0%
50,000	13,620	7,772	100%	10%	3%	3%	2%	2%	10%	10%	17%	26%	17%	1%
100,000	4,719	2,727	100%	14%	5%	3%	2%	2%	7%	7%	12%	20%	18%	11%

Source: SOI (1936).

Our estimates on top wage groups primarily rely on the data extracted from personal income tax records. This task is especially daunting given the absence of microdata, which only became available in the U.S. in the early 1960s. Thus, we depend on tabulations and interpolation techniques. Piketty and Saez (2003) provided these estimates from 1927 onward using the available information in the *Statistics of Income*

(SOI). Table 1 presents a summarized version of the original tabulated data, which shows the extreme details on the number of wage earners and their distribution among those with more than $5,000 in net income (i.e., roughly speaking, those within the top 1%). For all other returns, information is much more limited; therefore, Piketty and Saez had to estimate the wage distribution based on some assumptions about the number of earners (i.e., Column B in Table 1) and the relative distribution of wages within each bracket (Column C).

We start by revising and extending these estimates. Our amendments, which are inspired by the largely forgotten contribution of official authorities after the Second World War (Office of Business Economics, 1953), have four basic objectives. First, we improve the imputation method for the distribution of wages that fit below the top 1% of earners. Second, we extend the estimates to 1918 based on the close relationship that can be observed between the relative number of wage earners (per income bracket) and the variance in wages. Third, we move from the traditional unit of measurement (tax units) that was used by Piketty and Saez to another—individuals—that seems to be more correct conceptually. Finally, we include the sources of wage income that were exempted from the federal income tax, most notably, the earnings of local and state employees until 1938.

The precise details of this new estimation method are sufficiently complex and are explained further in another more technical paper, with comprehensive details provided in the accompanying appendix. The main insights can be summarized as follows. The first step involves computing the number of wage earners per income bracket using reports published by tax authorities in the 1930s and 1940s that provide more information than standard SOI reports. Furthermore, we estimate the distribution of wages from net income tabulations by using a procedure that we have named the *Constant Lorenz Curve*

Transformation. This method starts with the net income distribution and makes separate estimates on the wage distribution for persons fitting in each bracket by keeping constant the observed Lorenz curves for a closely related variable (i.e., total income) in one year (1936) with more detailed information. Then, we use these Lorenz curves as transition matrices for the years that have an unknown distribution (Liebenberg and Kaitz 1951). This method helps us prevent the assumption of identical rankings between wage and net income distributions, as in Piketty and Saez (2003). In the third step, we individualize wages in joint returns by using the information derived from the 1940 census. Finally, following the revision of Geloso et al. (2022), we include the sources of wage income that were exempted from the federal income tax by imputing the number of local employees and their respective wages above the tax threshold.

In this paper, we focus on the top 1% (above the 99th percentile) and the next 4% (from the 95th to 99th percentile) of wage earners. Both groups are well covered in income tax statistics since they almost always fit above the minimum filing threshold. We refrain from extending results beyond these percentiles (for example, up to the top 10%), as this would entail projecting estimates toward a group of workers that is not observed in the original source prior to 1939. The bottom 95% of the wage share is computed as a residual. We also restricted our analysis to 1947 due to a substantial tax reform in the subsequent year, which notably increased the number of joint returns and complicates the task of accruing wages within couples. Overall, as we illustrate in the Appendix, the combined research provides a more accurate perspective on the evolution of the top wage distribution during this period relative to the series elaborated by Piketty and Saez.

2.2 Top wage groups and occupations

The IPUMS sample can be used to compare the results derived from the personal income tax tabulations with the national censuses. Table 2 presents the estimates for the 95th to 99th percentiles in 1939 using both statistical sources. This comparison initially serves as a robustness check to evaluate the representativeness of the previously obtained series. Remarkably, our methodology reveals closely aligned percentiles, albeit with tax records slightly higher than those in the census. These disparities can be traced to the rounding effect evident in self-reported data within census records. For instance, this is illustrated by the most substantial disparity observed in the 97th percentile, which is $99 lower in census records due to rounding effects at a $3,000 yearly wage threshold. Nonetheless, the two key percentiles of interest are almost identical, with the 95th percentile displaying a discrepancy of $4 and the 99th percentile having a difference of $27.

Table 2. Estimated percentiles 95 to 99 through the SOI and census data, 1939.
Figures in current dollars

Percentile	**SOI**	**Census**
95	2,504	2,500
96	2,809	2,750
97	3,099	3,000
98	3,611	3,600
99	5,027	4,992

Source: SOI, IPUMS

Since both sets of records are highly comparable, we propose using them in a combined manner. This approach surpasses the typical analysis conducted solely through tax statistics, which, lacking socioeconomic profiles of taxpayers, restrains a comprehensive examination of top-earning groups. Census records precisely furnish this vital information, including details such as age, gender, race, occupation, and worker industry. On the other hand, by integrating the income tax data, we overcome the limitations of

previous studies that only used census records and had to confront the challenges of using top-coded data.

Nevertheless, the census provides data only on a decennial basis, limiting our understanding of the evolving profile of top earners, particularly within the remarkably volatile context of this period. To offer a more comprehensive and dynamic analysis, we concentrate on two key groups among high-paid workers for whom there are comparable records in terms of employment and compensation at the national level: corporate officers and top-earning proprietors. These distinctive groups provide unparalleled insights into the profile of the top 1% of the wage distribution. Corporate officers constitute the occupational group that is most significantly overrepresented at the top, and top-earning proprietors, although not formally classified as wage earners in conceptual terms, closely correspond to the group of workers associated with the top 1%.

Corporate officers, as defined herein, refer to individuals occupying the most senior managerial roles within corporations, including positions such as chairperson, president, treasurer, and other executive officers. Their compensation, both at the national level and in relation to their respective company's balance sheet size, is documented within corporate tax statistics. By amalgamating these data with regulatory filings of listed corporations, which are available from 1934 onward, one can accurately compute the actual count of officers and estimate the distribution of earnings among them. In performing these estimations, we have benefited substantially from the recent contribution of Kwon, Ma, and Zimmermann (2023), who gathered all the available information from corporate tax statistics, and from the Survey of American Listed Corporations (SALC), which is a unique dataset of the Securities and Exchange

Commission (1940) that includes information on corporate pay for almost 1,000 corporations. Further details on these estimates are provided in the Appendix.

In the case of proprietors, data on their employment and compensation by industry are readily available from the NIPA. Educational levels and other sociodemographic variables are reported in the 1940 and 1950 national censuses. The critical piece of information is an estimation of the income distribution of proprietors across major industries; in this calculation, we benefit from the detailed tabulations on sole proprietorships (which usually appear in SOI every two years) together with data on partnerships (with full data in separate statistics for 1939 and 1953 and more limited information in SOI for the years in between). More comprehensive details are available in the Appendix for interested readers.

3. Top groups: Shares and average wages

Our analysis starts with the examination of the wage shares (Figures 1A, 1B and 1C) to uncover the main trends in inequality. The shares evolved during the interwar period (1918-39) within some close bounds, as illustrated by the fact that the top 1% wage share was between 8% and 10% of the total, and the next 4% ranged between 10% and 12%. Overall, these are relatively high levels compared to those in other countries in which top wage shares have been estimated (Piketty 2001; Moriguchi and Saez 2008; Atkinson and Voitchovsky 2011), although not as extreme as those observed in the United States since the 1990s.

Figure 1. Wage Shares for the Top 1%, Next 4% and Bottom 95%

Figure 1A. Top 1%

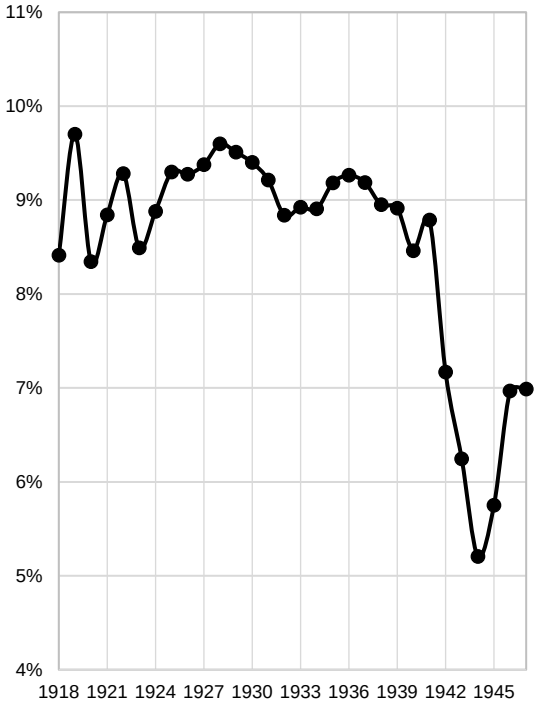


Figure 1B. Next 4%

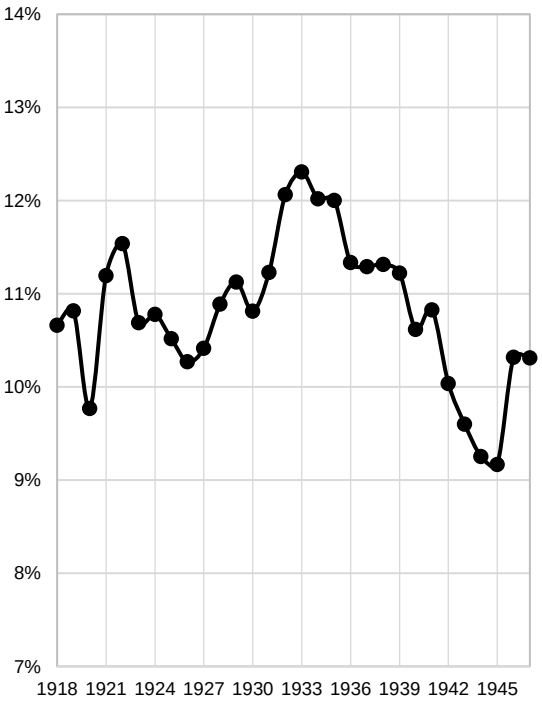
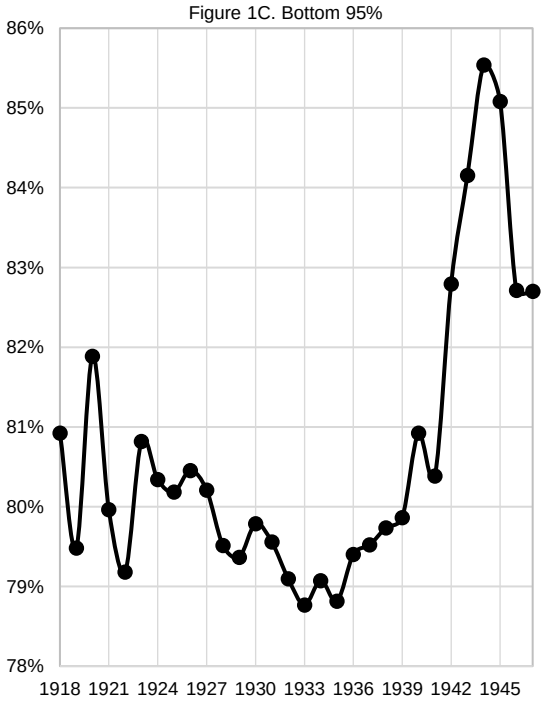


Figure 1C. Bottom 95%

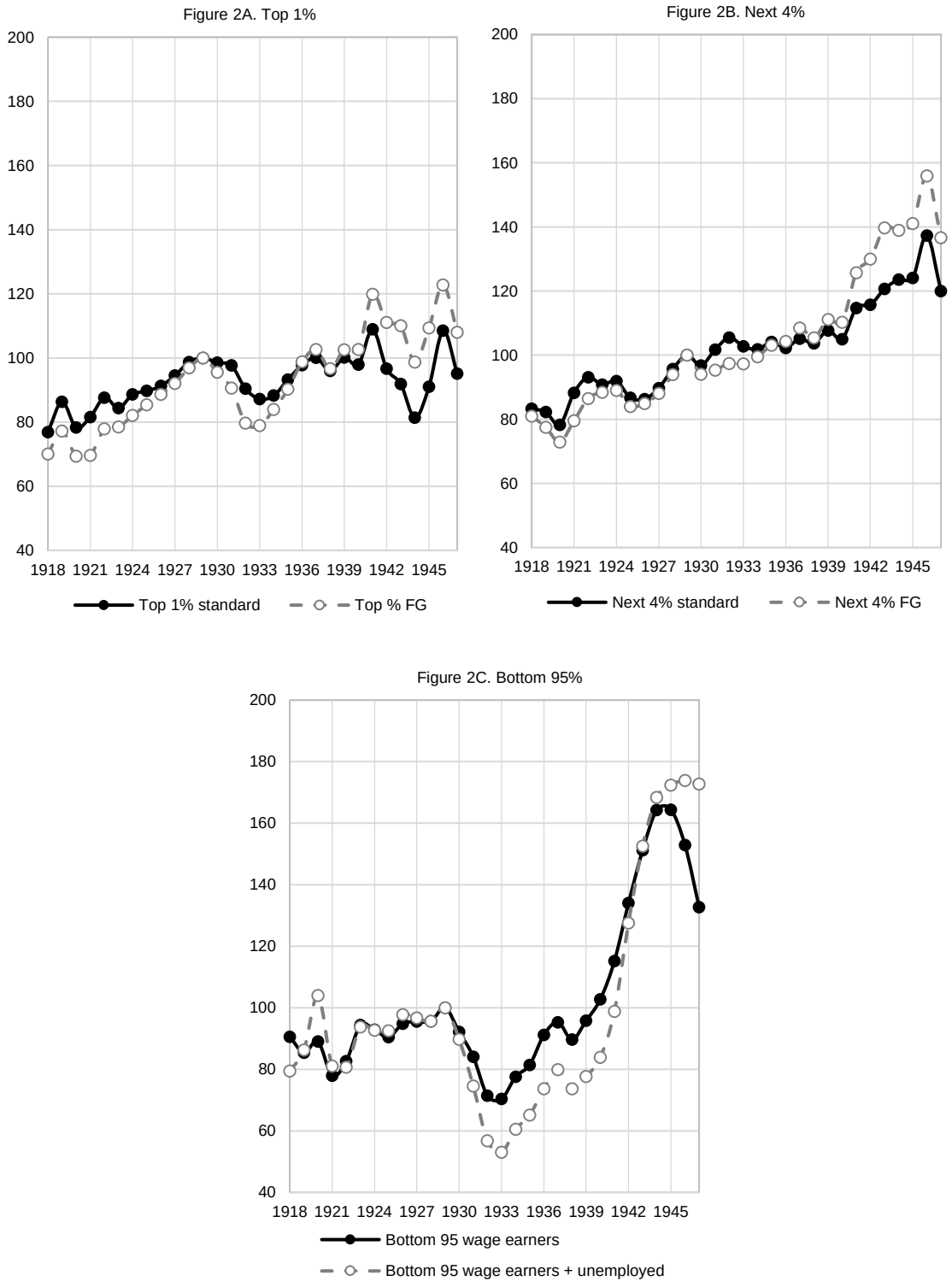


Sources: SOI.
Note: The wage shares of the top 1, next 4 and bottom 95% correspond, respectively, to groups between p99-100. p95-99, p0-95.

The top 1% wage share increased considerably in the 1920s (especially in the years 1924-8), which was reversed at the start of the Great Depression (1929-33). The next 4% group hardly increased its share during the 1920s boom, as it hovered at approximately 11% of the wage bill. Then, in the first years of the 1930s, the next 4% group experienced a very significant increase that was partly reversed in the second half of the decade. By construction, the bottom 95% commanded a dwindling share of wages from the early 1920s to 1933 and later stagnated during the rest of the Depression. The three series unequivocally show that, by far, the largest shock occurred during the Second World War. In this relatively short time, the top 1% share declined by almost 2 percentage points and stabilized at a permanently lower level. The next 4% group also experienced a decline, though with a weaker permanent effect, and by 1946-7, its share was roughly at the same level as that in the mid-1920s. Finally, the bottom 95% undoubtedly benefited the most from the war years.

Wage shares, in either definition, reveal more about the relative evolution of each group than about their absolute performance. To illustrate the evolution of earnings in relation to the business cycle, we calculate real average wages using two definitions: standard and fixed-group. While the first definition is based on the percentiles resulting from the number of employed workers in each year, for the fixed-group share, a permanent number of individuals is selected, and their respective average wage is computed (Krolage et al., 2022). Implicitly, this metric works under the assumption that sudden changes in employment fall entirely on the left tail of the distribution. In this paper, the fixed groups of the top 1% and next 4% are constructed relative to the employed population in 1939 (39.6 million wage and salary workers). This is a good reference point and of a similar order of magnitude to that of 1929 (37.7 million employees), roughly speaking, the maximum employed population of the interwar period.

Figure 2. Average Real Wages for the Top 1%, Next 4% and Bottom 95% Percentiles (1929=100).



Sources: SOI.

Notes: The fixed-group (FG) shares have been constructed relative to the employed population of 1939. By definition, this metric is equal to the standard share in this year. The average wage for the bottom 95% of the wage-earning population cannot be calculated through the fixed-group definition. Instead, we proxy its evolution by adding the number of unemployed individuals to the bottom 95% of the wage-earning population in each year.

Figures 2A and 2B display the mean real wages for the top wage groups and provide some relevant conclusions. First, in both cases, the difference between the average earnings in

the standard and fixed-group measures is especially relevant in critical junctures, such as the Second World War or the Great Depression. In the average wage derived through the standard shares, the actual wage growth of top earners in periods of net job creation is typically underestimated. The opposite happens in periods of job destruction. In contrast, the fixed-group share provides an upper bound in these periods.

Studying the real wages of the top 1% of earners is especially relevant, as their real earnings were significantly more volatile than those of the next 4% group, which enjoyed a much stronger downward rigidity. The real wages of the top 1% increased substantially in the 1920s (growing roughly by 30% in the fixed-group definition) and then stagnated for the following two decades. The two shocks in the Great Depression and the Second World War are probably not as surprising, although a comparison of the two definitions reveals a more complete perspective on the factors at play. From 1929 to 1933, the real earnings of the top 1% group declined substantially (by 13% in the standard definition or by 20% in the fixed group), thereby making the Great Depression the biggest negative shock in earnings, both in terms of depth and persistence. In the war years, the difference between the two metrics is also quite pronounced. Between 1939 and 1947, the real earnings of the top 1% declined by 5% according to the standard metric, while in the fixed-group definition, real earnings grew by 5%. Thus, the Second World War was not such a negative shock for top earners, but the difference lies in the radically different economic context, as shown by the performance of other groups.

Trends in the real earnings of the next 4% are substantially different. Overall, under both scenarios, this group of workers shows better relative performance at times of economic crisis, such as in the 1920-1 recession, and even more significantly, at the start of the Great Depression (1929-33). In fact, in the fixed group definition, real salaries hardly

decreased, indicating that this group was particularly isolated in terms of wage cuts at times of crisis. Again, the growth in earnings in the fixed group gauge was greater than that in the standard one, especially during the years from 1939 to 1943. This evidence elucidates that the war-induced boom did not particularly benefit this group and that their wages underperformed the broader economy.

For the bottom 95%, one cannot construct a fixed-population measure; however, a relatively good proxy is to compute a mean wage by dividing the aggregate compensation among wage earners and unemployed persons. The logic behind this exercise is that a substantial part of the adjustment of the bottom takes place through redundancies, and the average wage that results from counting the unemployed better reflects the actual earnings of the potentially working population. This metric may also better reflect a situation with a significant reduction in working hours and an increase in part-time employment, which was characteristic of the Great Depression (Bernanke 1986). Most likely, workers were entering and exiting the labor market more fluidly in this scenario (Hatton and Thomas 2010, 476).

Figure 2C depicts the real earnings in these two scenarios. The difference between them is not relevant to identifying the absence of significant growth in earnings during the Roaring Twenties (1924-9), but it helps to better understand the enormous decline of the Great Depression. Real wages experienced a decline of 30% based on the standard definition, while accounting for unemployment resulted in an approximate 50% decrease. The substantial disparity between the two figures underscores the significance of unemployment's influence on labor market adjustments during the crisis. These declines marked the most significant losses for any group and were not reversed until the United States entered the Second World War in 1941. Finally, it is crucial to note the remarkable

recovery observed toward the period's conclusion, with earnings soaring by more than 50%.

In summary, both standard and alternative measures confirm the overall decrease in wage inequality during this period. However, this trend was far from steady. The Roaring Twenties increased inequality, especially due to the greater increase in the wages of the top 1% and the stagnancy of the bottom. Afterward, the top group suffered from large losses during the Great Depression, but the impact of rising unemployment and the downward rigidity of the wages of the next 4% group prevented a reduction in wage inequality. The decisive turning point happened when the U.S. entered the Second World War. Real wages at the top stagnated, and the gains of the bottom 95% significantly outperformed the rest. This analysis confirms that while the Great Depression caused relevant changes, the war completely transformed wage inequality in the U.S.

4. Understanding the decline in top wages (1939-1949)

4.1 Industry and occupational classifications

To date, top earners have been categorized only by their wage level, but a deeper understanding can be established based on the 1940 and 1950 censuses. Table 3 shows personal characteristics (gender and race), education, industries, and occupations for the groups of interest (all wage earners, the next 4%, and the top 1%). As expected, top wage groups were overwhelmingly composed of white males with higher educational levels even if, as the literature has already highlighted (Goldin and Katz 2010), the college gap decreased over time. In 1940, only 12% of employees had at least one year of college, while among the top 1%, this share was more than 50%. Ten years later, the difference in schooling had declined to approximately two years, as the very top did not experience an increase in the share of graduates.

Table 3. Socioeconomic Characteristics of Wage Earners, 1939 and 1949

Group	Variables	All wage earners		Next 4%		Top 1%	
		1939	1949	1939	1949	1939	1949
Demographic	Male	72%	67%	93%	95%	97%	94%
	Race, white	90%	89%	99%	99%	100%	99%
Education	Years schooling	9.1	9.8	11.5	11.9	12.9	12.9
	College (% with)	12%	15%	35%	37%	51%	52%
Occupation	Professionals	7%	7%	21%	21%	22%	19%
	Managers and officials	5%	5%	25%	22%	51%	46%
	Clerical	12%	12%	11%	6%	6%	4%
	Sales workers	7%	6%	13%	11%	13%	15%
Industry	Manufacturing	27%	26%	27%	32%	32%	32%
	Transport	8%	8%	15%	10%	7%	6%
	Trade	15%	16%	15%	17%	17%	26%
	Finance	3%	3%	8%	5%	14%	8%
	Professional services	7%	7%	12%	8%	11%	7%
	Public sector	5%	6%	10%	8%	6%	5%
	Other	36%	34%	13%	20%	14%	16%

Source: IPUMS, 1940 and 1950 censuses.

The classification through industries does not show such a large difference, as all major activities were represented across the wage distribution. The next 4% group stands out for being overrepresented in the public sector and transportation, while the top 1% of earners are above average in finance, professional services, and manufacturing sectors. As noted in previous studies (Philippon and Reshef 2012), finance declined significantly in importance in both top groups, and trade became more important. The occupational structure highlights a defining cleavage between groups that transformed slowly between the two dates. White-collar workers (professionals, managers, salespeople, and clerical workers) constituted the core of the top 1% and the next 4% groups, but these groups exhibited some subtle differences. In 1939, approximately 22% of the top 1% and next 4% held professional occupations, and a similar trend occurred with sales posts (approximately 13%). Thus, the major difference lies in the relative weight of clerical workers, on the one hand, and managers, on the other. The first were more important in

the next 4% group but were almost residual at the very top. In the top percentile, managers and officials were the predominant group – constituting approximately half of the total – and within this group, jobs were mostly performed by those that the census ascribed as being "managers, officials, and proprietors, not elsewhere classified", which included most corporate officers and executives, as opposed to those who had well-described occupations (i.e., public officials, inspectors, store managers, etc.). Interestingly, both clerical and managerial jobs declined in importance at the top between 1940 and 1950.

Given that top wage groups had such a well-differentiated occupational mix, it is useful to examine previous population censuses to observe how shifts in employment may have impacted the performance of top earners. Table 4 summarizes the occupational structure in the U.S. economy from 1920 to 1950, differentiating between white-collar employment categories and separating between the headline number (all persons gainfully employed) and those working for wages. In the first instance, it is better to focus on this last group, as it is the target population of this paper.

The results reveal a very substantial increase in the number of clerical workers, whose ranks increased in every single period. Trends among managers and professionals are different: the 1920s witnessed the largest growth in their relative numbers, which was then followed by a stagnating job market for these occupations in the 1930s and a renewed increase throughout the 1940s. Goldin and Katz (2010) have argued that white-collar workers were still, at the beginning of the 20th century, a "noncompeting group" who earned higher wages due to the reduced supply of skilled workers. The enormous expansion in higher education and the substantial increase in their ranks from the 1910s onward led to the demise of this special condition. This argument is indeed an accurate description of clerical workers, but it does not seem to apply as clearly to managers and

professionals, at least in terms of timing and causes. Simply put, these groups faced different job markets. Their numbers did not increase as rapidly, and the factors that defined entry into these occupations were affected not only by educational criteria but also by other features, such as tenure (clerical posts could be accessed by workers starting a career, while managerial posts were almost always reserved for more experienced workers), or by formal barriers such as certification (as in the numerous professions that established bar exams). There is a third factor that is often forgotten. Table 4 shows that a significant portion of the managers and professionals were self-employed. In contrast, clerical workers and, to a great extent, sales workers primarily consisted of wage employees.

Table 4. Occupational Structure in the U.S., 1920-1950.

Panel A: All labor force (in thousands)

Year	Total	Professional, Technical	Managers, Officials, and Proprietors	*Management, main occupations*	*Managers, officials, and proprietors (n,e,c,)*	Clerical and Kindred	Sales workers	Other
1920	40,032	2,237	2,671	390	2,281	3,235	1,979	29,910
1930	47,477	3,269	3,554	463	3,091	4,195	3,134	33,326
1940	51,435	3,518	3,862	618	3,245	5,216	3,192	35,647
1950	61,056	5,194	5,221	725	4,496	7,410	4,156	39,075

Panel B: Wage earners (in thousands)

Year	Total	Professional, Technical	Managers, Officials, and Proprietors	*Management, main occupations*	*Managers, officials, and proprietors (n,e,c,)*	Clerical and Kindred	Sales workers	Other
1920	29,290	1,697	1,052	330	722	3,196	1,700	21,646
1930	37,283	2,674	1,670	413	1,257	4,163	2,853	25,922
1940	41,421	2,881	1,749	558	1,191	5,170	2,936	28,686
1950	51,105	4,496	2,562	662	1,900	7,327	3,713	33,007

Panel C: Share of wage earners, per occupation and census year

Year	Total	Professional, Technical	Managers, Officials, and Proprietors	*Management, main occupations*	*Managers, officials, and proprietors (n,e,c,)*	Clerical and Kindred	Sales workers	Other
1920	73%	76%	39%	85%	32%	99%	86%	72%
1930	79%	82%	47%	89%	41%	99%	91%	78%
1940	81%	82%	45%	90%	37%	99%	92%	80%
1950	84%	87%	49%	91%	42%	99%	89%	84%

Source: IPUMS.

Notes: Occupational and employment status data derived from the 1% IPUMS samples, with individual weights applied (Ruggles, Steven et al. 2021). They encompass individuals aged 15 years and older who are part of the labor force. Occupational status refers to the

individual's occupation during the census year (specifically, at the last week of March), classified according to the 1950 occupational classification system. Employment status distinguishes between wage or salary earners and all other workers.

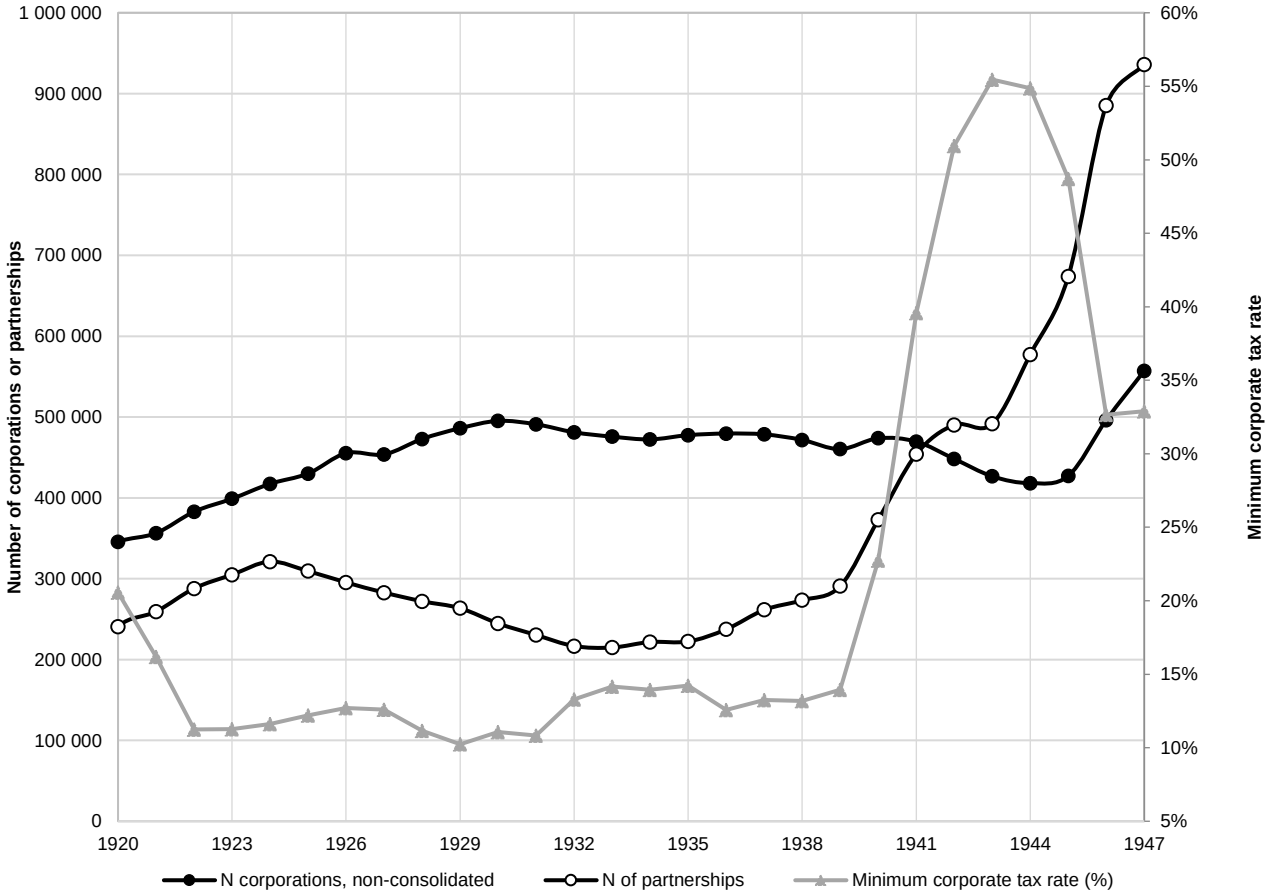
In the remainder of this paper, we explore this important dimension, namely, how the shift in the legal form of businesses affected the number of top-paid workers and the wage distribution. The main argument is that for many managers and professionals, there was a fluid frontier between serving as an employee for a corporation or being a proprietor. This potential shift most obviously impacted the people who could fit in the top wage groups and created a compositional effect. Moreover, this change affected not only relative earnings for the most obvious reasons (entry and exit effects) but also the stock of skills; thus, their returns could vary in both types of firms. Factors such as taxation or regulations would increasingly influence this decision. To conduct this analysis, in the next section, we begin examining the number and pay of corporate managers throughout this period, given their significance within the top 1% share and their potential transition into partnerships. Thereafter, we focus on the upsurge of proprietors during the Second World War, its causes, and its impact on top wage metrics.

4.2 Top corporate managers

To understand the market for corporate managers, corporate tax statistics provide the most comprehensive perspective on the number of active corporations (Figure 3). The trends are very similar to those for the number of managers according to the census, even if the number of managers grew more rapidly due to the increase in the average size of corporations. These series show that there was rapid growth in the employment opportunities of managers during the 1920s, followed by a stagnant market in the 1930s. The most surprising change happened throughout the 1940s, when the number of corporations initially experienced a significant contraction until the post-WW2 boom drove this figure upward again. The explanation for this change most likely lies in the transformation of a considerable number of small corporations into partnerships so that

shareholders could avoid the increase in the corporate tax. These two additional variables have also been included in Figure 3. The average effective corporate tax rate, which had stood at 13% in 1938, more than tripled to 55% by 1943. This shift is a critical factor that has been understudied to date, as it highlights that the number of corporate executive posts was heavily influenced by the decision of businesses to move from one legal status to another.

Figure 3. Number of Corporations and Partnerships Average Effective Corporate Tax Rate (in percentage), 1920-1947

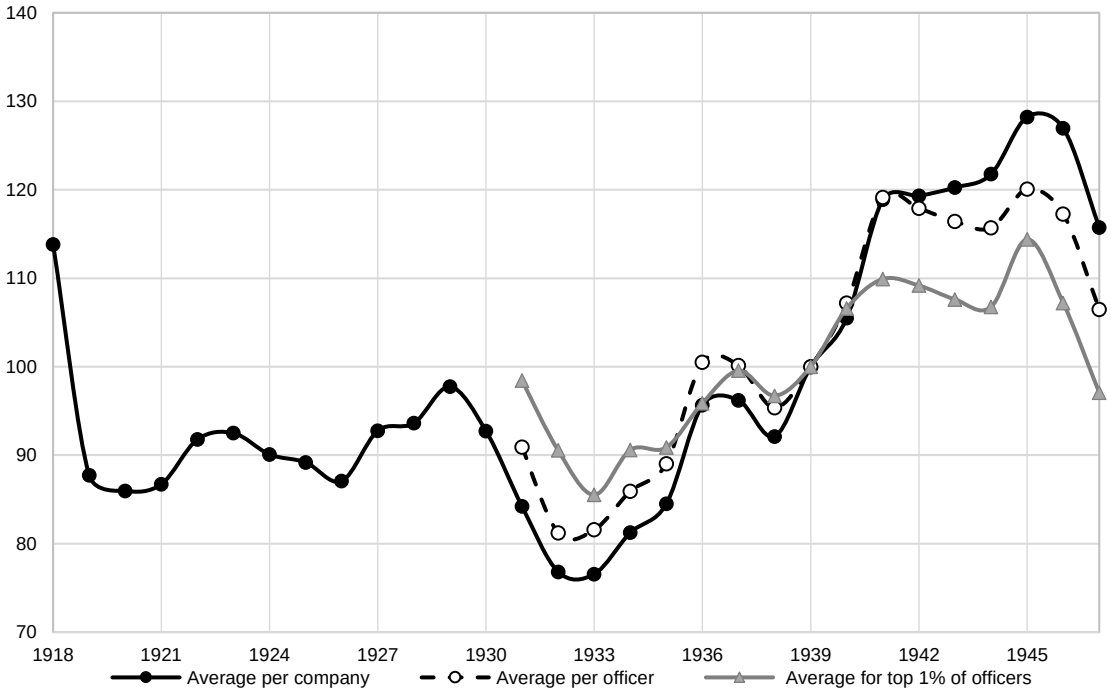


Sources: SOI.

The other metric of interest—compensation of officers—is presented in Figure 4 as the average pay per corporation and per officer. Additionally, from 1931 onward, it is possible to disaggregate these figures to consider the variance in pay among officers. The tabulations of the corporate tax report the compensation for different groups of corporations per asset size. Then, this evidence is matched with the information provided

on earnings at the individual level among officers in the Survey of American Listed Corporations (SALC), a unique dataset of the Securities and Exchange Commission (1940). From these records, we focus on the top 1% of corporate managers (roughly speaking, a group of approximately 7,000-10,000 persons).

Figure 4. Average Real-Officer Compensation, 1939=100
Average per company, per officer and among the top 1% highest-paid



Sources: Estimates based on SOI and SALC data

The three different metrics point in the same direction, even if sometimes they tell a slightly different story. During the 1920s, corporate pay increased steadily, and this rise was presumably more notable among the highest-paid officers since the average is driven downward due to the compositional effect arising from the proliferation of numerous small firms. Subsequently, from 1930 to 1933, compensation decreased substantially, followed by a recovery in subsequent years. These two distinctive trends coincide with the evolution in the average wage of the top 1% and support the idea that managers and officers were a group of decisive importance at the top.

The key question remains understanding the 1940s, the most critical juncture for top earners and corporate officers. Corporate tax records show that average pay among officers increased from 1938 to 1940-1, although this increase largely reverted in subsequent years among the top-paid officers. Overall, in the benchmark scenario, officer compensation stagnated (or slightly declined), following a similar evolution to the results of Frydman and Saks (2010). This subpar evolution has been explained by different mechanisms. Frydman and Molloy (2012) explain the fall in corporate compensation during the Second World War period as a result of declining returns to firm size and growing union power within companies. However, Piketty and Saez (2003) give more credit to the strict regulation of executives' compensation by the National War Labor Board and to the unprecedented surge in the marginal personal tax rate (Piketty 2014).

Building upon these findings, our research embraces a broader perspective. If the aim is to explain the performance of the top 1% of wage earners—a small group but still one made up of hundreds of thousands of workers—one cannot make strong claims by only looking at the highest paid executives of listed corporations, which constituted a far smaller group that would mostly fit in the top 0.01% (i.e., a few thousand persons). To understand the evolution of the top 1%, it is mandatory to consider a broader group of corporate officers. Furthermore, to ponder the impact of this group on wage inequality, two factors must be addressed: their comparative pay levels and their aggregate count, which is directly influenced by the establishment or dissolution of corporations. The variability in the number of corporations is, in turn, directly related to the economic cycle and to changes in the legal structure of businesses. Previous studies have primarily focused on top executive compensation levels, possibly assuming relative stability in the number of large corporations over time. However, as highlighted earlier in this section,

these constant levels cannot be assumed when examining a wider group of firms and the U.S. economy at large.

To analyze the relationship between executive compensation and top wages, we start with a basic decomposition expressed as:

$$S_t^{p=99} = \left(s_{o,t} * F_{o,t}^{p=99}\right) + \left(s_{non,t} * F_{non,t}^{p=99}\right) \qquad (2)$$

where *S* is the wage share above percentile 99 in year *t*. This share can be decomposed by computing the relative share (*s*) of the aggregate compensation derived by two major occupational groups in the economy: officers (*o*) and nonofficers (*non*). These two wage bills are then weighted to consider the share (*F*) that flows to workers above the defined percentile for the top 1% (*p=99*) following the standard definition, that is, defined in relation to the aggregated wage-earning population of each year. With the underlying data, one can measure the number of earners in the 99th percentile who are officers and nonofficers. Moreover, the U.S. corporate tax statistics data are sufficiently detailed to disentangle this contribution of officers' compensation according to the size of the firms. In the interest of simplicity, companies are grouped into three major categories that broadly refer to small, medium-sized, and large corporations.[1]

[1] These three groups are defined according to their total assets: under 1 million, between 1-10 million assets, and over 10 million dollars. To contextualize these figures, in 1938 the median size of a listed corporation stood slightly above 5 million dollars in assets, and only 17% of corporations had assets under one million dollars.

Figure 5A. Number of Corporate Officers in the Top 1% and Relative Share by Asset Size, 1931-1947

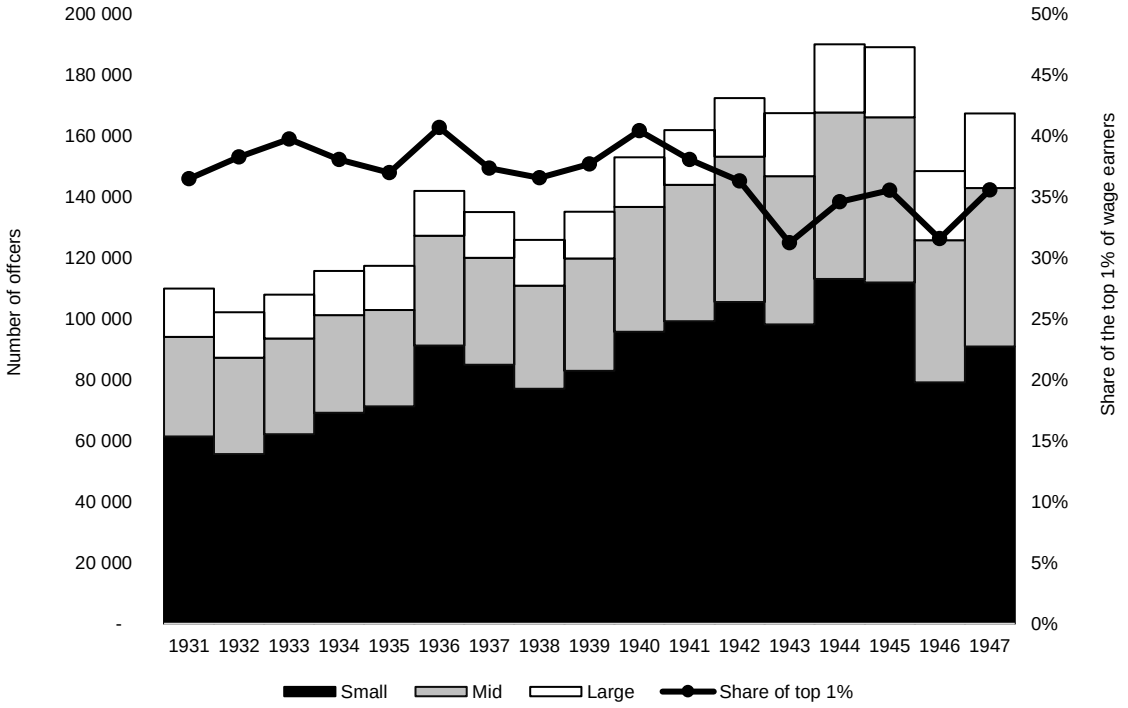


Figure 5B. Contribution of Officers' Compensation to the Top 1% Wage Share by Asset Size, 1931-1947

Sources: SOI, NIPA, Survey of American Listed Corporations (SEC) and Corporate Tax Statistics (Bureau of Internal Revenue).
Notes: Small corporations are defined as having assets under 1 million dollars, mid-sized corporations as having assets between 1 and 10 million, and large corporations as having assets greater than 10 million. Officers' compensation share is calculated over the wages and salaries reported in the NIPA.

Figure 5A and 5B present these results and highlight three relevant facts. First, during the pre-WW2 era, approximately 120,000 officers were in the top 1%, or close to 40% of the total group.[2] These officers contributed approximately 2.6 percentage points of the aggregate top 1% wage share, which typically stood at approximately 9%. The second main finding is that the top 1% officers' compensation derived from small and mid-sized corporations was far from trivial. For example, in the late 1930s, these companies contributed almost 90,000 of all the officers in the top 1% and 2 points of the 2.6 percentage points to the aggregate of officers' compensation in the top. The third finding relates to the shock that occurred during the Second World War. During these years, the contribution of officers' compensation to the top 1% wage share declined abruptly to 1.8% and later stabilized at approximately 2% in 1947. Moreover, if one decomposes the contribution according to the size of the corporation, the smallest ones declined steadily throughout this period, while the increase in the largest (both in terms of the number of officers and compensation) was not enough to compensate for this decrease.

Measuring changes in the relative weight of officers from small, medium, and large corporations is still a very crude metric. These figures omit shifts between groups, as typically the increase in the size of corporate balance sheets would move firms to a higher group. More importantly, they also neglect the exit and entry effect due to the net creation of corporations. Since we are especially interested in this last factor, we estimate the net increase (or decrease) in the number of corporations on a yearly basis for each group of

[2] The reader may note that these numbers are lower than those previously reported in the decennial censuses for salaried managers because in this section we focus only on corporate officers (the chairperson, president, and other executive positions). Nevertheless, results are very similar. In the 1940 census, managers represented c. 50% of the top 1%, while in this estimate, corporate officers represented approximately 38-40% of the total. The difference between the two shares can be mostly attached to those managers that were not corporate officers.

firms and their potential impact on the observed top 1% share. We perform this exercise by taking the number of corporations in an asset group (say, those with assets between 1 and 10 million dollars) and extrapolating the average change in their assets to factor in how many firms moved to another size group. In the second step, we compare those numbers with the figures reported in the corporate tax statistics and assume that the difference relates to the net creation (or destruction) of corporations. The details of these estimates are presented in the Appendix, but it is important to remember that the aggregate number of new corporations on a yearly basis is known from the statistics and that with this procedure, we simply distributed the net creation of corporations per asset bracket.

Figure 6A shows the change in the number of officers within the top 1%, and Figure 6B shows the contribution of these changes to the top 1% wage share. During the Second World War, the overall decrease in the number of corporations reduced the number of officers at the very top, and this change was largely dictated by changes among small firms. If one measures this shift in terms of the contribution to the top 1% wage share, the disappearance of small corporations explains a significant part (a fall of 0.1 percentage point) of the overall decline in the officers' compensation to the top wage share (which fell by 0.7 points). Nevertheless, one could argue that these changes were reverted by the equally significant increase in the number of small corporations in the postwar boom (1946 and 1947), which increased the number of corporate officers in the top 1% and their relative contribution to its wage share.

Figure 6A. Change in the Number of Officers among the Top 1% due to the Net Creation of Corporations by Asset Size, 1939-1947.

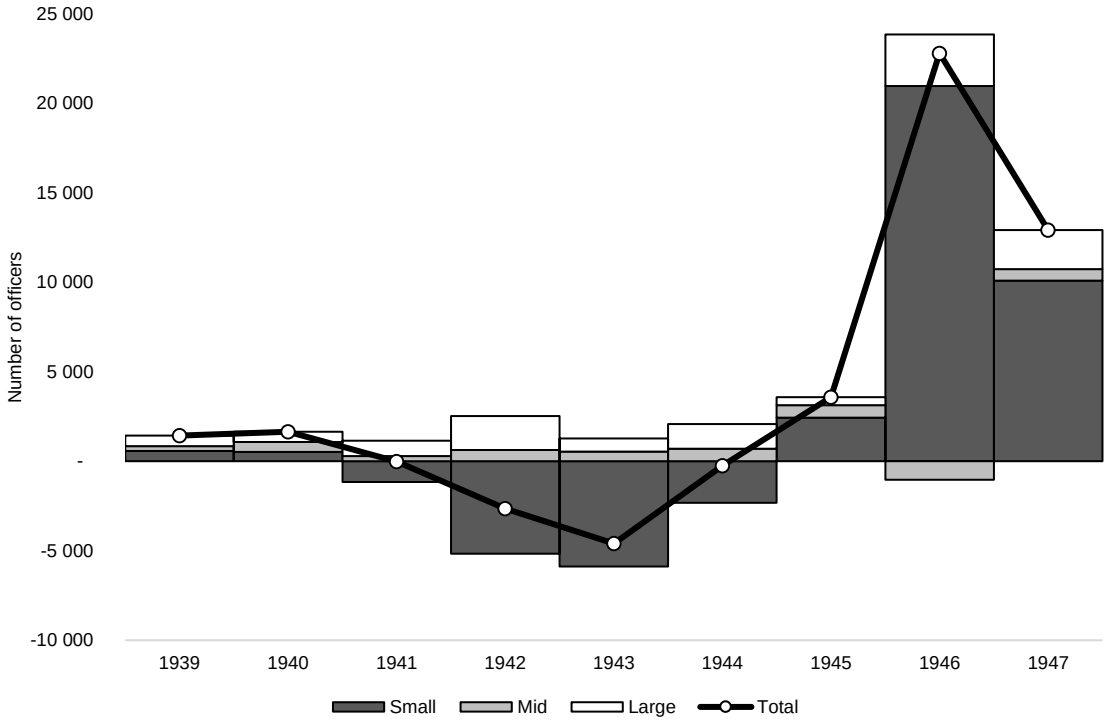
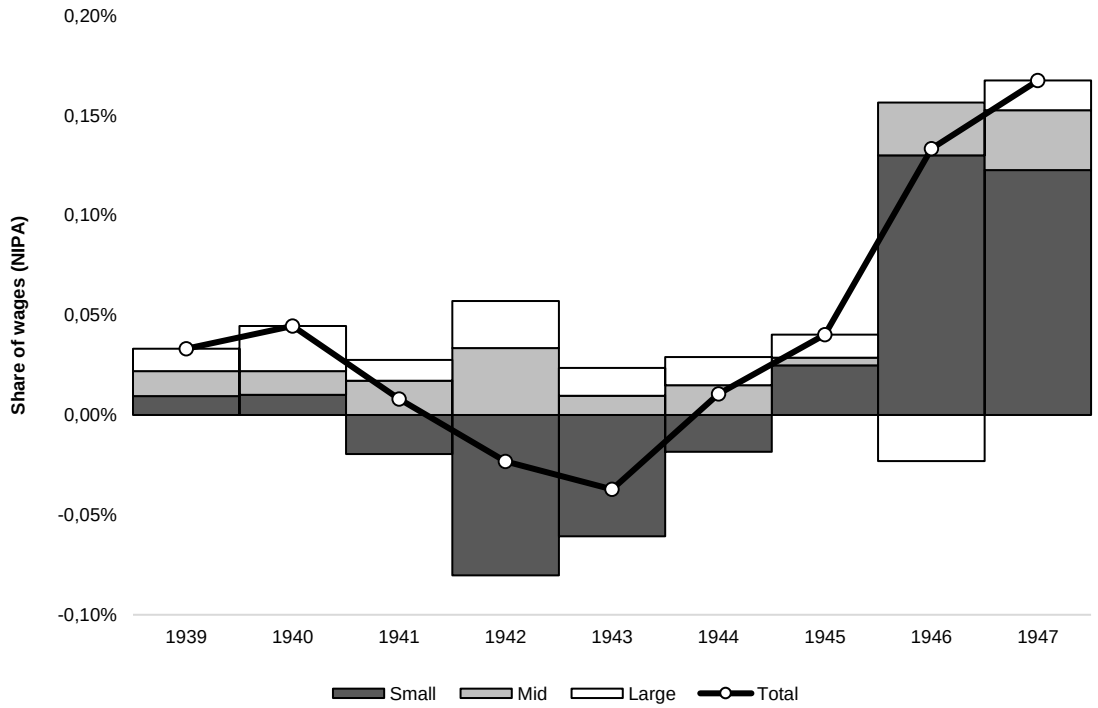


Figure 6B Change in the Officers' Compensation Component in the top 1% wage share due the Net Creation of Corporations, by Asset Size.



While this is indeed the observed trend, this path omits one key fact highlighted at the beginning of this section, namely, that the number of partnerships increased dramatically from 1939 to 1949. These firms tripled in number, while the figures on corporations increased at a much lower rate (by 25%). Thus, one can reasonably ask what would have happened to the top 1% wage share and to the relative contribution of officers' compensation if those newly created partnerships had opted to form a corporation. To carry out this exercise, we follow a conservative estimate, that is, what would happen if the newly created partnerships employed the same number of officers as small corporations (i.e., two officers, as in corporations with assets under 1 million dollars) and had the same relative pay levels both in terms of the average and the relative dispersion of earnings.

Figure 7 presents these results by comparing the observed number of officers in the top 1% and the counterfactual series based on these assumptions. The key finding is that the additional number of officers who would have entered the top 1% would have been substantially larger, almost doubling the observed number at the end of this period. That is, without the new preference for partnerships, there could have been almost 300,000 corporate officers at the top as opposed to the observed 170,000. With the aim of understanding the ultimate effect of this change in employment status on inequality, this transition between 1939 and 1949 is thoroughly analyzed in the following section.

Figure 7. Number of Corporate Officers in the Top 1, 1939-1947

Sources: Own estimates based on NIPA, SOI and SALC

4.3 Self-employment effect on the wage distribution

The 1940s witnessed an unprecedented shift in business preferences regarding their legal structure, prompting thousands of managers and other workers to alter their status from salaried employees to self-employed individuals. In this section, we estimate the actual earnings of proprietors to quantify the impact of this institutional shift on wage inequality and to understand its underlying causes. Studying the self-employed in the United States is a challenging task because proprietors constitute a minority within the nonfarm working population, and their earnings are typically not recorded in well-established sources, such as manufacturing censuses or labor force surveys. Moreover, even with accessible data, distinguishing between proprietors' income attributed to labor and capital is puzzling. Consequently, virtually all researchers have overlooked this group during this era.

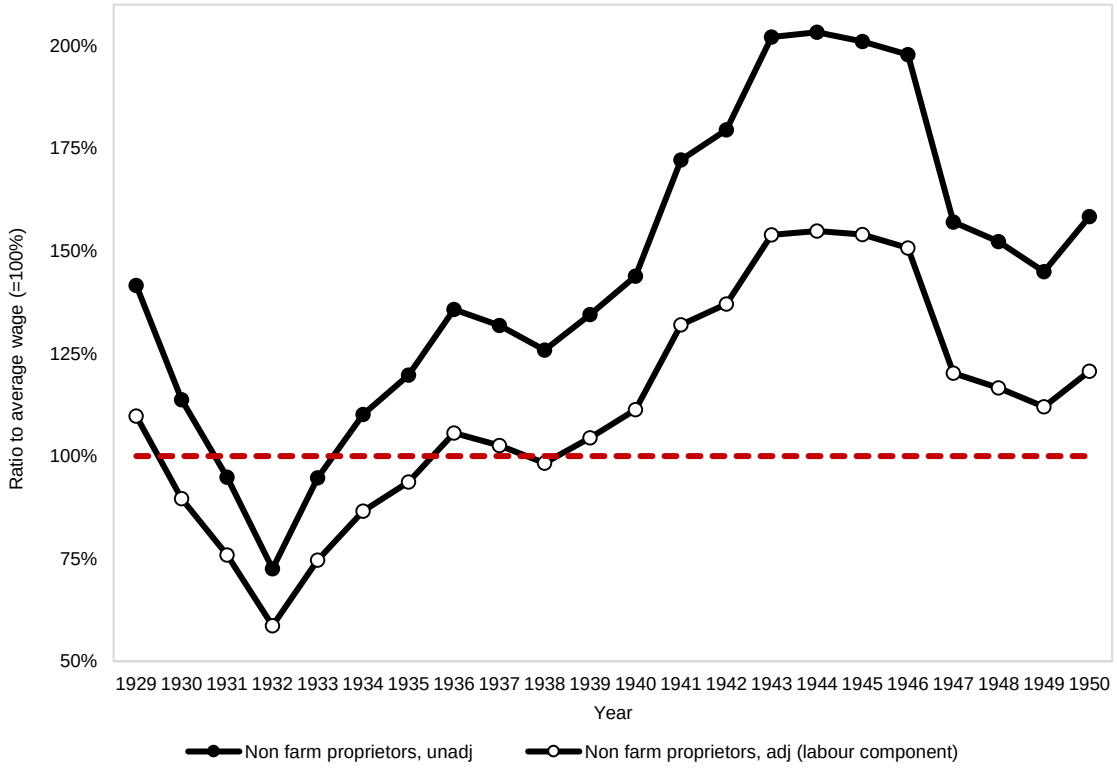
To calibrate the importance of high-earning proprietors, we follow a similar logic as in the previous sections, integrating national accounts, census records, and tax statistics. The

NIPA provides a yearly series on the number and aggregate earnings of self-employed classified by major industries. Census microdata offer detailed information on demographics (gender, age, etc.), education, and occupations. However, individual information on earnings appeared for the first time in 1950, with the usual top-coding limit, but was absent from the 1940 census. We turn to tax records to fill the gap in terms of proprietors' earnings before 1949 by using SOI reports, which include comprehensive data on the income of sole proprietors and partners by industry for some specific years. In this section, we exclude agricultural workers, as tax data are unreliable for measuring farmers' incomes (United States. Office of Business Economics 1953).

SOI also provides a meaningful way to separate proprietors' income between labor and capital, which is a matter of critical importance and a persistent challenge for economists. Among contemporary studies, Smith et al. (2019) match the accounting microdata of closely held businesses (S-corporations and partnerships) and, after factoring in industry and size effects, conclude that almost 75% of their profits constitute a "disguised wage" to owners. Saez and Zucman (2019) prefer to estimate factor shares in these firms by measuring the return to capital by taking the ratios of similar listed corporations. In this paper, we follow this logic by using the unique 1953 SOI report on partnerships, which included information on both income and balance sheets at the industry level for the first time. We calculate the rate of return of partnerships by dividing the net income by the capital of partners. In the second step, we compute the capital share of partnership profits by assuming that it should be equal to the return on equity (ROE) for corporations in the same industry. The residual share of profits represents the labor return. The results of this exercise are presented in the Appendix. Overall, the results align with the conventional assumption that labor constitutes approximately 70% of proprietors' income but with substantial variations across industries, ranging from 47% in sectors more intensive in

physical capital and natural resources (e.g., mining) to up to 87% in those more dependent on human capital (e.g., professional services). Given their consistency, these labor/capital ratios are extended to all proprietors (both sole proprietors and partners) for the preceding years.

Figure 8. Relative Earnings of Nonfarm Proprietors, 1929-1950



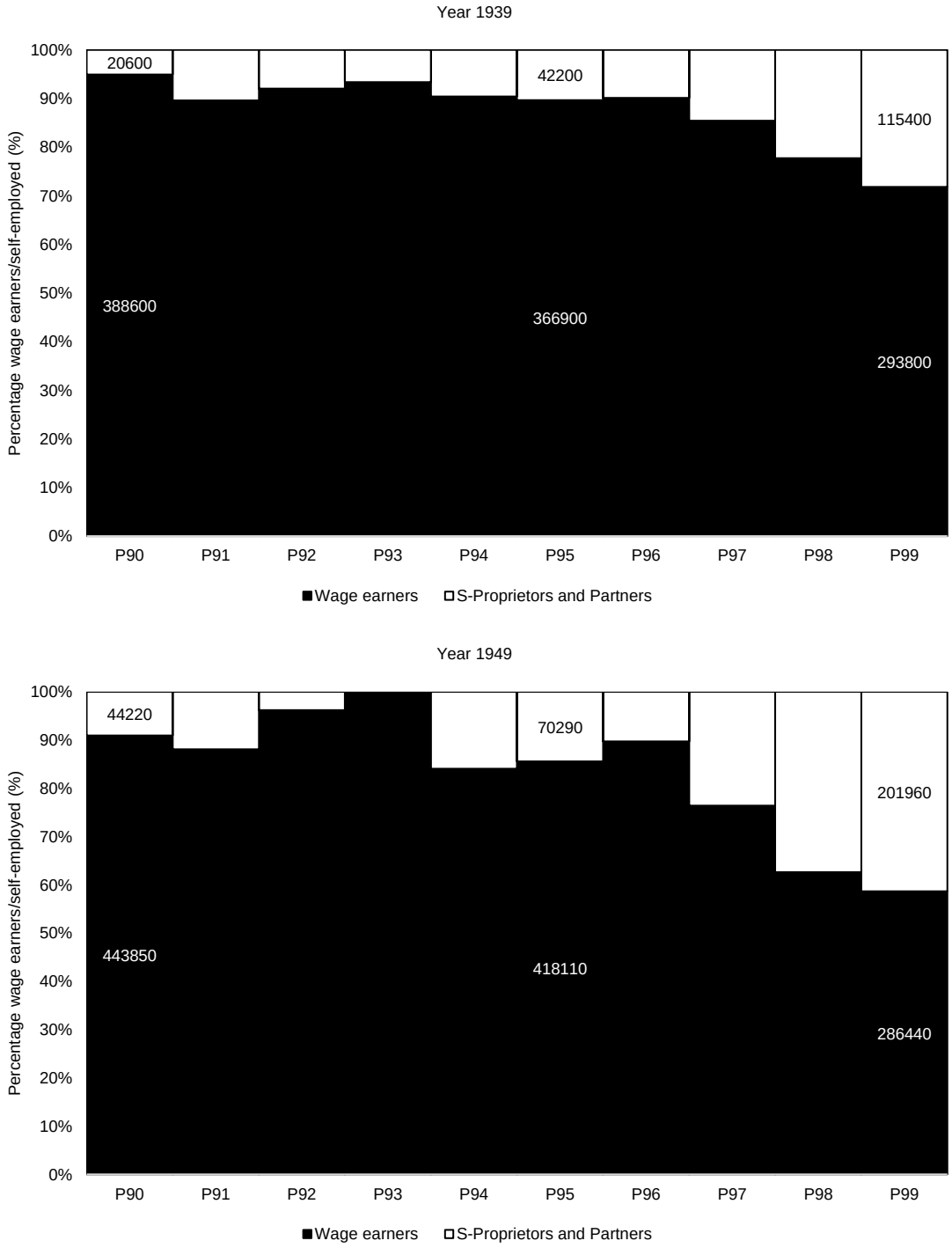
Sources: NIPA

Notes: Both series are calculated by dividing nonfarm proprietors' incomes by the wages and salaries of nonagricultural workers

The analysis starts with a comparison of the aggregate earnings of nonfarm proprietors with those of wage earners. The NIPA series shows that proprietors' incomes were on average higher than those of wage-earners, but if one adjusts for the labor component of the first, the labor incomes of the two groups in the pre-WW2 era were relatively close (Figure 8). Nevertheless, the war constituted an enormous demand and supply shock that resulted in the transformation of the labor composition of the employed population, a change in relative prices, and a shift in the legal form of businesses. All these factors could drive proprietors' earnings to a much higher level. In fact, even if this income

premium was reduced in the postwar years, proprietors' earnings were still substantially higher than before 1940.

Figure 9. The Top Decile of the Non-Farm Working Population by Percentile and Employment Status, 1939 and 1949



Source: SOI (1939) and IPUMS (1940 and 1950).

Note: Figures on the number of workers are provided for the 90th, 95th and 99th percentiles.

To ponder the relative weight of high-earning proprietors, we estimate a new labor earnings distribution that incorporates both wages and the labor income of proprietors. According to the 1950 census, merging wage-earners and proprietors is a straightforward exercise that only requires imputing incomes above the top-coded level. For 1940, we use the wage distribution from the census and proprietors' incomes from SOI tabulations following a similar logic. For both years, we applied the previously computed labor/capital ratios at the industry level to proprietors' income, thus ensuring that only labor incomes are incorporated in the analysis.

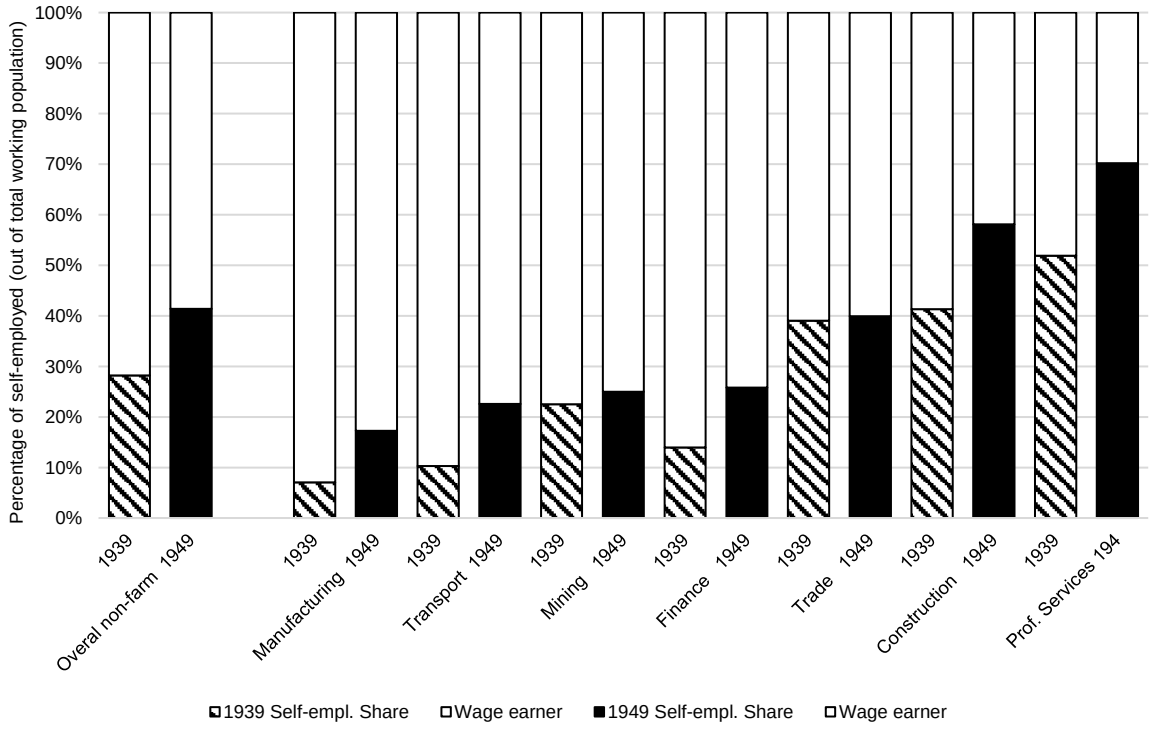
Figure 9 depicts the percentage of wage-earners and proprietors within the newly computed top decile of the labor earnings distribution in both years.[3] The results clearly indicate that proprietors were overrepresented at the very top of the labor income distribution. Self-employed individuals represented approximately 14% of the nonfarm workforce in both years, and their share hovered between 10% and 12% of the earners placed between the 90th and the 96th percentiles. In addition, their relative weight hardly increased in these percentiles after the Second World War. Between 1939 and 1949, the expansion of proprietors was thus concentrated in the top 3% of the wage distribution and more particularly in the top percentile.

Figure 10 represents the share of proprietors in the top 1% of the labor income distribution in both benchmark years, thereby confirming the significant change in the relative weight of the two groups. In the 1939 earnings distribution, proprietors constituted 28% of the top 1% of the nonfarm working population, but ten years later, this share experienced a substantial upsurge to 41%. Furthermore, when these shares are decomposed at the

[3] This figure utilizes the 99th percentile corresponding to the new labor earnings distribution. These percentiles increase from $5,246 and $8,950 to $5,869 and $11,249 in 1939 and 1949, respectively.

industry level, proprietors constituted, in some cases, most workers at the top of the distribution. This is the case, for example, in professional services before the war. In the post-war years the increase in the share of self-employed individuals occurred across all industries. Indeed, in those sectors that started with a high share of wage employment—such as manufacturing, transportation, and finance—the percentage of highly paid proprietors increased more rapidly, doubling their share in this ten-year period.

Figure 10. Percentage of Wage Earners and Self-Employed within the Top 1% of Labor Earnings, 1939 and 1949.



Source: SOI and IPUMS (1940 and 1950 census).
Notes: Only workers with positive wages or self-employed incomes were included. In 1949, workers with dual income (both wages and self-employed income) were categorized based on their self-classified employment status.

The rise in proprietors' incomes and their growing weight at the top of the distribution did not go unnoticed to the most acute observers of the time. One argument underscoring the growing significance of high-earning proprietors, as succinctly provided by Simon Kuznets in his seminal work, posits that wage-earners and proprietors would react differently to a sustained increase in prices (Kuznets 1953:73). Furthermore, Kuznets and

Friedman (1945) argued that the earnings of independent professionals (e.g., lawyers, medical doctors, accountants, etc.) were higher to compensate for the years of training, to account for the heightened risks and volatility in their income trajectory, and due to significant barriers to entry. These factors could make the 1940s a specifically favorable period for proprietors, pushing a greater number of them to the top of the distribution. A different kind of argument was provided by legal experts on tax and corporate matters, who noted that the unprecedented rise in corporate tax rates in the 1940s caused small corporations to bear a much heavier tax burden than income-equivalent proprietors. Stigler (1956, 55) noted that sole proprietorships and partnerships had a clear tax advantage over corporations in the booming service industry. Senator Humphrey observed that, in 1950, the hike in corporate taxes led to "flagrant war-boom tax avoidance", producing a jump in the number of partnerships to evade excess-profit taxes and to lower the individual income tax burden (Humphrey 1994, 633). The evidence presented in the previous section already supports the shift in preferences toward partnerships and the relative reduction in the number of corporations. An extension of this argument is that high-earning proprietors and salaried workers were not necessarily two mutually exclusive groups. In an economy dominated by wage employment, businesses could change their legal status for tax reasons, and a small but significant number of highly paid workers would switch their employment status.

One way to clarify this issue is to understand the determinants of self-employment and the returns to labor using 1950 census microdata. Specifically, we employ a probit model to explore the variables affecting the probability of being self-employed over being a wage earner and to test whether taxes play a role in defining this status. As the census does not provide information on individual tax payments, we use gross earnings as a

proxy for taxes.[4] We also control for covariates such as education, gender, marital status, family income, industry, and occupation. Considering the high incomes garnered by the self-employed, it is plausible that these individuals may be earning a premium following Kuznets' argument. If this were the case, the relationship between gross earnings (or the tax base) and self-employment might suffer from a simultaneity problem (reverse causality), leading to biased and inconsistent parameter estimates when using a conventional probit model. To address this issue, we adopt a simultaneous equation methodology. Specifically, we estimate a two-stage least squares model (2SLS) following Amemiya (1978), Heckman (1978) and Maddala (1983), where instruments are applied to substitute the two endogenous variables (i.e., gross income and the self-employment dummy).

Equations 3 and 4 present the second stage of the baseline model. Equation 3 represents a Mincer equation, where earnings (in natural logarithm) are explained by education, age, gender, and a binary variable with a value of 1 when the worker is self-employed and 0 otherwise. If proprietors genuinely represent a noncompetitive group, the coefficient of this last variable (γ_1) should be positive and significant, indicating an earnings premium. On the other hand, Equation 4 is estimated through a probit model aiming to assess the effect of demographic variables (education, gender, and age), industry, occupation and earnings on the likelihood of becoming self-employed; other control variables, such as family income and marital status, are included in the alternative models. If taxes were an incentive to become a proprietor, as aforementioned, the coefficient of earnings (γ_2) is expected to be positive and significant. The instruments for the endogenous variables are

[4] In the following regressions, we use both the overall income earned by proprietors, which is the best proxy for taxes since it represents the tax base, and the labor income using the previously estimated labor/capital ratios, as a robustness check.

estimated using all the exogenous covariates in the first stage. Subsequently, the predicted instruments are integrated into the second-stage equations.

$$lnE_i = \gamma_1 \widehat{Self_i} + \beta_1\, Educ_i + \beta_2\, Age_i + \beta_3\, Age_i^2 + \beta_4\, Gender_i + \varepsilon_i \quad (3)$$

$$Self_i = \gamma_2 \widehat{lnE_i} + \beta_1\, Educ_i + \beta_2\, Age_i + \beta_3\, Gender_i + \beta_4\, Capitalind_i + \beta_5\, Clerical_i + \beta_6\, Manager_i + \varepsilon_i \quad (4)$$

Table 5 presents the benchmark results of the second-stage models with corrected standard errors (the first stage is shown in Table 9 in the Appendix). The coefficients of the Mincer equation (Column 1) are significant and show the expected values. Both education and age contribute positively to earnings, although the age effect is nonlinear and diminishes over time. In addition, the data confirm a substantial gender pay gap, with women earning, on average, 67.6% less (approximately $1,673).

The most relevant point for this study is the absence of an earnings premium for self-employed individuals. Proprietors received 7.6% less (or $188) than equivalent wage-earners on average. This coefficient, which decreases to 3.8% if other control variables are included (Column 3), remains negative and robust to all the specifications discussed below. This result invalidates Kuznets' argument regarding barriers to entry and a risk-return tradeoff and demonstrates that the higher earnings of proprietors resulted from their advanced educational achievements and accumulated experience. Table 6 indeed shows the descriptive statistics for the variables included in the model for both self-employed and wage-earning workers. The table confirms that proprietors had much higher levels of education (the percentage holding at least a college education was twice as large) and had more experience (eight more years on average). Gender differences were also significant, with relatively fewer women being self-employed.

Table 5. Benchmark Models (Second Stage): Analysis of Determinants of Labor Returns and the Self-Employment Decision

	Model 1		Model 2		Model 3		Model 4	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VARIABLES	Mincer eq.	Probit eq.	Mincer eq.	Probit eq.	Mincer eq.	Probit eq.	Mincer eq.	Probit eq.
I_self	-0.076***		-0.039***		-0.038***		-0.123***	
	(0.004)		(0.004)		(0.004)		(0.004)	
I_lnE		0. 291***		0.219***		0.324***		0.291***
		(0.024)		(0.030)		(0.024)		(0.025)
Educ	0.062***	0.001	0.061***	-0.003	0.066***	0.003	0.062***	0.001
	(0.001)	(0.002)	(0.001)	(0.002)	(0.001)	(0.002)	(0.001)	(0.002)
Age	0.107***	0.0152***	0.105***	0.015***	0.106***	0.015***	0.108***	0.015***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)	(0.001)	(0.000)
Age2	-0.001***		-0.001***		-0.001***		-0.001***	
	(0.000)		(0.000)		(0.000)		(0.000)	
Gender	0. 676***	0.128***	0.657***	0.098***	0.680***	0.132***	0.684***	0.131***
	(0.005)	(0.021)	(0.005)	(0.021)	(0.005)	(0.021)	(0.005)	(0.021)
Marital				0.134***				
				(0.017)				
Capitalind		-0.462***		-0. 454***		-0.482***		-0.463***
		(0.013)		(0.013)		(0.013)		(0.013)
Clerical		-0.871***		-0.850***		-0.882***		-0.873***
		(0.031)		(0.031)		(0.030)		(0.031)
Manager		1.276***		1.28***		1.211***		1.304***
		(0.015)		(0.015)		(0.015)		(0.014)
Faminc				-0.001**				
				(0.000)				
Constant	3.876***	-5.273***	4.018***	-3.868***	3.916***	-4.581***	3.736***	-4.308***
	(0.024)	(0.149)	(0.023)	(0.195)	(0.024)	(0.145)	(0.024)	(0.150)
Observations	145,961	145,961	145,961	145,961	148,325	148,325	145,961	145,961
R^2	0.2551		0.2537		0.2668		0.2461	
Pseudo R^2		0.2388		0.2416		0.2431		0.2388

The corrected standard errors are in parentheses.

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

Notes: All estimated models follow the 2SLS approach outlined by Amemiya (1978) and Maddala (1983) utilizing data from the individuals in the 'sample-line' of the 1% IPUMS sample corresponding to the 1950 census. In the first stage, each endogenous dependent variable is regressed on all exogenous variables. Subsequently, instruments are predicted for these variables and introduced in the second-stage equations (see Appendix). The second-stage model consists of two simultaneous equations. The first estimates a Mincer equation using OLS, and the second estimates the likelihood of becoming a self-employed worker through a Probit model. The models target the nonagricultural U.S. economy. Only workers with positive wages and/or self-employed incomes were included. Models 1 and 2 exclude top-coded income (earnings exceeding $10,000) and utilize original gross earnings. Model 4 incorporates top-coded incomes and uses original gross earnings. Model 4 again excludes top-coded income, but it adjusts for self-employed income to account for labor income.

On the other hand, the probit model (Column 2) indicates that in 1950, factors such as gross earnings, age, male gender, managerial occupation, and marriage (an additional control variable in Model 2) increased the likelihood of becoming a proprietor. In contrast, employment in a capital-intensive industry (i.e., mining, manufacturing, and construction), engaging in clerical tasks, and having higher family income (an extra control variable in the second model) diminish this probability, while educational attainment appears to have no discernible effect. Accordingly, the model confirms that gross earnings, and consequently the taxable base, were a relevant factor in the decision to become self-employed. These results are robust to including top-coded incomes (Model 3), adjusting self-employed earnings to account only for labor earnings (Model 4), and narrowing the focus to managerial occupations (Table 10 in the Appendix). Additionally, parallel models have been estimated for each specific industry, consistently finding a positive correlation between gross earnings and self-employment across all sectors except mining (Table 10 in the Appendix).[5]

[5] While tax incentives were uniform across industries, the lack of a significant effect in this specific sector most likely suggests that the transition to self-employment might have been not feasible or excessively costly in this industry.

Table 6. Main Features of Nonagricultural Workers by Employment Status, 1939/40 and 1949/50

		1940	1950
Gainfully employed	Wage earners, share	85%	87%
	Proprietors, share	15%	13%
Average income, 1939 dollars	Wage earners	1,275	1,640
	Proprietors	1,584	2,284
	Proprietors/Wage earners	1.2	1.4
High school or higher education (% of total)	Wage earners	33%	42%
	Proprietors	35%	46%
Completed college (% of total)	Wage earners	6%	7%
	Proprietors	11%	13%
Experience (years)	Wage earners	22.7	23.0
	Proprietors	30.8	30.4
Female share (% of total)	Wage earners	29%	32%
	Proprietors	17%	16%

Source: IPUMS (1940 and 1950) and NIPA.

Notes: All variables, except those on income, refer to 1940 and 1950. Wages and proprietor incomes are reported relative to those of the previous year (1939 and 1949). Education measures have been estimated using the share of workers with this education level out of the overall full-time wage or self-employed workers. Experience has been calculated as age minus years of education minus six. The table excludes agricultural workers.

This evidence is helpful for reevaluating the position of top earners and wage inequality during the Great Compression. Taking this broader definition of labor earnings, we re-estimate the average income of the major groups of interest—the top 1%, next 4%, and bottom 95%—as well as their corresponding shares. Table 7 compares these results with the previous estimates that were built by considering only wages and the wage-earning population. The inclusion of self-employed individuals provides two significant changes. First, due to strict regulations, the notion that top-earning workers barely increased their incomes over the 1940s is challenged. In the standard analysis that includes only wages, the top 1% experienced a loss in real terms, while the next 4% experienced sluggish growth in earnings. In contrast, when adjusting for proprietors' labor incomes, both

groups experienced an additional increase of approximately 20% in their cumulative real earnings growth (with the top 1% growing 12.2% and the next 4% increasing 27%). This change in magnitude is relevant for reconciling the fortune of top earners with the boom in the U.S. economy and the substantial increase in labor productivity over the decade.

Table 7. Labor Inequality Measures for Incomes Reported in 1939 and 1949.

	Wages			Wages and labor income of the self-employed		
Panel A: Average income (nominal dollars)						
	1939	1949	Real cumulative growth	1939	1949	Real cumulative growth
Bottom 95%	909	2199	71%	909	2243	76%
Next 4%	3305	5912	8%	3560	7054	27%
Top 1%	10835	17378	-11%	11896	21816	12%
Panel B: Shares						
	1939	1949	Percentage change 1939-49	1939	1949	Percentage change 1939-49
Bottom 95%	78,3%	83,6%	7%	77,1%	81,3%	5%
Next 4%	11,9%	9,5%	-20%	12,3%	10,4%	-16%
Top 1%	9,7%	6,8%	-30%	10,6%	8,3%	-21%

The second major conclusion is helpful for calibrating the extent of the Great Compression. Overall, by focusing only on wage workers, scholars have significantly overestimated the decline in inequality during this critical period. Including the self-employed does not change this trend but does significantly reduce its magnitude. The decline in the top 1% share is reduced by 0.6 percentage points (that is, from a reduction of 30% to 21%), and the next 4% also witnessed a significant change of 0.5 percentage points (or from a decrease of 20% to 15%). This evidence ultimately highlights the exceptional performance of workers at the bottom—rather than the poor performance of the top—as the ultimate cause of the Great Compression.

5. Conclusion

This paper examines wage inequality by using tax records, censuses, and other administrative data, analyzing the critical period spanning from the conclusion of the First

World War to 1949. The results confirm the extent of a large decline in wage inequality during the 1940s –the Great Compression– identified by previous generations of scholars but provide a more comprehensive account of the secular dynamics and the causes of this sudden shock.

The results of our analysis show that high levels of wage inequality were a constant feature of the interwar period (1918-39), with no evidence of a long-term levelling trend. Tax records indicate that different groups across the wage distribution performed differently during the business cycle. The top 1% was the group that fared better during the Roaring Twenties but later experienced a substantial contraction in earnings during the Great Depression. In contrast, the next 4% exhibited stronger wage stickiness, and their relative situation improved during times of crisis (1920-1 and 1929-33). The share of the bottom 95% stagnated during the 1920s and declined markedly during the first years of the Depression due to the impact of unemployment and wage cuts. In the 1940s, these trends were strongly reversed, as the bottom 95% significantly outperformed all the other groups.

For a better understanding of this process, tax records are supplemented with other sources that allow us to reconcile top wage groups with specific occupations. This analysis illustrates that officers and managers of small and mid-size corporations played a significant role in the decline of the top 1%. This was mainly due to a shift in preferences regarding the legal status of companies that started to take place from 1938 onward, when an unprecedented rise in corporate tax rates fueled a shift toward the formation of partnerships. This change thereby altered the employment status of the highest-paid salaried workers, turning many highly skilled workers (managers, officers, and professionals) into proprietors. When factoring in this process, we find that the

performance of top earners was better than previously noted, as they adjusted their incomes upward despite war labor regulations. Moreover, this paper demonstrates that this shift overstates the decline in wage inequality during the Great Compression. The drops in the top 1% and the next 4% earnings share are reduced by 0.6 and 1 percentage point, respectively, when the self-employed are incorporated.

This study has broader implications as it connects two topics that have often been considered separately: wage inequality and business preferences regarding their legal status. It argues that levels of wage inequality are influenced by private businesses' decisions to organize as corporations (resulting in highly paid officers) or as partnerships (resulting in proprietors earning a mixed income that falls outside the scope of most studies). Consequently, significant shifts in these preferences should be taken into account, and scholars should strive to uncover the complexities imposed by these legal definitions. A promising avenue of research would involve gaining a better understanding of whether differences in top wage shares between countries are determined by the substantial variations in the relative prevalence of corporations. For instance, if corporations were much more important in the United States than in other advanced economies at the beginning of the 20th century –as illustrated by Hannah (2015) –, it would follow that the observed disparities in wage inequality levels could be attributed to this factor. At the same time, the legal form of business also carries profound social implications, potentially facilitating the mobility of certain groups while hindering others. These changes should be integrated into our understanding of racial and gender wage differences during the 20th century.

References

Amemiya, Takeshi. 1978. “The Estimation of a Simultaneous Equation Generalized Probit Model.” *Econometrica: Journal of the Econometric Society*: 1193–1205.

Atkinson, A. B., and A. J. Harrison. 1978. *Distribution of Personal Wealth in Britain*. Cambridge ; New York: Cambridge University Press.

Atkinson, A. B., and S. Voitchovsky. 2011. “The Distribution of Top Earnings in the UK since the Second World War.” *Economica* 78(311): 440–59.

Auten, Gerald, and David Splinter. 2023. “Income Inequality in the United States: Using Tax Data to Measure Long-Term Trends.” *Journal of Political Economy*. doi:10.1086/728741.

Autor, David, Claudia Goldin, and Lawrence F. Katz. 2020. “Extending the Race between Education and Technology.” In *AEA Papers and Proceedings*, American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203, 347–51.

Bailey, Martha J., and William J. Collins. 2006. “The Wage Gains of African-American Women in the 1940s.” *The Journal of Economic History* 66(3): 737–77. doi:10.1017/S0022050706000313.

Bernanke, Ben S. 1986. “Employment, Hours, and Earnings in the Depression: An Analysis of Eight Manufacturing Industries.” *The American Economic Review* 76(1): 82–109.

Blanchet, Thomas, Juliette Fournier, and Thomas Piketty. 2022. “Generalized Pareto Curves: Theory and Applications.” *Review of Income and Wealth* 68(1): 263–88. doi:10.1111/roiw.12510.

Bureau of Internal Revenue Statistical Section Income Tax. 1945a. *Preliminary Study of Individual Income Tax Returns for 1944, Forms W-2 and Forms 1040 with Adjusted Gross Income Under $25,000*. Bureau of Internal Revenue, Statistical Section, Income Tax Unit.

———. 1945b. *Supplement to Statistics of Income for 1939, Part 1: Compiled from Partnership Returns of Income*. Washington: U.S. Government Printing Office.

Collins, William J., and Gregory T. Niemesh. 2019. “Unions and the Great Compression of Wage Inequality in the US at Mid-Century: Evidence from Local Labour Markets.” *The Economic History Review* 72(2): 691–715. doi:10.1111/ehr.12744.

Department of the Treasury. 1940. “The Estimated Number and Compensation of Federal, State, and Local Officers and Employees, 1937-1938.” *Bulletin of the Treasury Department*: 1–3.

Fisher-Post, Matthew. 2020. “Examining the Great Leveling: New Evidence on Mid-Century American Inequality.” *WID.world Working Papers series* (01).

Frydman, Carola, and Raven Molloy. 2012. “Pay Cuts for the Boss: Executive Compensation in the 1940s.” *The Journal of Economic History* 72(01): 225–51. doi:10.1017/S002205071100249X.

Frydman, Carola, and Raven E. Saks. 2010. “Executive Compensation: A New View from a Long-Term Perspective, 1936-2005.” *Review of Financial Studies* 23(5): 2099–2138.

Geloso, Vincent, and Phillip Magness. 2020. “The Great Overestimation: Tax Data and Inequality Measurements in the United States, 1913–1943.” *Economic Inquiry* 58(2): 834–55. doi:10.1111/ecin.12865.

Geloso, Vincent, Phillip Magness, John Moore, and Philip Schlosser. 2022. “How Pronounced Is the U-Curve? Revisiting Income Inequality in the United States, 1917–60.” *The Economic Journal* 132(647): 2366–91. doi:10.1093/ej/ueac020.

Goldin, Claudia, and Lawrence F. Katz. 2010. *The Race between Education and Technology*. Cambridge, Mass.: The Belknap Press of Harvard University Press.

Goldin, Claudia, and Robert A. Margo. 1992. “The Great Compression: The Wage Structure in the United States at Mid-Century.” *The Quarterly Journal of Economics* 107(1): 1–34. doi:10.2307/2118322.

Gómez León, María, and Herman J. De Jong. 2019. “Inequality in Turbulent Times: Income Distribution in Germany and Britain, 1900–50.” *The Economic History Review* 72(3): 1073–98. doi:10.1111/ehr.12770.

Hanes, Christopher. 2000. “Nominal Wage Rigidity and Industry Characteristics in the Downturns of 1893, 1929, and 1981.” *American Economic Review* 90(5): 1432–46. doi:10.1257/aer.90.5.1432.

Hannah, Leslie. 2015. “A Global Corporate Census: Publicly Traded and Close Companies in 1910.” *The Economic History Review* 68(2): 548–73.

Hatton, Timothy J., and Mark Thomas. 2010. “Labour Markets in the Interwar Period and Economic Recovery in the UK and the USA.” *Oxford Review of Economic Policy* 26(3): 463–85. doi:10.1093/oxrep/grq023.

Heckman, James J. 1978. “Dummy Endogenous Variables in a Simultaneous Equation System.” *Econometrica* (46): 931–59.

Humphrey, Hubert H. 1994. “August 29, 1950: Revenue Act of 1950.” In *The Senate 1789-1989. Classical Speeches 1830-1993. Volume Three. Bicentennial Edition*, Washington D.C.: U.S. Government Printing Office, 625–50.

Jaworski, Taylor, and Gregory T. Niemesh. 2018. “Revisiting the Great Compression: Wage Inequality in the United States, 1940–1960.” *Historical Methods: A Journal of Quantitative and Interdisciplinary History* 51(1): 39–48. doi:10.1080/01615440.2017.1393360.

Krolage, Carla, Andreas Peichl, and Daniel Waldenström. 2022. “Long-Run Trends in Top Income Shares: The Role of Income and Population Growth.” *The Journal of Economic Inequality* 20(1): 97–118. doi:10.1007/s10888-021-09520-8.

Kuznets, Simon. 1953. *Shares of Upper Income Groups in Income and Savings*. New York: NBER.

———. 1954. *National Income and Its Composition, 1919-1938*. National Bureau of Economic Research.

Kwon, Spencer Yongwook, Yueran Ma, and Kaspar Zimmermann. 2023. “100 Years of Rising Corporate Concentration.” *University of Chicago, Becker Friedman Institute for Economics Working Paper* 2023(20). doi:10.2139/ssrn.3936799.

Liebenberg, Maurice, and Hyman Kaitz. 1951. “An Income Size Distribution from Income Tax and Survey Data, 1944.” In *Studies in Income and Wealth, Volume 13*, NBER, 378–462.

Lindert, Peter H, and Jeffrey G Williamson. 1983. “Reinterpreting Britain’s Social Tables, 1688–1913.” *Explorations in Economic History* 20(1): 94–109. doi:10.1016/0014-4983(83)90044-X.

Lindert, Peter H., and Jeffrey G. Williamson. 2016. *Unequal Gains: American Growth and Inequality since 1700*. Princeton, NJ: Princeton University Press.

Maddala, Gangadharrao S. 1983. *Limited-Dependent and Qualitative Variables in Econometrics*. Cambridge: Cambridge University Press.

Maloney, Thomas N. 1994. “Wage Compression and Wage Inequality Between Black and White Males in the United States, 1940–1960.” *The Journal of Economic History* 54(2): 358–81. doi:10.1017/S0022050700014522.

Mas, Alexandre. 2019. “Does Disclosure Affect CEO Pay Setting?: Evidence from the Passage of the 1934 Securities and Exchange Act.” *Princeton University, Department of Economics, Industrial Relations Section* (632).

Milanovic, Branko, Peter H. Lindert, and Jeffrey G. Williamson. 2011. “Pre-Industrial Inequality.” *The Economic Journal* 121(551): 255–72. doi:10.1111/j.1468-0297.2010.02403.x.

Modalsli, Jørgen. 2015. “Inequality in the Very Long Run: Inferring Inequality from Data on Social Groups.” *The Journal of Economic Inequality* 13(2): 225–47. doi:10.1007/s10888-014-9279-6.

Moriguchi, Chiaki, and Emmanuel Saez. 2008. “The Evolution of Income Concentration in Japan, 1886–2005: Evidence from Income Tax Statistics.” *Review of Economics and Statistics* 90(4): 713–34. doi:10.1162/rest.90.4.713.

Philippon, Thomas, and Ariell Reshef. 2012. “Wages and Human Capital in the U.S. Finance Industry: 1909–2006.” *The Quarterly Journal of Economics* 127(4): 1551–1609. doi:10.1093/qje/qjs030.

Piketty, Thomas. 2001. *Les hauts revenus en France au XXe siècle: inégalités et redistributions, 1901-1998*. Paris: B. Grasset.

———. 2014. *Capital in the Twenty-First Century*. Cambridge, Mass.: Belknap Press of Harvard University Press.

Piketty, Thomas, and Emmanuel Saez. 2003. "Income Inequality in the United States, 1913–1998." *The Quarterly Journal of Economics* 118(1): 1–41.

———. 2007. "Income and Wage Inequality in the United States, 1913-2002." In *Top Incomes over the Twentieth Century: A Contrast between Continental European and English-Speaking Countries*, eds. Anthony B. Atkinson and Thomas Piketty. Oxford; New York: Oxford University Press, 141–225.

Securities and Exchange Commission. 1935. *Official Summary of Holdings of Officers, Directors and Principal Stockholders*. Washington D.C.: Government Printing Office.

———. 1940. *Survey of American Listed Corporations*. New York: Securities and Exchange Commission.

Stigler, George J. 1956. *Employment in the Service Industries*. Princeton, NJ: Princeton University Press.

United States, Bureau of Internal Revenue. Various years. *Statistics of Income. Compiled from Corporation Income and Excess-Profits Tax Returns*. Washington D.C.: U.S. Government Printing Office.

———. Various years. *Statistics of Income. Compiled from Income Tax Returns and Including Statistics from Estate Tax Returns and Gift Tax Returns*. Washington D.C.: U.S. Government Printing Office.

United States. Office of Business Economics. 1953. *Income Distribution in the United States by Size, 1944-1950*. Washington D.C.: U.S. Government Printing Office.

U.S. Treasury Department, Division of Tax Research. 1940. *Statistics of Income Supplement Compiled from Federal Income Tax Returns for 1936: Individual Incomes*. Washington.

U.S. Treasury Department, Internal Revenue Service. 1957. *Statistics of Income: Partnership Returns 1953*. Washington D.C.: Government Printing Office.

US Treasury, United States Research and Statistics Division. 1938. *Statistics of Income Supplement, Compiled from Federal Income Tax Returns of Individuals for the Income Year 1934. Section 1: Number of Individual Income Tax Returns for 1934 Classified by Counties, and Cities of 25,000 and Over Population by Net Income Classes*. Washington: Government Printing Office.

Appendix: Top Earners and the Great Compression:

New Estimates based on Tax Records

Wages in personal income tax returns 52
Wage-earners in tax returns 54
The distribution of wages in tax returns 57
Accruing wages to couples filling joint returns. 62
Adding the tax-exempted population: Local government employees 71
A comparison with Piketty and Saez's series 72
Corporate officers' compensation 74
Corporate tax statistics and SEC data 74
Average pay: per company and per individual 77
The distribution of officers' compensation 81
Calculating the net change in the number of corporations 82
Proprietors' income: tax and census data 84
Tax records on sole-proprietorships and partnerships 84
The 1950 census records 85
Labor and capital shares in proprietors' incomes 85
Sensitivity analysis 87
References 92

This paper provides estimates on wage inequality in the United States for the 1918-49 period, primarily based on tax records. The appendix, derived from a technical working paper by Artola and Gómez-Blanco, explains step-by-step the methodology used to perform the computations. We first address the key data inputs and adjustments to the wages reported in income tax tabulations to obtain the definitive distribution of wages. Second, we present our estimates on corporate officers' compensation during the period based on two additional sources (i.e., corporate tax statistics and a special survey conducted by the SEC). Third, we deal with the sources used to estimate the number and incomes of high-earning proprietors. Finally, we present the fixed group shares and provide several robustness checks and additional insights of the regressions included in the main text.

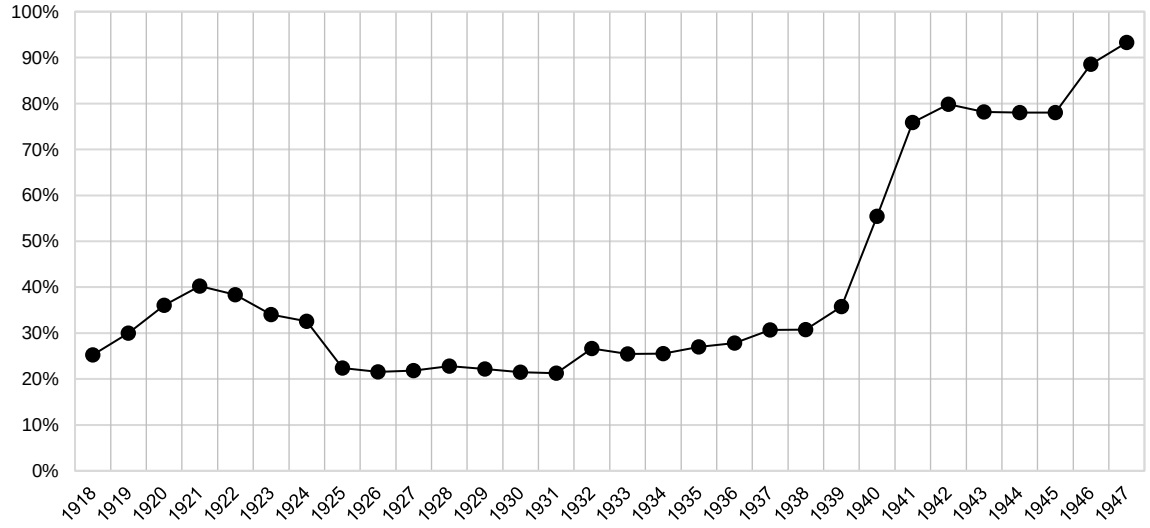
Wages in personal income tax returns

The federal income tax, enacted in 1913, is levied on the aggregate income of individuals or married couples. The published tabulations of the Bureau of Internal Revenue typically provide information on the number of returns classified either in terms of adjusted gross income or net income.[6] However, we focus solely on one component, namely the wages declared in tax returns. This concept is defined in sufficiently broad terms (i.e., wages, salaries, tips, commissions, bonuses, etc.) to align with the definition used in the US national income and product accounts (NIPA). A useful insight is to compare the wages reported in the official *Statistics of Income* (SOI) with the national accounts aggregate

[6] Three related income concepts are used in this appendix. Total income refers to all positive sources of income such as wages, interests, dividends, unincorporated business profits, capital gains, etc. Adjusted gross income (AGI) deducts from total income the deficit reported by unincorporated businesses and capital losses. Net income starts from AGI and additionally deducts donations to charities, interest on mortgages and personal loans, local and state taxes, plus other items authorized by the law.

(Figure 1).[7] The picture that emerges from this exercise is that, prior to the huge expansion of the income tax during the Second World War, there was already a significant share of wages reported in tax returns, especially when the filling threshold was relatively low (i.e., at $2,500 for married couples in the 1918-24 and 1932-39 years).

Figure 1. Wages reported in tax returns relative to the national aggregate (NIPA), 1918-1947



Sources: SOI and NIPA.

From this raw data, we carry three additional adjustments: First, we estimate the number of wage-earners among the tax-filling population. With this information in hand, one can accurately compute an average wage among earners. In a second step, we simulate a distribution of wages at the smallest scale possible, that is, among the earners that fit into a single income-bracket. Note that for this process, we use the information directly provided by SOI for the returns with net income over $5,000 from 1927, thereby ensuring that the wages of the top 1% are perfectly captured in tax records. However, for returns

[7] Wages from national accounts are based from 1929 onward on the official figures provided by the Bureau of Economic Analysis (BEA), and before this date on very similar estimates (Kuznets 1954).

below this amount, we need to impute some information. As a third step, we carry some additional amendments. The most critical is the splitting up of wages reported in joint returns when both spouses were actively employed. This adjustment is important if one aims to fully reconcile the tax information with other sources, for example, census records. Another relevant adjustment is to add the wages (and earners) of the tax-exempted population, most notably of local and state government employees before 1938.

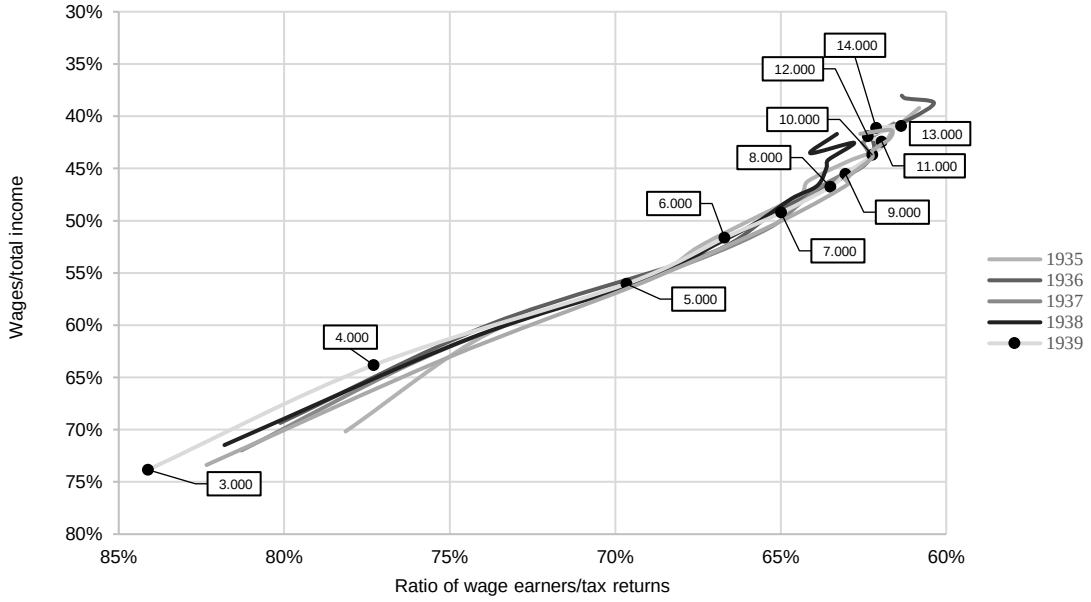
Wage-earners in tax returns

Tax statistics typically tabulate information of tax returns in terms of aggregate income. From this information, we aim to build a distribution of tax returns in terms of wages. To achieve this, our initial step involves determining the count of wage recipients among the tax-filing population. In the case of the United States, the *Statistics of Income* (United States, Bureau of Internal Revenue Various years) include this information on a regular basis from 1939 onward. Prior to that date, there are some additional reports that provides this information in 1934, 1936 and 1937 (US Treasury 1938; U.S. Treasury Department, Division of Tax Research 1940). Furthermore, SOI reports provide the number of returns with wage income for taxpayers with net income over $5,000, either on a per bracket basis (i.e., for returns with net income between $5,000-$6,000, $6,000-$7,000, etc.) in the period 1934-38, or in aggregate terms, as in the years from 1927 to 1933. This additional information for the $5,000 group is provided because these returns were normally audited in the central bureau of the revenue service in Washington, and therefore they were examined at great length.

For all returns in years 1918-26, as well as for returns with net income under $5,000 in the years with missing information (1927-33, 1935, and 1938), we need to estimate the number of wage recipients. The easiest and most consistent way to address this challenge is to focus on the information already available in SOI. From the sources, it is easy to determine that, for each bracket, the share of wages over total income is related to the number of wage recipients out of

all taxpayers. This relationship is also stable over time (see Figure 2). We, therefore, use the wages reported on a bracket basis to estimate the number of recipients.

Figure 2. Relation between wage recipients (as a share of returns) and wages (as a share of income)
Information from tax returns by net income brackets ($3,000-$14,000) and years (1935-1940)



Sources: SOI.
Notes: Axis have been sorted in decreasing order to match them with an increasing trend in income level (recall that the relationship between income and the two ratios is negative).

SOI reports points that, in this period, the percentage of wage earners per tax return decreases as wages in net income shrinks, until wage earners represents on average between 60% and 65% of tax returns. At that point, which corresponds to $8,000 of net income, the share of wage earners remains stable, even though wages over net income still fall further. Using the yearly information on wages and total income, we estimate the number of wage earners filling a tax return prior to 1927 (or 1934 for tax returns below $5,000) by using the relation mentioned above. We run a pooled OLS for the Equation (1) in the years with comprehensive information (1934-40).

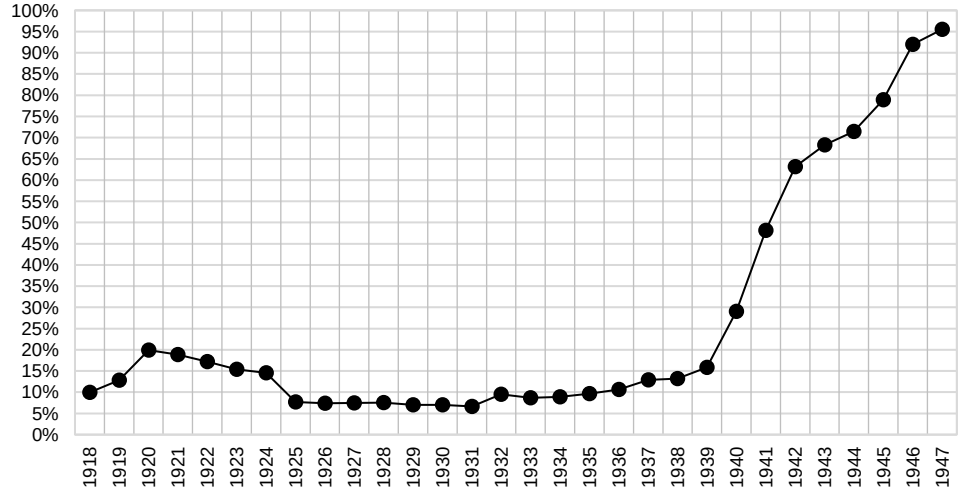
$$\frac{Wage\ earners_t^i}{Tax\ returns_t^i} = \alpha^i + \beta^i \frac{Wages_t^i}{Total\ income\ _t^i} + \varepsilon_t^i \qquad (1)$$

where *i* stands for the income interval (bracket) and *t* is the year.

We thus compute the shares of wage earners over total tax returns per net-income bracket using the coefficients obtained from Equation 1. With this information, calculating the number of wage earners becomes straightforward -we multiply the corresponding shares by the number of returns provided by SOI each year.

The results are displayed in Figure 3, where the number of tax returns reporting wages is expressed as a percentage of the national wage-earning population provided by NIPA. The general picture is similar to the one presented in Figure 1. Although it was only after the Second World War that the income tax started to cover the bulk of the wage-earning population, prior to this date, there was still a significant share (around 10-15%) recorded, and this never fell below 5%. This data constraint explains why our estimates are never projected below the 95th percentile. Finally, we should caution the reader that this series requires an additional adjustment (i.e., dividing wages reported in joint returns by couples with two earners), which pushes upward the percentage of the population covered in tax statistics.

Figure 3. Tax returns reporting wages relative to the national wage earning population (NIPA), 1918-1947



Sources: SOI and NIPA.

In addition to this basic check, some external controls demonstrate that the errors generated by this method are very small. As explained, SOI provides the number of wage earners with net income over $5,000 per bracket for the period 1927-34. We can estimate these figures using the previous method and compare them with the official statistics. It is important to emphasize that this group (i.e., top-paid taxpayers) is of a smaller size, and its share of wages on income is lower, which increases variability and potential errors. Yet, even under these conditions, the gap between imputed and actual values is smaller than 1% when pooling all brackets, as shown in Table 1. Consequently, this method proves highly robust.

Table 1 Number of tax returns with net income over 5,000$ and reporting wages

Year	(A) Observed	(B) Estimated	Ratio (B/A)
1927	540.893	538.421	99,54%
1928	612.988	601.823	98,18%
1929	623.793	621.443	99,62%
1930	515.718	513.153	99,50%
1931	386.979	383.387	99,07%
1932	235.828	234.037	99,24%
1933	214.021	213.136	99,59%

Sources: SOI.

The distribution of wages in tax returns

The second step requires a more in-depth analysis over the distribution of earnings. Ideally, we aim to have information as detailed as possible on the distribution of tax returns both in terms of income and wages. To illustrate this point, Table 2 provides the information for 1946. From these records, it is relatively easy to estimate the wage distribution by using the generalized Pareto curve interpolation method (Blanchet et al.,

2022). This procedure can be used for the aggregate tax-filling population, or for one specific group (e.g., for returns with income between $3,000 and $4,000).

Table 2 Cross-tabulation of income tax returns reporting wages, USA 1946

Income brackets (Adjusted Gross income)	Number of returns (000s)	Wage brackets										
		1	500	1.000	1.500	2.000	2.500	3.000	4.000	5.000	10.000	25.000
1	4.280	99%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
500	5.143	4%	95%	0%	0%	0%	0%	0%	0%	0%	0%	0%
1.000	6.248	2%	4%	93%	1%	0%	0%	0%	0%	0%	0%	0%
1.500	7.041	1%	2%	3%	93%	1%	0%	0%	0%	0%	0%	0%
2.000	6.743	1%	1%	2%	4%	92%	1%	0%	0%	0%	0%	0%
2.500	5.463	1%	1%	1%	2%	5%	89%	1%	0%	0%	0%	0%
3.000	6.305	1%	1%	1%	1%	2%	5%	89%	1%	0%	0%	0%
4.000	2.366	1%	1%	1%	1%	1%	2%	9%	83%	1%	0%	0%
5.000	1.584	3%	2%	1%	1%	2%	2%	5%	8%	76%	0%	0%
10.000	386	5%	3%	2%	2%	2%	2%	4%	3%	18%	59%	0%
25.000	105	5%	3%	2%	2%	2%	2%	3%	2%	11%	26%	42%

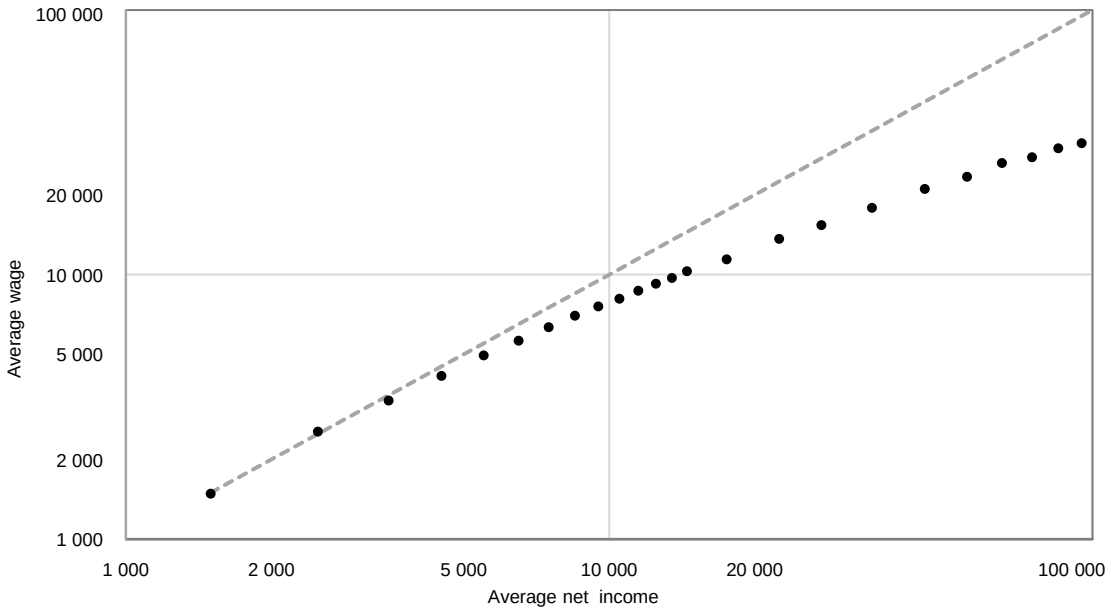
Sources: SOI.

Unfortunately, such comprehensive information is only available since the end of the Second World War. Prior to this date, similar statistics exist only for tax returns with net income over $5,000 from 1927 onward. In contrast, for tax returns under this amount, we need to estimate the wage distribution based on aggregate income tabulations. To do so, we employ a methodology called *constant Lorenz curve transformation*, that differs from other approaches such as the one applied by Piketty and Saez. In fact, our method is more akin to the techniques carried by official authorities after the Second World War (Liebenberg and Kaitz 1951).

This methodology relies primarily on two basic assumptions. First, the distribution of wages for a closely defined population group (i.e., returns that fit within a single net income bracket) remains constant across time. The underlying reasoning is that wages constituted most of taxpayers' income with less than $5,000 of net income, and these wages should fit within narrow limits as net income brackets are closely defined (i.e., between narrowly lower and upper thresholds). Changes throughout time are then most likely due to fluctuations between groups or a shifting combination among wages and

other sources of income. In practice, this factor evolves very slowly through time, and there is no discernible trend in the period covered in this paper. Figure 4 reflects this premise by relating the observations in terms of wages and net income in 1936. Despite there is an increasing divergence relative to the aggregate income, low-income observations, which are the focus of this estimation, fit very near the 45-degree line.

Figure 4. Tax returns with wages, by net income brackets, 1936.



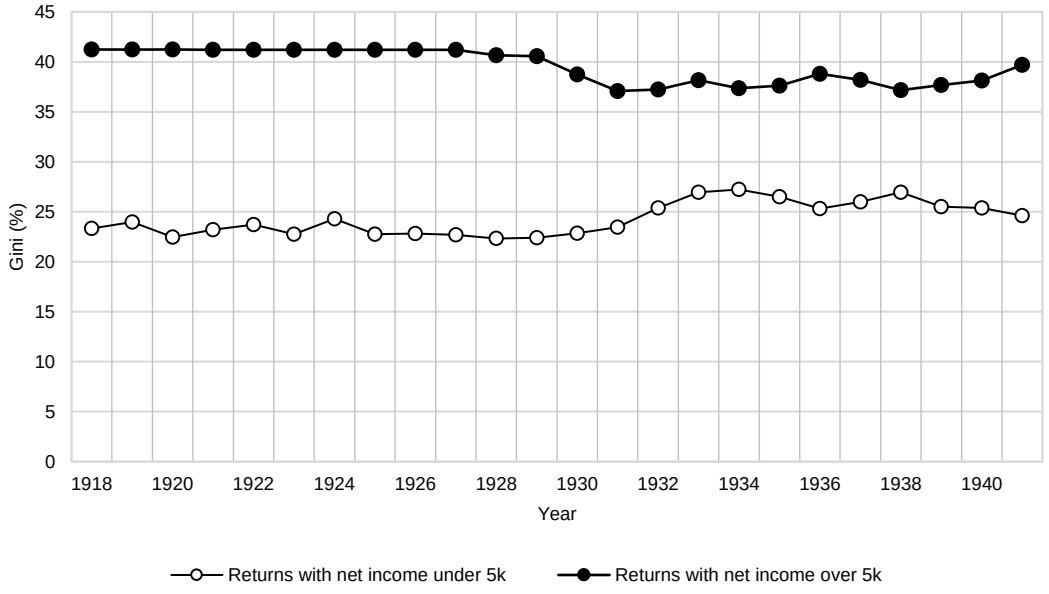
Sources: SOI.
Note: Observations grouped for the available brackets in terms of net income.

The second assumption is that, for each of these closely defined groups, the wage distribution can be approximated by taking the variation in terms of total income. This is largely validated by available records. For example, in the matrix shown in Table 2, the diagonal –pointing to a very close match of adjusted gross income and wages– has values near 100% among the lowest-income brackets. To illustrate this point, the Gini index observed from the wage distribution for the population with net income over $5,000 hovers between 37% and 40% in the period of 1927-41 (Figure 5). This is a relatively small gap if one factors in the relatively heterogeneity of the group, which has no upper bound. The inclusion of top earners moreover increases the variability of

wages. For groups placed relatively lower in the wage distribution, changes are most likely smaller throughout time.

An accurate distribution in terms of total income is especially useful. For this purpose, we benefit from a special report for the year 1936 that included a matrix relating returns in terms of total and net income and calculate a Lorenz curve of total income for each net income bracket ($1,000-$2,000, $2,000-$3,000, and so on). Note that the resulting distribution shows greater variation due to the heterogeneous use of deductions among taxpayers. Finally, the resulting Lorenz curves for each net income bracket are extrapolated to other years. It is worth noting that in this exercise, we use the original annual information reported in SOI on net income, deductions, wages, and the number of earners. The only component that is extrapolated is the distribution of wage earners in terms of net income by bracket.

Figure 5. Gini index for wage income: returns with net income under and over $5,000



Sources: SOI.
Notes: The Gini index for returns with net income over $5,000 has been kept constant from 1918 to 1926 by taking the value observed in 1927.

The results from these estimates are summarized in Figure 5. At first glance, the wage distribution for the two main subgroups of earners (i.e., under and over $5,000 in net income terms) moves within some close bounds. This is mostly because the average nominal wage for both groups

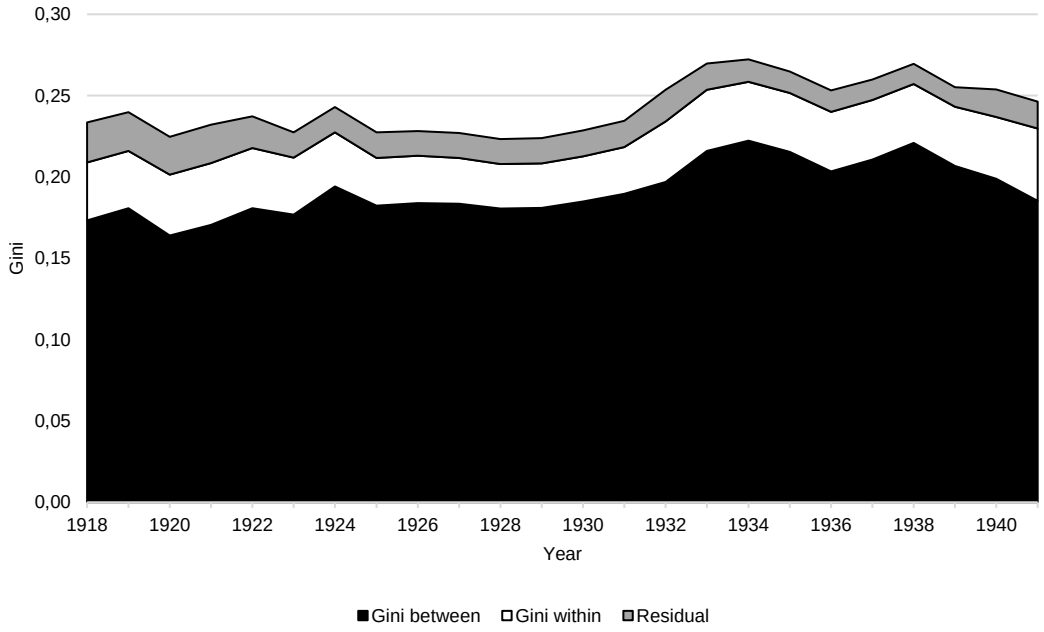
barely budges during this period. This fact is explained because the compositional effect caused by the entry/exit of tax-filers is larger than the underlying changes in long-standing earners due to macroeconomic conditions. Still, one should not dismiss some significant variations, most importantly among the subgroup that is imputed (i.e., returns with less than $5,000 in net income).

This method presents significant advantages over other ways of estimating the wage distribution from tax records at a time with no available microdata. For example, the adjustment carried out by Piketty and Saez (2003) takes a similar starting point (i.e., income in lower brackets depends almost entirely on wages), but they impute the wage distribution of each subgroup based on the distribution in terms of net income (as opposed to total income).[8] The difference between this method and our proposal is illustrated in Figure 6. Both methods would correctly estimate the Gini component between groups (i.e., the difference in the average wage of earners in one bracket versus other brackets). However, their method would impute a smaller within component (as net income shows a smaller variation than total income).

Furthermore, by definition, their method does not allow for a residual (or overlap) component (a person could have a higher wage than someone in the next income bracket). Therefore, the major advantage of our method is that it provides higher, more reliable Gini coefficients owing to Lorenz curve convexity and re-ranking of tax returns.

[8] This assumption is better understood with an example. If one focuses in the net income bracket of $2,000-$3,000, the average net income is $2,500, and the average wage among its recipients is $2,400. Consequently, the wage bracket is defined by the ratio between the two ($\frac{2400}{2500} = 0.96$), to make a new lowerbound ($1,920) and upperbound ($2,880)

Figure 6. Gini coefficient and its components in returns with net income of less than 5,000$



Source: Author's own elaboration based on data of SOI.

Despite these advantages, our technique also has some drawbacks. First and foremost, it assumes that the relationship between net and total income by interval is the same as in 1936. This may overstate (or understate) inequality if total income becomes less (or more) concentrated with respect to the wage distribution. However, the implications of this assumption are quite limited because the correlation between wage and total income is very high, and imputations are based on narrow income brackets. Another potential limitation is that, for returns with net income over $5,000, we keep the distribution constant from 1918 to 1926. Again, the potential limitations of this procedure are small, as the observed variation of the Gini is indeed very low for this group

Accruing wages to couples filing joint returns.

During the analyzed period, married couples could choose to file a tax return either jointly or separately, and this choice was solely based on their personal decision. In separate returns, each spouse would report the income they had earned, individualizing wages

accordingly. In joint returns, the reported compensation could be that of one spouse (if only one person was employed) or both spouses. In this section, we first delve into the available information aimed at addressing the magnitude of the issue at hand -specifically, the share of married women in employment. Second, we examine the structure of the US tax system to gain a better understanding of the incentives for married couples to opt for either filing system. Finally, we propose an adjustment process to individualize earnings for the treatment group.

In the interwar period, income tax returns rarely specified whether joint returns with wages had one or two earners. However, given the low labor participation rate of married women during that time, it is reasonable to assume that, in most cases, there was only one earner. The microdata of the 1940 census (Ruggles et al. 2021) allows for a more detailed analysis of employed married women. According to this source, 27.8% of adult women earned wages in the United States, but the ratio decreases to 17.8% when considering only married women.

Presumably, it is reasonable to believe that more affluent households had even lower female participation rates, but one must be careful when testing this hypothesis. If we consider aggregate wage income reported at the household level, the participation rates for married women decrease to 16% for households among the top 5% of the wage distribution. However, this ratio surprisingly increases to 20.2% in households among the top 1% of the distribution. At this point, it is essential to note that households in which both spouses were employed are more likely to fall into the highest percentiles simply because of the combination of two sources of income. Therefore, it is more useful to study the labor participation decision of married women as a function of the husband's wage income. To explore these patterns, we employ a logit model with a dummy dependent

variable equal to one if the wife works and zero otherwise.[9] The main independent variable is the husband's wage income, although the wife's educational level is also included as a control. Given that business and capital incomes were not reported in the 1940 census, only households in which husbands had positive wage income were included.[10] Non-married women were also excluded for obvious reasons. The output of the regression is displayed in Table 3.[11]

Table 3. Logit model on married women being on wage employment,

	(1)
VARIABLES	Work wife
Husband income	-0.418***
	(0.00823)
Wife educ	0.317***
	(0.00817)
Constant	-1.813***
	(0.00698)
Observations	181,245

Robust standard errors in parentheses

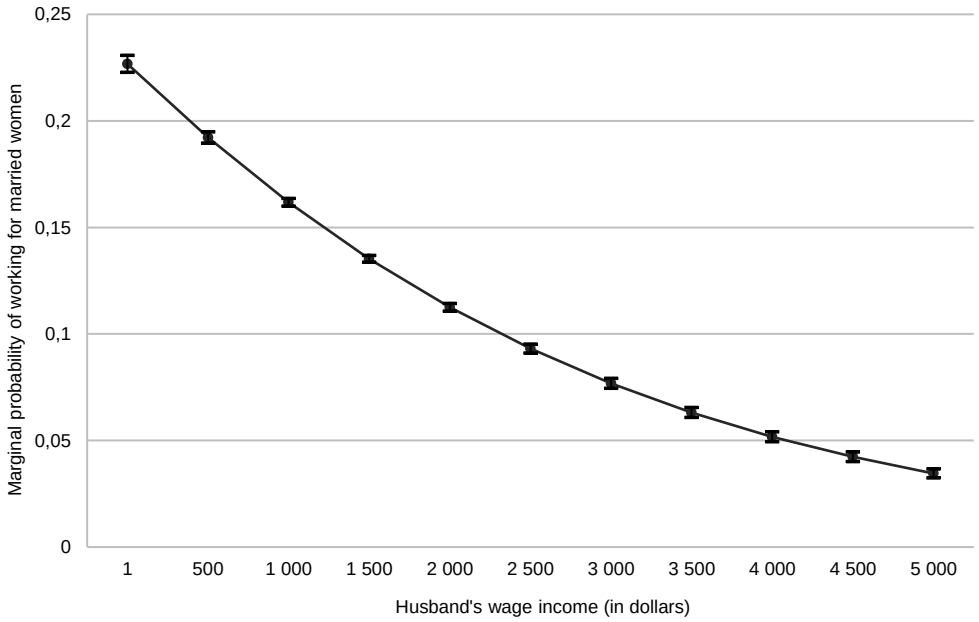
*** p<0.01, ** p<0.05, * p<0.1

The model shows that women were less prone to be employed as husband's earnings increased given any educational level. In Figure 7, we estimate the predicted probability based on the husband's wage at different values. This probability starts approximately at 23% for women with the lowest husbands' wages and decreases up to less than 4% when

[9] A probit specification was also applied. The results were barely the same, but the logit model showed slightly higher (in absolute terms) AIC and BIC values.
[10] This way we avoid potential biases that wealthy businessmen with zero wages could produce.
[11] Variables has been standardized to ease the comparison between coefficients.

husbands' earnings reach $5,000. Note that, in this paper, the focus is on returns reporting at least $2,500, thus the potential number of double-earnings returns should be between 3.4% and 9.3%.

Figure 7. Probability of being employed of married women dependent upon their husband's wage income. United States, 1939



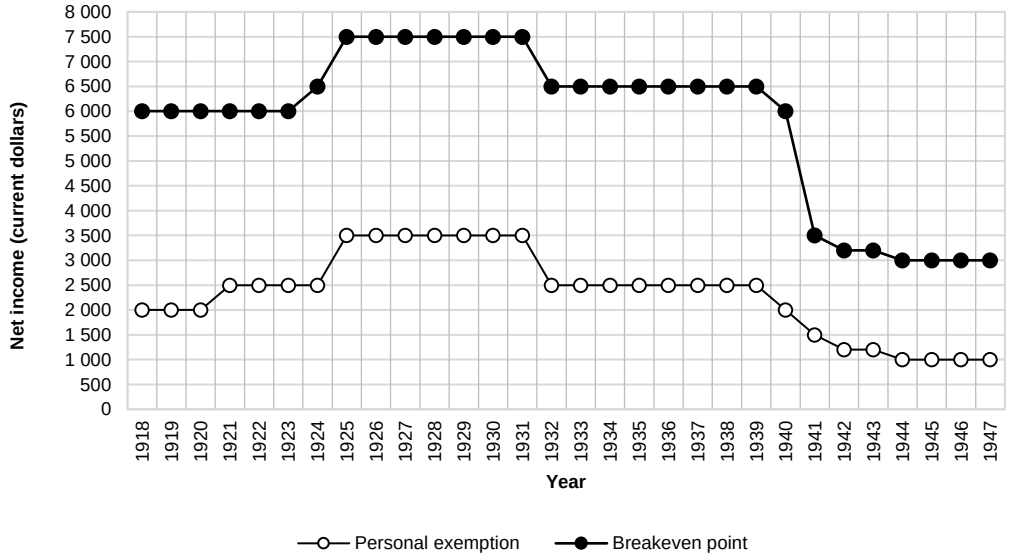
Sources: Author's own elaboration based on the 1940 census (IPUMS USA).

With this information in mind, it is critical to correctly apportion wages in joint statements. Piketty and Saez decided to leave the information provided in joint tax returns unchanged, and instead deducted from the denominator the number of married women in employment. This procedure, however, implicitly assumes that wages are ascribed to one spouse when they should be sometimes divided between both of them. In addition, this makes it more difficult to establish a comparison with other sources, such as census data, since it defines tax units (i.e., single individuals plus married couples) as the unit of analysis, in opposition to persons.

In this paper, we opt for an alternative method that can be summarized in the following manner. We start by analysing the potential tax benefits that could be derived from filing a joint tax return to comprehend the preference of couples to opt for this system or for separate returns. We then estimate how wages earned in married couples would split between men and women in the 1940 census. Finally, we split joint returns into individual salaries by using the previous information and the SOI tabulations of joint tax returns.

The first step is purely based on the analysis of the advantages stemming from the income tax legislation. From 1913 to 1947, in couples in which both spouses were wage-earners, the decision to opt for joint or separate returns was dictated by two variables: the amount deducted in terms of the personal exemption, and the cut-off point (or breakeven point) at which progressive tax rates started to work (i.e., the amount affected by the second tax bracket). Both are presented in Figure 8. Couples whose joint earnings fell between these two variables had an incentive to file jointly, as the tax guides of that time explained in detailed (Household Finance Corporation 1944, 6). By contrast, those whose joint earnings exceeded this threshold would file separately. Lastly, those who earned below the personal exemption would most likely not file a tax return, and consequently these observations are disregarded as being non-representative.

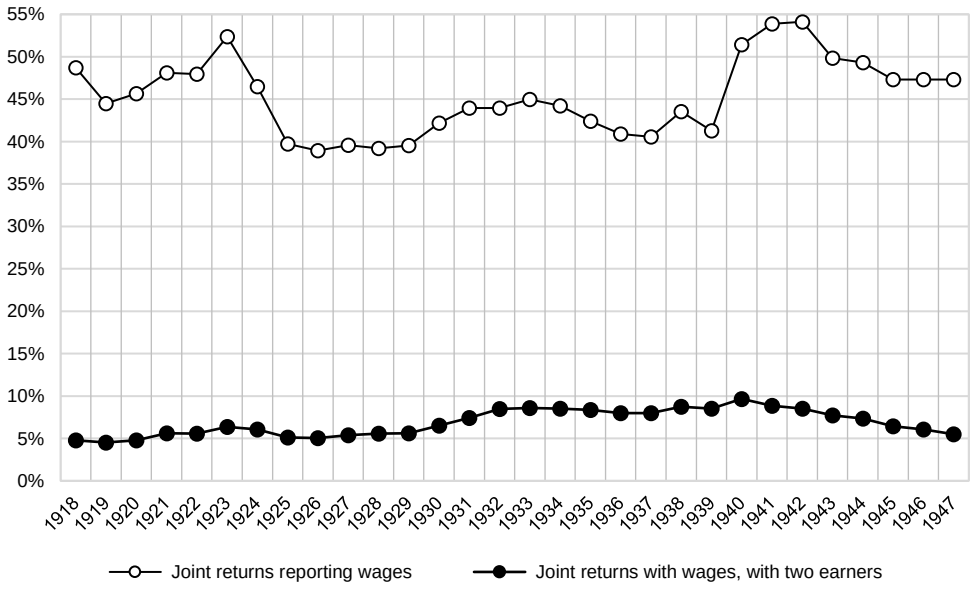
Figure 8. Personal exemption and breakeven point for married couples filing a tax return



To better understand this reasoning, it is worth recapping how both variables worked. The first one acted as the personal allowance granted to married couples. Since this amount was higher than for other individuals (single or widowed), this made that couples had an incentive to file jointly. This minimum threshold also works as the lower bound to use tax returns, as couples earning below this amount were most likely not filing a return. The second variable relates to the progressivity of the tax schedule. Married couples whose joint income was higher than this level had an incentive to file separately since otherwise they would be taxed at higher rates. Note that, in this situation, couples could split their personal exemption to the best of their advantage, to the point that one spouse could use the full amount and the other had no exemption at all. Furthermore, the tax provisions only granted the right to split income by accruing to each spouse the money that he (and she) earned. Capital income could be easily divided, but wages were apportioned to the person who had earned them. This provision ensures that wages reported in joint returns with incomes above the breakeven point are to be considered as only earned by one spouse.

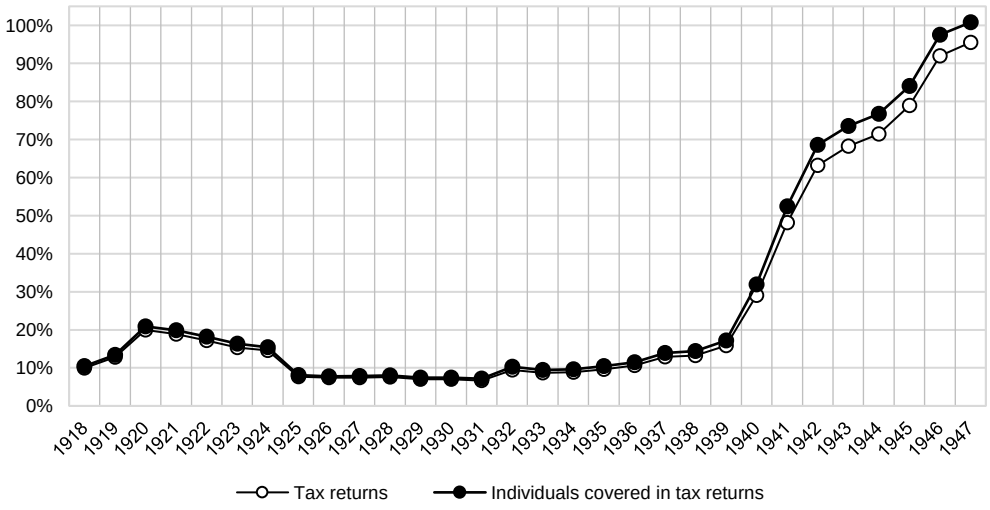
The second step consists in identifying the potential number that needs to be adjusted. To do so, we first quantify the number of joint returns that report wages and those with two earners. In this process we have benefited enormously from the information provided in the first government estimates (United States. Office of Business Economics 1953, 32–37) that included these data for the years 1943 to 1947. A very similar procedure is followed in 1936-9 by using a special set of tabulations of the 1936 income tax returns and the 1940 census microdata. Specifically, we use the IPUMS dataset (Ruggles et al. 2021) to compute couples' earnings when both spouses were wage-earners, and match them with their respective net income bracket. Finally, for other years, the trends are extrapolated by using the data available on the number of married women that were employed in the US economy. The results from these estimates are shown in Figure 9. One might note that the share of joint returns varies within a close bound (40-55%), and that changes are mainly due to variations in the personal exemption (as in 1924 or 1940). The share of joint returns with two earners is relatively low (between 5% and 10%), but its impact is far from negligible given that these returns usually have above-average wages (by definition, they combine two salaries). In fact, when one factors this, as Figure 10 shows, the overall coverage of the wage-earning population is more comprehensive than if one only takes the number of tax returns.

Figure 9. Joint returns with wages.
As a percentage of all returns reporting wages



Sources: SOI

Figure 10. Wage earners in tax returns relative to the national wage earning population (NIPA), 1918-1947



Sources: SOI and NIPA

The third step involves splitting wages in joint tax returns with two earners. Again, we make an extensive use of the information provided in the mid-1940s government reports (United States. Office of Business Economics 1953) and from the 1940 census. In this last case, we focus on couples whose joint wage fitted in the same income brackets

provided by the SOI tabulations and estimate the share of each partner. Table 4 shows average shares by spouse and bracket.

Table 4. Wife's and husband's share in total wage of the couple, 1939.

Couples' earnings (dollars)	Wife's share	Husbands' share
1-499	34.1%	65.9%
500-999	35.3%	64.7%
1000-1499	34.7%	65.3%
1500-1999	34.5%	65.5%
2000-2499	34.9%	65.1%
2500-2999	36.4%	63.6%
3000-3499	37.2%	62.8%
3500-3999	36.8%	63.2%
4000-4499	37.6%	62.4%
4500-4999	37.3%	62.7%
>5000	34.6%	65.4%
Mean	35.1%	64.9%

Sources: IPUMS USA (1939 census).

This table illustrates that, on average, 65% of couples' wages were earned by husbands, and 35% by wives. This estimate is consistent with the gender pay gap computed for the same period by other scholars (Goldin, 1990). Finally, we apply these ratios to the estimated number of joint returns with two earners in each bracket. In practical terms, this means deducting the original observations and adding a double set of new ones that, by definition, have lower wages. Overall, although this adjustment is a bit of an oversimplification (because it establishes a fixed ratio to split earnings in couples), it still seems to provide better estimates than the original set. Most importantly, after this amendment all observations correspond to individuals.

Adding the tax-exempted population: Local government employees

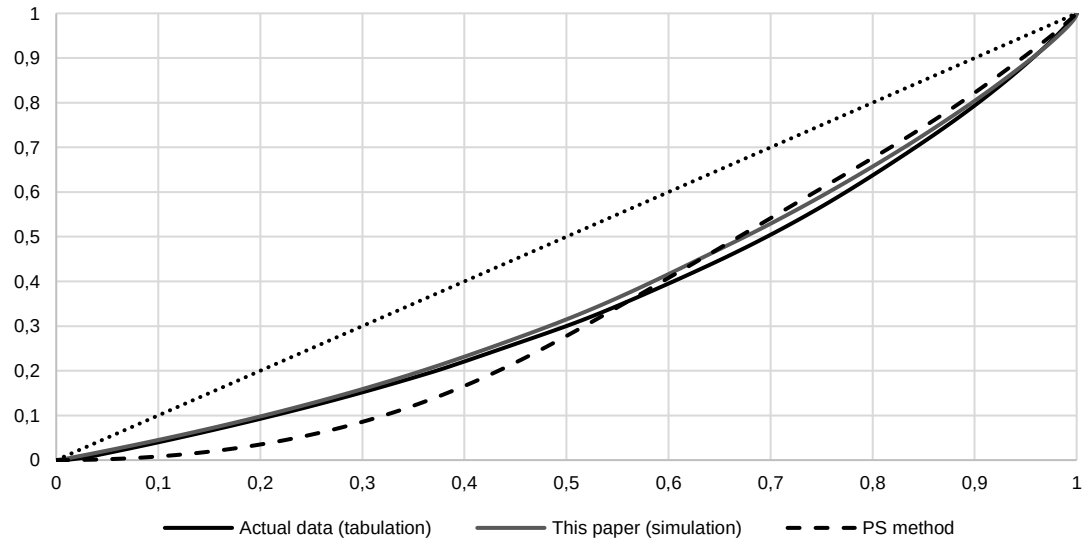
The US income tax had a relatively good legal design that covered most forms of labor earnings. In this period, there were only two major sources legally exempted. First, military compensation for active servicemen earning up to $1,500 was always untaxed (and unreported). In principle, this omission could only affect estimates from 1942 onward, when the filing threshold fell below this income level but not before. To avoid such a bias, $1,500 was kept as the minimum threshold of wage income to be used from tax records for the years 1942-7.

The second major concern relates to the exemption granted to the wages of local and state employees until 1938. In this case, the issue is more complex than with military pay as there is no threshold of reference. As noted by other scholars (Kuznets 1953; Geloso et al. 2022), the best solution is to estimate the incomes of this population. However, this is especially daunting as it involves computing not only wages and salaries, but also other secondary sources (interest, rents, etc.). Given that this paper is only interested in wages, a simpler assumption was made by first taking the number of local employees and their average compensation from national accounts. Second, we simulated a distribution each year by keeping constant the Lorenz curve observed in the year 1937 (Department of the Treasury 1940). Finally, those employees who had a salary above the minimum threshold ($2,500 for married couples during most of the period) were included. Overall, although these assumptions may seem strong, in practical terms, it amounts to adding approximately 250,000-300,000 new observations (i.e., equivalent to the top 10% of local and state government employees).

A comparison with Piketty and Saez's series

As explained in the previous sections, this paper introduces numerous amendments to Piketty and Saez's top wage estimates. The most critical revision relates to the way in which wages are distributed among the tax-filing population with net income under 5,000 dollars, as this is a relatively large group that compromises those earners typically fitting between percentiles 95 to 99 of the wage distribution. To check the reliability of the two procedures, in this section we present an additional point of comparison for year 1936. In that year, the already mentioned special SOI report separately included the distribution of wages of all returns with a net income of less than $5,000 (U.S. Treasury Department, Division of Tax Research 1940). We can hence compare the real distribution with the two alternatives: the procedure just explained and that resulting from the methodology followed by Piketty and Saez. As Figure 11 shows, our procedure provides a very accurate result relative to the actual Lorenz curve. In contrast, the result based on the Piketty and Saez procedure imputes too many observations below the median and too few above.

Figure 11 Lorenz curve for the distribution of wage income in returns with less than $5,000 net income

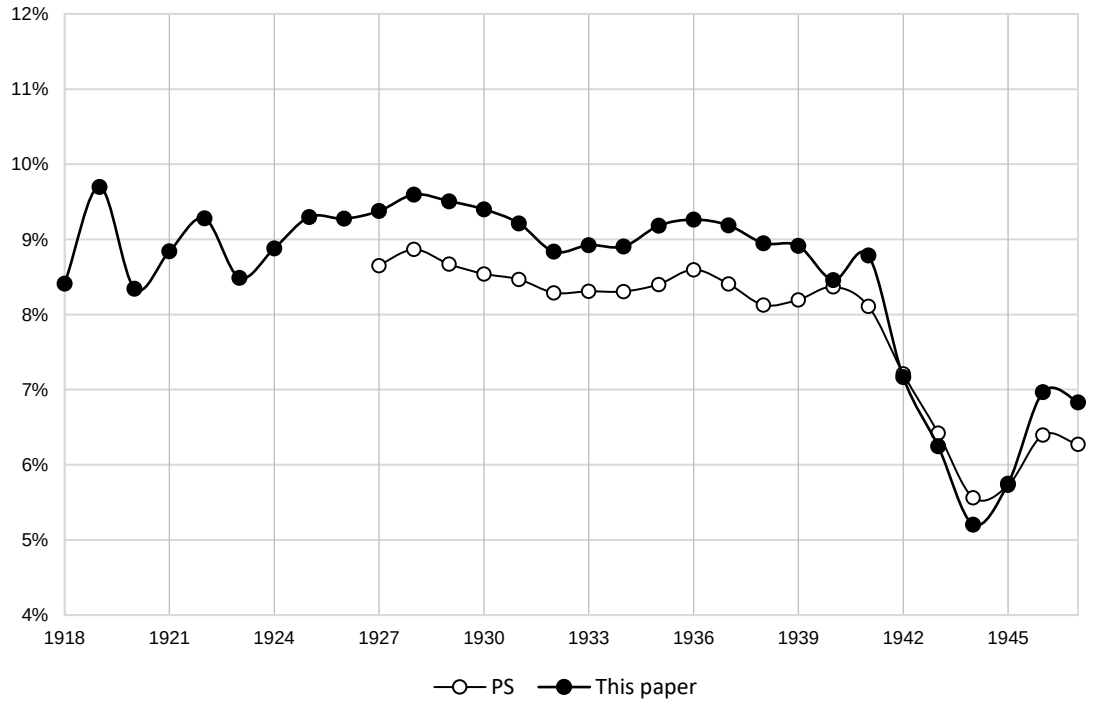


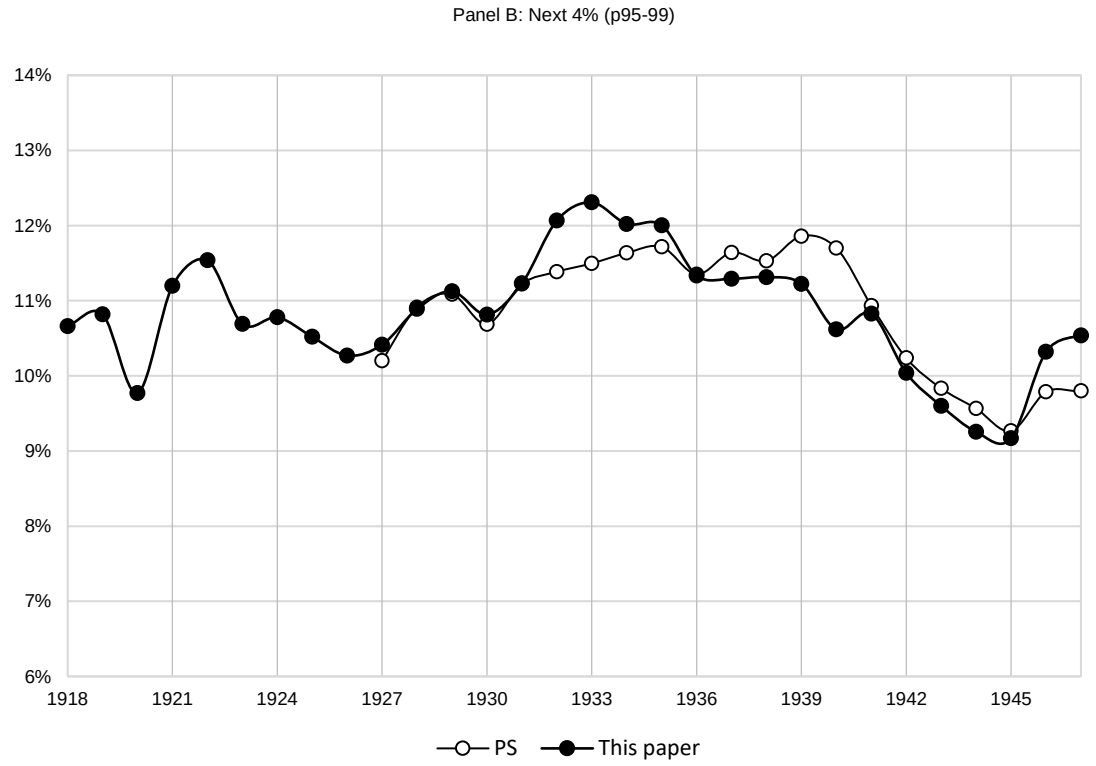
Sources: Own estimates based on the data provided in Piketty and Saez, (2007); U.S. Treasury Department, Division of Tax Research, (1940).
Note: The Lorenz curve is computed among those reporting wages and net income under 5,000$, not all returns.

One final point of this comparison is presented in Figures 12A and 12B, displaying the top 1% and next 4% income shares of wage earners estimated by both Piketty and Saez (2003) and by us. As observed, our estimates for the 99th percentile are generally one percentage point higher until the 1940s, when both coefficients converge. On the contrary, the next 4% is not systematically above or below but is clearly more volatile than Piketty and Saez's estimation. These differences are explained due to the different imputation of wages for returns with net income under 5,000$, the inclusion of local employees pay, and to the correct accruing of wages reported in joint returns. The first two factors typically increase our top wage shares, while the second factor reduces our metrics versus their estimates.

Figure 12. A comparison of top wage shares, Piketty and Saez (2003) and this paper

Panel A: Top 1% (p99-100)





Corporate officers' compensation

In a specific section of the paper, we quantify the contribution of corporate officers' compensation to the evolution of the top 1% wage share. To conduct this analysis, we utilize two additional sources: corporate tax statistics and a specialized survey from the Securities and Exchange Commission. From these sources, we introduce additional variables for analysis, including the number of active corporations, the count of officers, and their total compensation. Additionally, we estimate the distribution of corporate compensation by holding constant the fundamental parameters that influenced pay variation within each company.

Corporate tax statistics and SEC data

Our starting point is the corporate tax statistics published by the Bureau of Internal Revenue (United States, Bureau of Internal Revenue, various years). Within these records, we specifically focus on a particular category of the profit and loss account,

namely, “the compensation of officers”. Although the statistics do not provide an exact definition of this item, the corporate tax returns clearly state that it should include all forms of compensation (i.e., “whatever form paid”). Furthermore, the requirement for officers to be fully identified and the disclosure of their personal shareholdings clarify that only the very top managers (i.e., officers) are included. The advantage of this accounting item lies in its long-time span, as it was reported since the inception of the corporate tax statistical series at the end of the First World War. Moreover, starting from 1931, the corporate tax records report the number of companies (and corporate pay) by the asset size of their balance sheet. This information is especially useful to differentiate between the pay of small, medium, and large companies.

The second source used in this paper is the Survey of American Listed Corporations (SALC) elaborated by the Securities Exchange Commission for the years 1934-40 (Securities and Exchange Commission 1940). The survey, as detailed by Mas (2019), covered approximately 1,500 corporations, mostly from manufacturing and some service sector firms (department stores, grocery stores, motion pictures, etc.). Corporations regulated by other federal institutions, such as in utilities, transportation, banking, or insurance, were not included. The survey includes individual data from the 10-K reports of 973 corporations, providing the aggregate executive compensation and the individual earnings of the three highest-paid executives. SALC also includes standard accounting information, among which we take the reported assets for each firm.

Table 5. Corporations classified by their assets in 1938-9: SALC and SOI

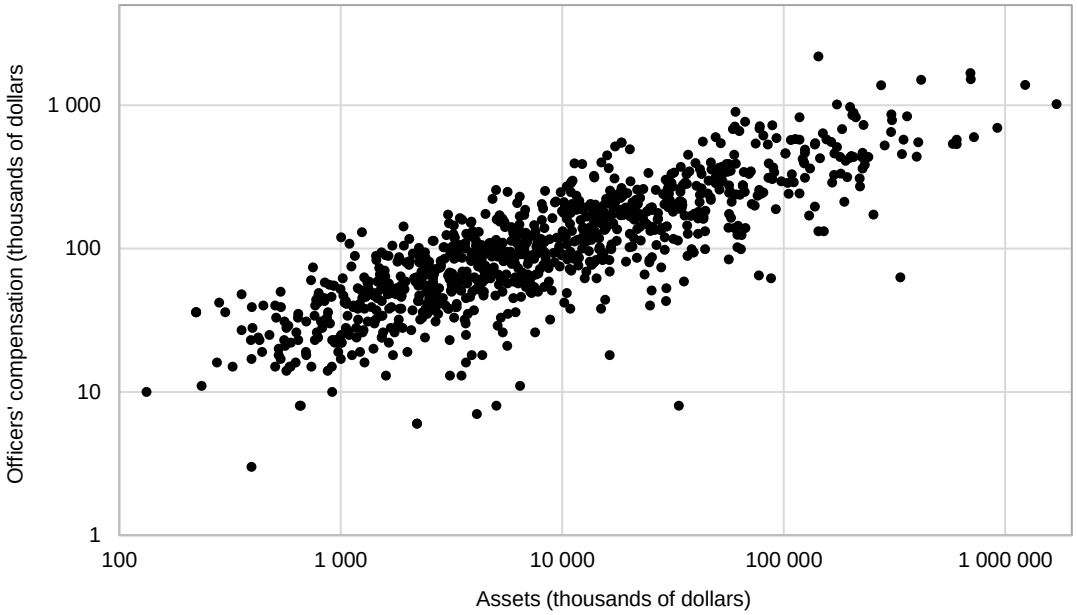
Assets (in 000s)	*SEC Survey (SALC)*			*SOI Corporate tax*
	Year 1938	*Year 1939*	*Total 1938-9*	*Year 1938*
-	-			227.491
50	1	-	1	59.582
100	3	-	3	57.733
250	12	4	16	27.371
500	42	20	62	17.079
1.000	226	84	310	17.187
5.000	110	46	156	2.542
10.000	219	52	271	2.213
50.000	56	8	64	349
100.000	75	3	78	394
Not reported	11	1	12	N.A.
Total	**755**	**218**	**973**	**411.941**

Note: For 218 companies, there is not microdata prior to 1939. Thereby we built a dataset by merging the observations of 1938 and 1939, assuming that the balance sheet size and corporate compensation would not change between these two years.
Sources: SOI and SEC.

A comparison between the SALC dataset and the complete universe of corporations registered for the corporate tax (Table 5) shows that the companies in the SALC sample are substantially larger. Consequently, as we will detail later, for all the estimates we extrapolate the trends observed in each subgroup of corporations (per asset size) in the SALC dataset to their corresponding national aggregates.

The basic principle underlying our estimates is shown in Figure 13. The SALC dataset shows a strong correlation between the size of the balance sheet and the aggregate corporate compensation.

Figure 13. Corporate assets and aggregate officers' compensation. SALC firms, 1938-9



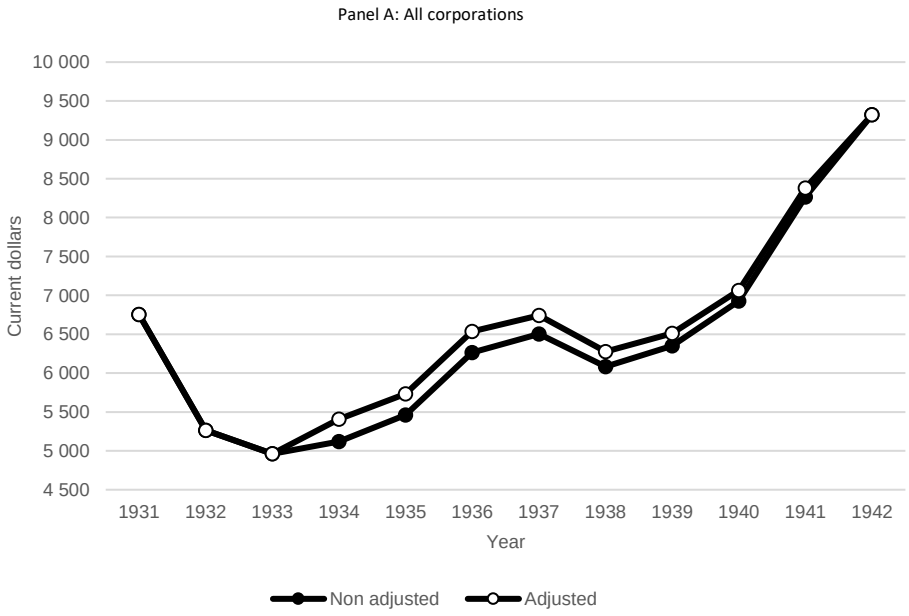
Sources: SEC (SALC).

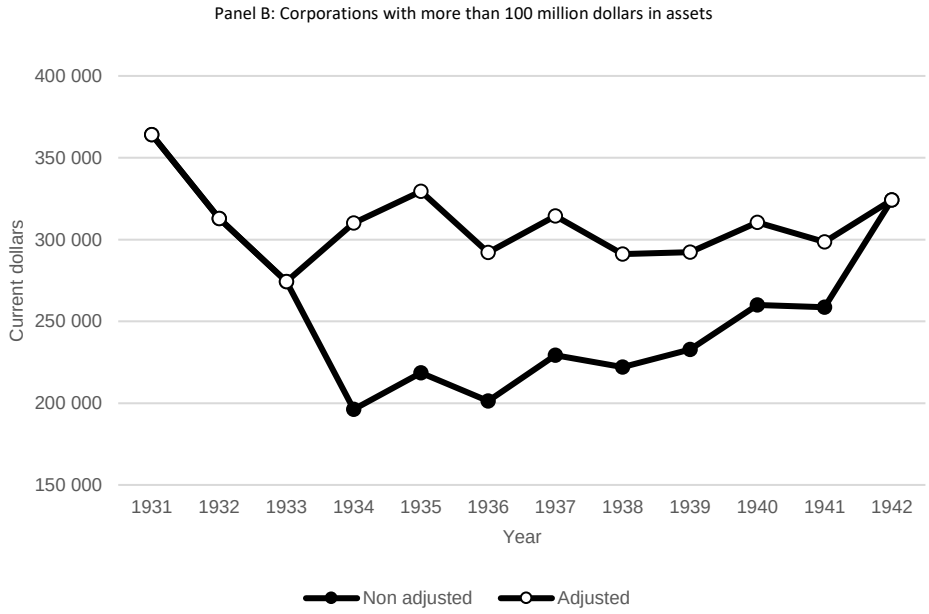
Average pay: per company and per individual

From these data inputs, we first estimate the average pay, whether at a company level or per officer. For the average pay per company, we need to estimate the number of corporations at the national level, both in total and for each bracket on asset size. This information is available in the corporate tax statistics but is affected by changes in the tax code regarding the right of corporations to file a consolidated return. As explained in a recent study (Kwon, Ma, and Zimmermann 2023, 68–70), until 1933 corporations had the right to consolidate with almost no restrictions, but from 1934 to 1941 this right was limited to railroad companies. From 1942 onward, corporations could again consolidate if they fulfilled certain criteria (basically, the ownership threshold of subsidiaries was set at 95%). Since in this study we are interested in using data at a consolidated level, we correspondingly adjust the 1934-41 data. To do so, we take the number of corporations in the two pair of years with changing criteria (1933-34 and 1941-42) and assume that the corresponding increase (or decrease) in the number of corporations is due to the change

in the tax code. Finally, we reassign the corporate compensation of corporations that are subsidiaries (identified as the brackets in which the number of firms decrease) and reassign this amount to the bracket of their parent company (brackets in which the number of companies increase). Figure 14 displays the average pay, both in unadjusted and adjusted terms, for all corporations and for those with assets of more than 100 million dollars. The data shows that the adjustment is almost irrelevant for the aggregate series but is of key importance to measure correctly the compensation paid by the largest companies.

Figure 14. Compensation of officers. Average pay per company, 1931-1942

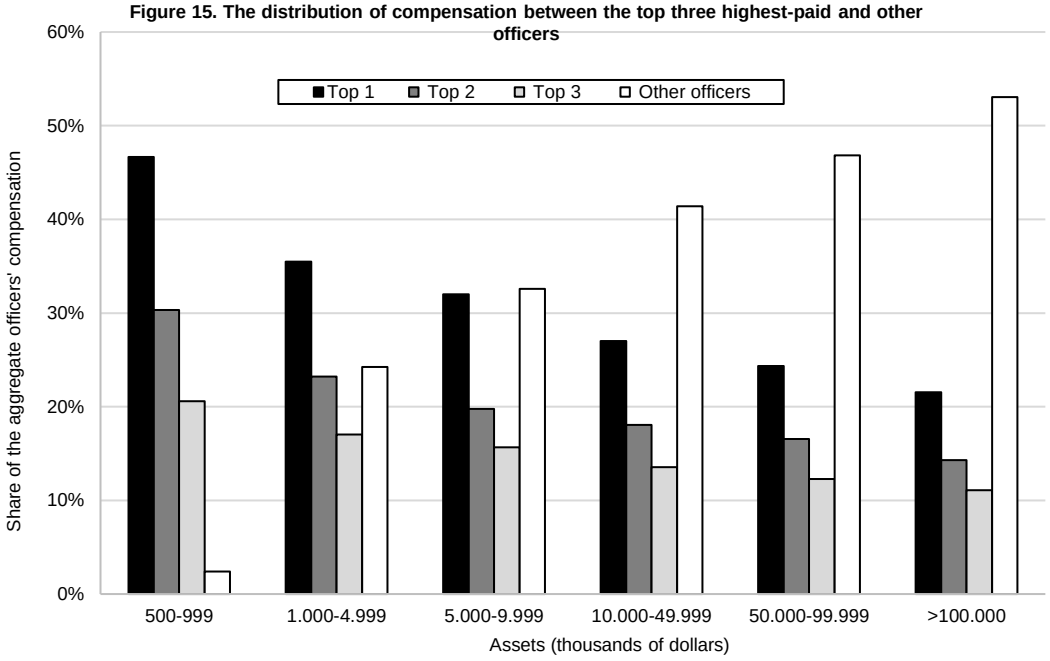




Estimates of the average pay per officer follow a similar logic, with the caveat that one requires information on the number of employed officers. Unfortunately, this information is not provided in the corporate tax statistics. As an alternative, we use the SALC dataset. In this source, we begin by computing the share of total compensation accruing to each of the three highest-paid executives and then calculate the remaining share. Figure 15 shows the results, indicating that larger corporations unequivocally had to employ a significantly higher number of officers than small ones. To reconcile these estimates, we divide the residual component by the share accruing to the third highest-paid official and round up to the nearest whole number. Using this method, we estimate that the number of officers ranges from three (in corporations with assets between half and one million dollars) to eight (in firms with more than 100 million dollars in assets). For the smallest corporations (i.e., those with assets of less than half a million dollars), we rely on the

observed correlation for the rest of the companies and assume that the number of officers oscillated between 1.5 and 2.3.[12]

Figure 15. The distribution of compensation between the top three highest-paid and other officers



Sources: SEC (SALC).

s an additional robustness check, we have computed the actual number of officers enumerated in the first complete register of officers and directors for a subsample of listed companies (Securities and Exchange Commission 1935), and the results are very consistent (Table 6).

[12] We assume that corporations follow a log-linear relationship in terms of asset size and number of officers: 1-50,000 dollars with 1.5 officers, 50,000 to 100,000 dollars 2 officers, 100,000 to 250,000 dollars with 2.3 officers and between 250,000 and 500,000 dollars with 2.8 officers.

Table 6 Number of officers per corporation, by asset size. 1936

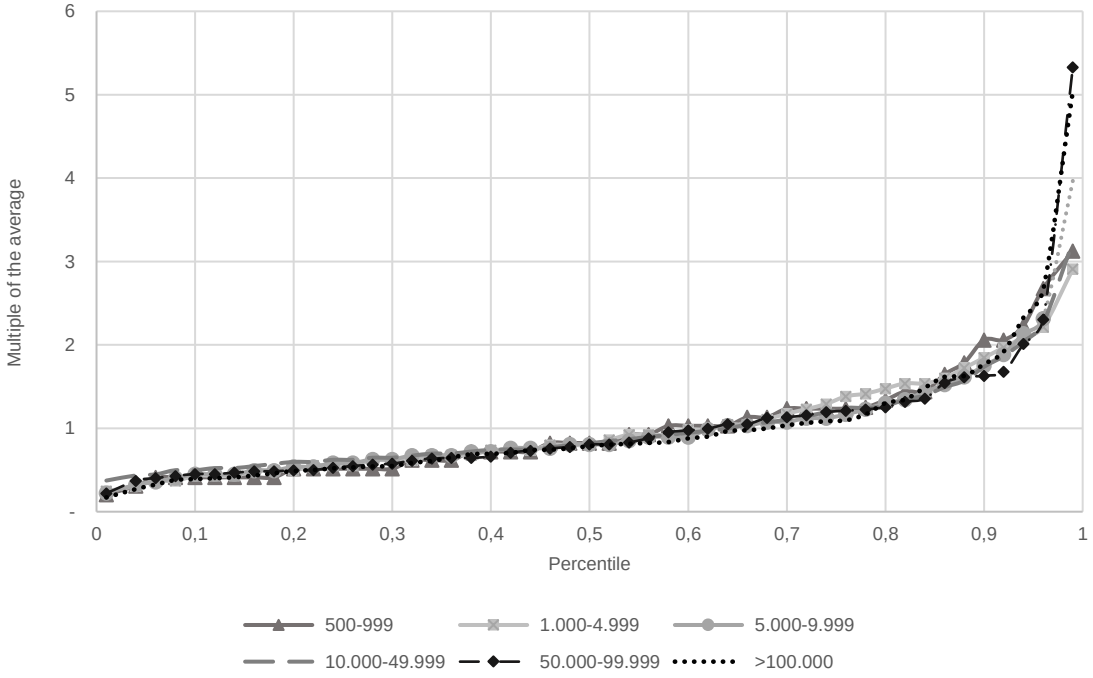
Assets (in 000s)	Number of officers
100	2,5
500	3,3
1.000	4,1
5.000	5,4
10.000	6,1
50.000	7,6
100.000	9,6

Source: Securities and Exchange Commission (1935, 1940)

The distribution of officers' compensation

The last input of our estimates refers to the variance in officers' compensation. From the SALC dataset, we draw on the data of the individual compensation for the three highest paid executives in each subgroup of corporations (per asset size) and calculate the earnings at different percentiles relative to the average of that subgroup. The results are shown in Figure 16. Note that the trends are relatively similar, irrespective of the assets size. For example, median earnings fit around 0.79-0.82 of the average. The only discernible difference is that there is slightly more variance within the group of large companies, given the existence of extremely highly paid officers (with compensation rising to 5 times the average for the 99th percentile). Thereafter, we extrapolate these trends, per asset size subgroup, to the dataset of companies covered in the corporate tax. For small-sized firms, we simply keep constant the distribution observed for SALC companies with assets between half and one million dollars.

Figure 16. Officers' compensation for different subgroups of corporations (per asset size) and percentile. SALC dataset, 1938.



Sources: SEC (SALC).

Calculating the net change in the number of corporations

The previous method provides consistent estimates of the compensation of officers for each year and group of corporations (per asset class). However, on a year-to-year basis, business activity determines the expansion (or reduction) in the balance sheets of corporations, thereby increasing (or decreasing) the number of firms in each bracket. Simultaneously, new corporations are created, and some existing ones are dissolved. Overall, the net change in the number of firms can be readily calculated from the statistics by measuring the difference between the count of corporations filing tax returns in one year compared to the subsequent year. This change is, however, not as informative at face value for the number of corporations per asset groups. The differences observed in the number of corporations in a single bracket (say, between one and five million dollars in assets) in two pairs of years are determined by i) the movement of corporations between

brackets (changes in asset size among the same population) and ii) the net creation of firms.

Typically, the number of corporations fluctuates to a small degree between years, and this issue could be mostly dismissed. However, during the Second World War period, corporate tax statistics indicate a decline in the number of corporations, together with a notable surge in the nominal volume of assets. Since these two transformations have profound implications for our understanding of the evolution of compensation for officers, we have estimated the change in the increase in balance sheets (per group of firms), and the residual as the net change in the number of active corporations. This way, we can approximate the change in corporate compensation as a residual. In this regard, the total net change in the count of corporations is distributed across brackets, and this figure can be positive or negative in each case.

To estimate variations in the corporate balance sheet, we begin by focusing on the largest group of firms –those with more than half a million dollars in assets. Utilizing both the previously mentioned SALC survey and the CRSP dataset, we observe the evolution of assets for approximately 1,000 manufacturing companies on a year-to-year basis. Table 7 presents the distribution of these companies per asset group. For all companies in a subgroup and year, we observe the change in total assets in the following year. Accordingly, we compute the percentage of companies that move between brackets (to an upper or lower group) and those that remain in the same bracket. These percentages are then extrapolated to the total population of firms recorded in the corporate tax statistics.

Table 7. Manufacturing corporations in the Survey of American Listed Corporations, 1939-1942

Asset groups (lower limit, in thousand dollars)	Year			
	1939	1940	1941	1942
500	64	64	50	38
1.000	385	376	347	309
5.000	177	190	206	213
10.000	262	269	312	336
50.000	53	66	67	70
100.000	71	74	88	104
Total	**1012**	**1039**	**1070**	**1070**

Proprietors' income: tax and census data

In the last section of the paper, we estimate the labor income of high-earning proprietors. Our aim is not to provide consistent estimates for the full period (1918-49) in the same detailed manner as with wages. Instead, we focus on obtaining two consistent estimates for the years covered by the census (1939 and 1949). This approach allows us to measure the relative change in the number of highly paid salaried and self-employed workers.

Tax records on sole-proprietorships and partnerships

The 1940 census did not provide information on proprietors' incomes; thus, we rely on income tax data. In the income tax, earnings from sole proprietorships and partnerships were reported separately. In 1939, the Bureau of Internal Revenue conducted an extensive survey encompassing all partnerships registered in the country (Bureau of Internal Revenue Statistical Section Income Tax 1945b), irrespective of whether individual partners filed a tax return. Notably, this detailed report offers invaluable insights into the number of partnerships across various industries, the count of partners, and the net income distributed across brackets. We construct an estimation of partnership income distribution at the individual level for each industry by assuming the same distribution of partnership income among firms as among its partners. This estimated distribution of partnership income is very close to the actual observations that can be derived from the individual income tax report for that year but has the advantage of preserving the original information of the industry.

To estimate sole-proprietorship income, we first use the traditional SOI tabulations, which provided the distribution of business income at the industry level in 1938 and 1939 for returns with over $5,000 in net income. Then, in a final step, we can estimate the income of sole proprietors with a net income under $5,000 by assuming that, on aggregate, they earned the difference between the NIPA totals (by industry) and the estimated income of partnerships and sole proprietors reported in SOI.

The 1950 census records

Labor and capital shares in proprietors' incomes

Self-employed income remunerates both the labor and capital provided by the proprietor. Economists have long debated the most effective method to distinguish between these types of returns. Overall, the approach in macroeconomic studies that primarily uses national accounts is substantially different than in distributional studies, which mostly use microdata. In this paper, our methodology aligns with recent studies utilizing accounting data reported by proprietors and comparing them to corporations. Under the assumption that both types of entities should exhibit relatively similar business ratios, such as the return on equity, we infer a capital and labor share within proprietors' income. This adjustment process is more precise the more detailed the accounting information is in terms of the industry classification.

This adjustment process can only be done for one single year at the end of the analysed period. In 1953, the IRS published a detailed report on partnerships that included, for the first time, information on the balance sheet of these firms (U.S. Treasury Department, Internal Revenue Service 1957). With this information, one can easily compute a rate of return by dividing the net income by the capital of partners.

In a second step, we derive the pre-tax return on equity of corporations, following the same industry classification. Finally, in a third step, partnership income is divided between capital and labor by assuming that firms should have the same return on capital as the corporations in the same industry, and the residual constitutes the labor share of partnership income. Table 8 summarizes these results. It is important for the reader to note that the labor and capital shares attributed to major industries (such as manufacturing, finance, services, etc.) are derived proportional to the weightings of minor groups in partnership income. Since this level of detail is not available for sole-proprietorship and partnership income in the preceding years, we rely on the labor and capital shares at the major industry groups in 1953.

Table 8. Imputed labor and capital shares of partnership income, 1953

Industries	Corporations	Partnerships		
	Pretax ROE	Pretax rate of return (net income / capital)	*Imputed capital share*	*Imputed labor share*
Mining	11,8%	11,8%	52,9%	47,1%
Construction	13,8%	44,8%	23,6%	76,4%
Manufacturing	18,7%	35,9%	30,7%	69,3%
Beverages	13,1%	30,5%	30,0%	70,0%
Food and kindred products	14,0%	21,3%	39,6%	60,4%
Textile products	7,3%	25,6%	22,2%	77,8%
Apparel and products made from fabrics	8,5%	45,0%	15,8%	84,2%
Lumber and wood products, except furniture	11,5%	13,2%	46,7%	53,3%
Furniture and fixtures	15,2%	42,1%	26,5%	73,5%
Printing, publishing, and allied industries	18,7%	56,0%	25,0%	75,0%
Chemicals and allied products	15,1%	23,5%	39,2%	60,8%
Stone, clay, and glass products	20,2%	52,1%	27,9%	72,1%
Fabricated metal products, exc. machinery & trans. equip.	19,1%	52,1%	26,8%	73,2%
Machinery, except transportation equipment and electrical	20,9%	57,0%	26,8%	73,2%
Other manufacturing industries	29,1%	44,2%	39,7%	60,3%
Transportation, communication and public utilities	10,3%	41,1%	20,0%	80,0%
Wholeshale trade	12,4%	29,9%	29,2%	70,8%
Retail trade	11,9%	31,5%	27,3%	72,7%
Finance, insurance and real estate	13,8%	19,5%	34,9%	65,1%
Security and commodity exchange brokers and dealers	12,0%	23,4%	33,9%	66,1%
Other finance	9,7%	11,0%	46,7%	53,3%
Insurance	32,0%	84,5%	27,5%	72,5%
Real estate	8,8%	15,4%	36,4%	63,6%
Combination real estate and insurance	8,8%	15,5%	36,2%	63,8%
Services	12,5%	77,4%	12,9%	87,1%
Hotels and other lodging places	9,3%	6,6%	58,6%	41,4%
Total personal services	10,6%	49,6%	17,6%	82,4%
Business services	18,1%	109,3%	14,2%	85,8%
Automobile repair services and garages	11,7%	71,0%	14,2%	85,8%
Miscellaneous repair services	13,3%	72,5%	15,5%	84,5%
Motion picture theaters	7,6%	17,3%	30,4%	69,6%
Other amusements and recreational services	18,4%	42,2%	30,3%	69,7%
Medical services	17,3%	209,4%	7,6%	92,4%
Legal services	17,3%	236,0%	6,8%	93,2%
Educational services	17,3%	73,2%	19,1%	80,9%
Engineering and architectural services	17,3%	121,7%	12,4%	87,6%

Sensitivity analysis

Four simultaneous equation models have been estimated in the main paper. All specifications follow Amemiya (1978), Heckman (1978), and Maddala (1983). These authors developed two-equation theoretical models in which the dependent variables (one continuous and the other dichotomous) are jointly determined.[13] This introduces a potential challenge of reverse causality, necessitating correction through the employment of instruments for the dependent (endogenous) variables.

[13] See Alesina et al., (1996); Keshk (20039; Keshk et al., (2004); Hegre et al., (2010) for empirical appplications.

The estimation process is as follows: i) In the first-stage regressions, the dependent variables are regressed on all exogenous variables; ii) Predicted values for the endogenous variables are derived from the first-stage regressions and incorporated as regressors, alongside the exogenous variables, in the second-stage equations; iii) Standard errors are corrected, as the initial standard errors were based on the instrumented endogenous variables rather than the appropriate original variables.

Following this process, Equations 3 and 4 represent the first stage of the benchmark model:

$$\widehat{lnE_i} = \beta_1\, Educ_i + \beta_2\, Age_i + \beta_3\, Age_i^2 + \beta_4\, Gender_i + \beta_5\, Capitalind_i + \beta_6\, Clerical_i + \beta_7\, Manager_i + \varepsilon_i \quad (3)$$

$$\widehat{Self_i} = \beta_1\, Educ_i + \beta_2\, Age_i + \beta_3\, Age_i^2 + \beta_4\, Gender_i + \beta_5\, Capitalind_i + \beta_6\, Clerical_i + \beta_7\, Manager_i + \varepsilon_i \quad (4)$$

where $\widehat{lnE_i}$ is the instrument for the log of earnings (the sum of wages and salaries and business owner income), $\widehat{Self_i}$ represents the instrument for the dummy variable indicating self-employment. The remaining variables are exogenous factors as outlined in the main paper. β_i are estimated coefficients in the first-stage regression and ε_i represents the residuals.

From these equations, the instruments are estimated and included in the second-stage regressions (Equations 5 and 6):

$$lnE_i = \gamma_1 \widehat{Self_i} + \beta_1\, Educ_i + \beta_2\, Age_i + \beta_3\, Age_i^2 + \beta_4\, Gender_i + \varepsilon_i \quad (5)$$

$$Self_i = \gamma_2 \widehat{lnE_i} + \beta_1\, Educ_i + \beta_2\, Age_i + \beta_3\, Gender_i + \beta_4\, Capitalind_i + \beta_5\, Clerical_i + \beta_6\, Manager_i + \varepsilon_i \quad (6)$$

Table 9 presents the first-stage regressions of the models introduced in the main paper. Additionally, we provide several alternative models in Table 10 as robustness checks. All these models use the baseline specification, excluding top-coded incomes. In model 4, we focus on the manager and proprietor occupational group. As previously mentioned, this occupational category was heavily overrepresented both in the top 1% of the wage distribution and among proprietors. As shown, the results of this model barely change from those in the Table 5, lending support to the robustness of the main results.

The remaining models included in Table 10 estimate the benchmark specification again, but they incorporate only the data of one specific industry. Due to space constraints, we only present the results for the professional service, finance, and construction industries, as they had a higher presence of proprietors. As demonstrated, we still observe a positive effect of the tax base on the likelihood of being a proprietor in all sectors except mining (which is not included). Furthermore, the absence of a positive earnings premium for proprietors is evident in professional service and finance, although this does not hold true in other industries such as construction, transport, and manufacturing.

Table 9. Benchmark models (first-stage regressions)

	Model 1		Model 2		Model 3		Model 4	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VARIABLES	Mincer eq.	Probit eq.	Mincer eq.	Probit eq.	Mincer eq.	Probit eq.	Mincer eq.	Probit eq.
I_self								
I_lnE								
Educ	0.058***	0.016***	0.056***	0.016***	0.062***	0.023***	0.058 ***	0.017***
	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.002)
Age	0.102***	0.045***	0.085***	0.034***	0.102***	0.048***	0.101***	0.045***
	(0.001)	(0.002)	(0.001)	(0.002)	(0.001)	(0.002)	(0.001)	(0.002)
Age2	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***
	(0.000)	(0.000)	(0.000)	(0.017)	(0.000)	(0.000)	(0.000)	(0.000)
Gender	0. 653***	0.318***	0.466***	0.200***	0.667***	0.348***	0.644***	0.318***
	(0.005)	(0.014)	(0.006)	(0.017)	(0.005)	(0.014)	(0.005)	(0.014)
Marital			0.062***	0.148***				
			(0.005)	(0.016)				
Capitalind	0.181***	0.121***	0.169***	-0.417***	0.179***	-0.423***	0.187***	-0.409***
	(0.004)	(0.012)	(0.004)	(0.012)	(0.004)	(0.012)	(0.004)	(0.012)
Clerical	0.295***	-0.785***	0.293***	-0.786***	0.288***	-0.789***	0.030***	-0.785***
	(0.006)	(0.030)	(0.006)	(0.030)	(0.006)	(0.029)	(0.006)	(0.030)
Manager	0.227***	1.341***	0.198***	1.327***	0.299***	1.308***	0.128***	1.342***
	(0.007)	(0.013)	(0.007)	(0.013)	(0.007)	(0.013)	(0.007)	(0.013)
Faminc			-0.001***	-0.001***				
			(0.000)	(0.000)				
Constant	4.06***	-3.125***	4.668***	-2.847***	3.987***	-3.289***	4.073***	-3.125***
	(0.018)	(0.056)	(0.021)	(0.064)	(0.018)	(0.055)	(0.018)	(0.056)
Observations	145,961	145,961	145,961	145,961	148,325	148,325	145,961	145,961
R^2	0.2754		0.2945		0.2900		0.2622	
Pseudo R^2		0.2388		0.2416		0.2431		0.2388

Standard errors in parentheses

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

Notes: This table shows first stage regressions corresponding to the model presented in Table 5 (main text). All estimated models follow the 2SLS approach outlined by Amemiya (1978) and Maddala (1983) utilizing data from the individuals in the 'sample-line' of the 1% IPUMS sample corresponding to the 1950 census, and focusing on the non-agricultural U.S. economy. Each model comprises two simultaneous equations. The first one estimates a Mincer equation using OLS and the second one estimates the likelihood to become a self-employed worker. Only workers with positive wage and/or self-employed income were included in gross earnings. Models 1 and 2 exclude top-coded income (earnings exceeding $10,000) and use original gross earnings. Model 4 includes top-coded incomes and uses original gross earnings. Model 4 again excludes top-coded income, but it adjusts self-employed income to account for labor income only.

Table 10. Alternative models (second-stage regressions)

	Model 4		Model 5		Model 6		Model 7	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VARIABLES	Mincer eq.	Probit eq.	Mincer eq.	Probit eq.	Mincer eq.	Probit eq.	Mincer eq.	Probit eq.
I_self	-0.245***		0.132***		-0.041**		-0.052**	
	(0.074)		(0.020)		(0.017)		(0.023)	
I_lnE		0. 314***		0.256***		0.386***		0.493***
		(0.001)		(0.076)		(0.043)		(0.153)
Educ	0.033***	-0.086	0.058***	0.017**	0.085***	0.003	0.060***	0.032***
	(0.006)	(0.000)	(0.0025)	(0.007)	(0.001)	(0.002)	(0.003)	(0.012)
Age	0.077***	0.004***	0.089***	0.019***	0.109***	0.015***	0.086***	0.024***
	(0.004)	(0.000)	(0.003)	(0.001)	(0.002)	(0.001)	(0.004)	(0.002)
Age2	-0.001***		-0.001***		-0.001***		-0.001***	
	(0.000)		(0.000)		(0.000)		(0.000)	
Gender	0. 735***	0.003	0.274***	-0.273 *	0.721***	0.343***	0.432***	-0.031
	(0.026)	(0.971)	(0.055)	(0.158)	(0.016)	(0.036)	(0.027)	(0.104)
Marital		0.074***						
		(0.041)						
Capitalind		-0.248***						
		(0.032)						
Clerical				-1.174***		-0.777***		-1.174*
				(0.200)		(0.060)		(0.031)
Manager				1.310***		0.915***		0.128***
				(0.055)		(0.376)		(0.072)
Faminc								
Constant	5.105***	-1.538***	4.997***	-3.792***	3.358***	-4.789***	4.637***	-6.491***
	(0.085)	(0.617)	(0.119)	(0.516)	(0.077)	(0.242)	(0.152)	(1.038)
Observations	12,757	12,757	10,676	10,676	28,717	28,717	5,377	5,377
R^2	0.1515		0.1414		0.3008		0.2254	
Pseudo R^2		0.0383		0.1359		0.1535		0.2023

Corrected standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Notes: All estimated models follow the 2SLS approach outlined by Amemiya (1978) and Maddala (1983) utilizing data from the individuals in the 'sample-line' of the 1% IPUMS sample corresponding to the 1950 census, and focusing on the non-agricultural U.S. economy. The table only shows the second-stage regressions. Each model comprises two simultaneous equations. The first one estimates a Mincer equation using OLS and the second one estimates the likelihood to become a self-employed worker. Only workers with positive wage and/or self-employed income were included in gross earnings. Models 4 only includes managers in the analysis (occupational variables cannot hence be included). Model 5 shows the benchmark regressions using only observations from the construction industry. Model 6 uses only data from the professional service sector. Model 7 includes only observations of the financial industry.

References

Alesina, Alberto, Sule Özler, Nouriel Roubini, and Phillip Swagel. 1996. "Political Instability and Economic Growth." *Journal of Economic Growth* 1(2): 189–211.

Amemiya, Takeshi. 1978. "The Estimation of a Simultaneous Equation Generalized Probit Model." *Econometrica: Journal of the Econometric Society*: 1193–1205.

Atkinson, A. B., and A. J. Harrison. 1978. *Distribution of Personal Wealth in Britain*. Cambridge ; New York: Cambridge University Press.

Atkinson, A. B., and S. Voitchovsky. 2011. "The Distribution of Top Earnings in the UK since the Second World War." *Economica* 78(311): 440–59.

Auten, Gerald, and David Splinter. 2023. "Income Inequality in the United States: Using Tax Data to Measure Long-Term Trends." *Journal of Political Economy*. doi:10.1086/728741.

Autor, David, Claudia Goldin, and Lawrence F. Katz. 2020. "Extending the Race between Education and Technology." In *AEA Papers and Proceedings*, American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203, 347–51.

Bailey, Martha J., and William J. Collins. 2006. "The Wage Gains of African-American Women in the 1940s." *The Journal of Economic History* 66(3): 737–77. doi:10.1017/S0022050706000313.

Bernanke, Ben S. 1986. "Employment, Hours, and Earnings in the Depression: An Analysis of Eight Manufacturing Industries." *The American Economic Review* 76(1): 82–109.

Blanchet, Thomas, Juliette Fournier, and Thomas Piketty. 2022. "Generalized Pareto Curves: Theory and Applications." *Review of Income and Wealth* 68(1): 263–88. doi:10.1111/roiw.12510.

Bureau of Internal Revenue Statistical Section Income Tax. 1945a. *Preliminary Study of Individual Income Tax Returns for 1944, Forms W-2 and Forms 1040 with Adjusted Gross Income Under $25,000*. Bureau of Internal Revenue, Statistical Section, Income Tax Unit.

———. 1945b. *Supplement to Statistics of Income for 1939, Part 1: Compiled from Partnership Returns of Income*. Washington: U.S. Government Printing Office.

Collins, William J., and Gregory T. Niemesh. 2019. "Unions and the Great Compression of Wage Inequality in the US at Mid-Century: Evidence from Local Labour Markets." *The Economic History Review* 72(2): 691–715. doi:10.1111/ehr.12744.

Department of the Treasury. 1940. “The Estimated Number and Compensation of Federal, State, and Local Officers and Employees, 1937-1938.” *Bulletin of the Treasury Department*: 1–3.

Fisher-Post, Matthew. 2020. “Examining the Great Leveling: New Evidence on Mid-Century American Inequality.” *WID.world Working Papers series* (01).

Frydman, Carola, and Raven Molloy. 2012. “Pay Cuts for the Boss: Executive Compensation in the 1940s.” *The Journal of Economic History* 72(01): 225–51. doi:10.1017/S002205071100249X.

Frydman, Carola, and Raven E. Saks. 2010. “Executive Compensation: A New View from a Long-Term Perspective, 1936-2005.” *Review of Financial Studies* 23(5): 2099–2138.

Geloso, Vincent, and Phillip Magness. 2020. “The Great Overestimation: Tax Data and Inequality Measurements in the United States, 1913–1943.” *Economic Inquiry* 58(2): 834–55. doi:10.1111/ecin.12865.

Geloso, Vincent, Phillip Magness, John Moore, and Philip Schlosser. 2022. “How Pronounced Is the U-Curve? Revisiting Income Inequality in the United States, 1917–60.” *The Economic Journal* 132(647): 2366–91. doi:10.1093/ej/ueac020.

Goldin, Claudia, and Lawrence F. Katz. 2010. *The Race between Education and Technology*. Cambridge, Mass.: The Belknap Press of Harvard University Press.

Goldin, Claudia, and Robert A. Margo. 1992. “The Great Compression: The Wage Structure in the United States at Mid-Century.” *The Quarterly Journal of Economics* 107(1): 1–34. doi:10.2307/2118322.

Gómez León, María, and Herman J. De Jong. 2019. “Inequality in Turbulent Times: Income Distribution in Germany and Britain, 1900–50.” *The Economic History Review* 72(3): 1073–98. doi:10.1111/ehr.12770.

Hanes, Christopher. 2000. “Nominal Wage Rigidity and Industry Characteristics in the Downturns of 1893, 1929, and 1981.” *American Economic Review* 90(5): 1432–46. doi:10.1257/aer.90.5.1432.

Hannah, Leslie. 2015. “A Global Corporate Census: Publicly Traded and Close Companies in 1910.” *The Economic History Review* 68(2): 548–73.

Hatton, Timothy J., and Mark Thomas. 2010. “Labour Markets in the Interwar Period and Economic Recovery in the UK and the USA.” *Oxford Review of Economic Policy* 26(3): 463–85. doi:10.1093/oxrep/grq023.

Heckman, James J. 1978. “Dummy Endogenous Variables in a Simultaneous Equation System.” *Econometrica* (46): 931–59.

Hegre, Håvard, John R Oneal, and Bruce Russett. 2010. “Trade Does Promote Peace: New Simultaneous Estimates of the Reciprocal Effects of Trade and Conflict.” *Journal of Peace Research* 47(6): 763–74. doi:10.1177/0022343310385995.

Household Finance Corporation. 1944. *Income Tax Guide*. Chicago: Household Finance Corporation.

Humphrey, Hubert H. 1994. “August 29, 1950: Revenue Act of 1950.” In *The Senate 1789-1989. Classical Speeches 1830-1993. Volume Three. Bicentennial Edition*, Washington D.C.: U.S. Government Printing Office, 625–50.

Jaworski, Taylor, and Gregory T. Niemesh. 2018. “Revisiting the Great Compression: Wage Inequality in the United States, 1940–1960.” *Historical Methods: A Journal of Quantitative and Interdisciplinary History* 51(1): 39–48. doi:10.1080/01615440.2017.1393360.

Keshk, Omar M. G. 2003. “CDSIMEQ: A Program to Implement Two-Stage Probit Least Squares.” *Stata Journal* 3(2): 157–67.

Keshk, Omar M. G., Brian M. Pollins, and Rafael Reuveny. 2004. “Trade Still Follows the Flag: The Primacy of Politics in a Simultaneous Model of Interdependence and Armed Conflict.” *The Journal of Politics* 66(4): 1155–79. doi:10.1111/j.0022-3816.2004.00294.x.

Krolage, Carla, Andreas Peichl, and Daniel Waldenström. 2022. “Long-Run Trends in Top Income Shares: The Role of Income and Population Growth.” *The Journal of Economic Inequality* 20(1): 97–118. doi:10.1007/s10888-021-09520-8.

Kuznets, Simon. 1953. *Shares of Upper Income Groups in Income and Savings*. New York: NBER.

———. 1954. *National Income and Its Composition, 1919-1938*. National Bureau of Economic Research.

Kwon, Spencer Yongwook, Yueran Ma, and Kaspar Zimmermann. 2023. “100 Years of Rising Corporate Concentration.” *University of Chicago, Becker Friedman Institute for Economics Working Paper* 2023(20). doi:10.2139/ssrn.3936799.

Liebenberg, Maurice, and Hyman Kaitz. 1951. “An Income Size Distribution from Income Tax and Survey Data, 1944.” In *Studies in Income and Wealth, Volume 13*, NBER, 378–462.

Lindert, Peter H, and Jeffrey G Williamson. 1983. “Reinterpreting Britain’s Social Tables, 1688–1913.” *Explorations in Economic History* 20(1): 94–109. doi:10.1016/0014-4983(83)90044-X.

Lindert, Peter H., and Jeffrey G. Williamson. 2016. *Unequal Gains: American Growth and Inequality since 1700*. Princeton, NJ: Princeton University Press.

Maddala, Gangadharrao S. 1983. *Limited-Dependent and Qualitative Variables in Econometrics*. Cambridge: Cambridge University Press.

Maloney, Thomas N. 1994. "Wage Compression and Wage Inequality Between Black and White Males in the United States, 1940–1960." *The Journal of Economic History* 54(2): 358–81. doi:10.1017/S0022050700014522.

Mas, Alexandre. 2019. "Does Disclosure Affect CEO Pay Setting?: Evidence from the Passage of the 1934 Securities and Exchange Act." *Princeton University, Department of Economics, Industrial Relations Section* (632).

Milanovic, Branko, Peter H. Lindert, and Jeffrey G. Williamson. 2011. "Pre-Industrial Inequality." *The Economic Journal* 121(551): 255–72. doi:10.1111/j.1468-0297.2010.02403.x.

Modalsli, Jørgen. 2015. "Inequality in the Very Long Run: Inferring Inequality from Data on Social Groups." *The Journal of Economic Inequality* 13(2): 225–47. doi:10.1007/s10888-014-9279-6.

Moriguchi, Chiaki, and Emmanuel Saez. 2008. "The Evolution of Income Concentration in Japan, 1886–2005: Evidence from Income Tax Statistics." *Review of Economics and Statistics* 90(4): 713–34. doi:10.1162/rest.90.4.713.

Philippon, Thomas, and Ariell Reshef. 2012. "Wages and Human Capital in the U.S. Finance Industry: 1909–2006." *The Quarterly Journal of Economics* 127(4): 1551–1609. doi:10.1093/qje/qjs030.

Piketty, Thomas. 2001. *Les hauts revenus en France au XXe siècle: inégalités et redistributions, 1901-1998*. Paris: B. Grasset.

———. 2014. *Capital in the Twenty-First Century*. Cambridge, Mass.: Belknap Press of Harvard University Press.

Piketty, Thomas, and Emmanuel Saez. 2003. "Income Inequality in the United States, 1913–1998." *The Quarterly Journal of Economics* 118(1): 1–41.

———. 2007. "Income and Wage Inequality in the United States, 1913-2002." In *Top Incomes over the Twentieth Century: A Contrast between Continental European and English-Speaking Countries*, eds. Anthony B. Atkinson and Thomas Piketty. Oxford; New York: Oxford University Press, 141–225.

Ruggles, Steven et al. 2021. "IPUMS Ancestry Full Count Data: Version 3.0." *Minneapolis, MN: IPUMS*.

Securities and Exchange Commission. 1935. *Official Summary of Holdings of Officers, Directors and Principal Stockholders*. Washington D.C.: Government Printing Office.

———. 1940. *Survey of American Listed Corporations*. New York: Securities and Exchange Commission.

Stigler, George J. 1956. *Employment in the Service Industries*. Princeton, NJ: Princeton University Press.

United States, Bureau of Internal Revenue. Various years. *Statistics of Income. Compiled from Corporation Income and Excess-Profits Tax Returns*. Washington D.C.: U.S. Government Printing Office.

———. Various years. *Statistics of Income. Compiled from Income Tax Returns and Including Statistics from Estate Tax Returns and Gift Tax Returns*. Washington D.C.: U.S. Government Printing Office.

United States. Office of Business Economics. 1953. *Income Distribution in the United States by Size, 1944-1950*. Washington D.C.: U.S. Government Printing Office.

U.S. Treasury Department, Division of Tax Research. 1940. *Statistics of Income Supplement Compiled from Federal Income Tax Returns for 1936: Individual Incomes*. Washington.

U.S. Treasury Department, Internal Revenue Service. 1957. *Statistics of Income: Partnership Returns 1953*. Washington D.C.: Government Printing Office.

US Treasury, United States Research and Statistics Division. 1938. *Statistics of Income Supplement, Compiled from Federal Income Tax Returns of Individuals for the Income Year 1934. Section 1: Number of Individual Income Tax Returns for 1934 Classified by Counties, and Cities of 25,000 and Over Population by Net Income Classes*. Washington: Government Printing Office.