

Do Entrepreneurs Care About Tax?

A U.S. study on the impact of taxation on entrepreneurial activity and its implications for economic inequality

Hanna Johansson (25001) & Markus Lind (25443)

Stockholm School of Economics - Spring 2026

Abstract: This thesis examines the relationship between federal taxation, entrepreneurship, and economic inequality in the U.S. over the period 1989–2022. Using data from the SCF, CBO, and NBER, the study investigates whether entrepreneurs are disproportionately concentrated at the top of the income and wealth distributions, whether changes in tax rates affect households' decisions to become or remain entrepreneurs, and whether entrepreneurship constitutes a relevant channel through which taxation may influence economic inequality. The empirical analysis combines multiple model specifications and complementary methodological approaches to assess long-run distributional patterns and associations between taxation and entrepreneurial activity; descriptive distributional analysis, household-level regressions, first-difference estimations, event-study specifications, and matched synthetic panel and aggregate cell-level analyses. Analyses of the impact of tax progressivity, lagged taxation effects, and the decomposition of different federal taxes provide additional nuance.

The study confirms previous research's findings: entrepreneurial households are strongly overrepresented among high-income and high-wealth households, highlighting the importance of understanding entrepreneurial activity as an inequality-relevant outcome. However, the evidence on the relationship between taxation and entrepreneurship is more mixed, with results suggesting that taxation's direct influence on a household's entrepreneurial activity varies strongly across model specifications and different tax measures. Linking entrepreneurship back to inequality measures beyond descriptive statistics generates puzzling findings, providing limited evidence of a relationship between income inequality and entrepreneurship rates. Overall, the findings suggest that entrepreneurship could be an important mechanism linking taxation and inequality together, but empirical evidence connecting behavioral effects of taxation on entrepreneurial decisions appears limited and context-dependent with the available data.

Acknowledgements

We would like to express our sincere gratitude to our supervisor, Gualtiero Azzalini, for the invaluable support and patience throughout the writing of this thesis. Your encouragement, availability, feedback, and continuous engagement have not gone unnoticed and been greatly appreciated. Your insights and guidance have not only played an essential role in the completion of this work but also contributed greatly to our academic development throughout this process.

Thank you.

Keywords: Federal Taxation, Entrepreneurship, Inequality, United States

JEL: D31, H24, H31, L26

Table of Contents

1. Introduction	1
1.1. Background	1
1.2. Literature Review	2
1.3. Our contribution	5
1.4. Research questions	5
1.5. Thesis outline	5
2. Methodology	6
2.1. Empirical method	6
2.2. A note on the U.S. tax system	7
2.3. Data	8
2.4. Variable construction	9
2.4.1. Entrepreneurs	9
2.4.2. Gini coefficients	10
2.4.3. TAXSIM tax rates	11
2.5. Summary statistics	11
2.6. Model identifications	14
2.6.1. Pooled income-group regressions	14
2.6.2. Tax progressivity regressions	16
2.6.3. Policy-event study	17
2.6.4. Matched synthetic household panel	17
2.6.5. Effective tax rate cohort-by-education cells	19
2.6.6. Marginal tax rate cohort-by-education cells	19
2.6.7. Tax progressivity and inequality regressions	20
2.7. Statistical remarks	21
3. Entrepreneurs and taxation	23
3.1. Pooled income-group regressions	23
3.2. Tax progressivity regressions	25
3.3. Policy-event study	27
3.4. Matched synthetic household panel	30
3.5. Effective tax rate cohort-by-education cells	32
3.6. Marginal tax rate cohort-by-education cells	33
4. The link to inequality	34
4.1. Entrepreneurs and inequality	34
4.2. Taxation and inequality	38
5. Conclusions	43
6. References	45
7. Appendix A - Model specifications	47
7.A. Income and wealth	47
7.B. TAXSIM tax rates	47

7.C. Matching algorithm	47
7.D. Standard errors correction	49
8. Appendix B - Complementary statistics	50
9. Appendix C - Robustness specifications	52
9.A. Tax progressivity	52
9.B. Lagged TAXSIM tax	53
9.C. Top-tail mobility	54
9.D. Wealth inequality	55

1. Introduction

1.1. Background

The notion of rapidly growing inequality has gained momentum in recent years, with strong voices calling for actions against the deterioration of the middle class due to a rapid wealth accumulation by the ultra-rich in several western countries.¹ This has been especially prevalent since the COVID-19 pandemic, during and after which the accumulation rate saw unprecedented speed,² with the top 10 percent now holding 75 percent of the total global wealth.³ The implications of vast inequality are disputed, with innovation and job-creation advocates on one hand, and fairness and working-class protectors on the other.⁴ The innovation argument in particular has gotten a central role in the debate, highlighting the need for economic incentives to innovate and become entrepreneurial.⁵

How to tackle growing inequality is an additional aspect scholars as well as practitioners are not fully agreeing on, but taxation has grown to be a generally accepted method of correcting for the gap amongst income groups.⁶ Although, the most effective type of taxation is debated.⁷ The extent of redistributive effects generated from wealth taxes and (or) increased capital gains and income taxes are not fully understood, especially through which channels such redistributive effects primarily operate. One theory is that entrepreneurs, commonly being overrepresented in the top percentiles of the wealth and income distributions, are an essential fraction of the population through which inequality could be channeled. Whether an additional tax burden (or relief) would impact the fraction of people who choose to establish or maintain a business, and thereby redistributing income and wealth and have extended effects on inequality, is what this thesis is intended to examine.

Our results indicate that entrepreneurial households are disproportionately concentrated at the top of both the income and wealth distributions and have, during the sample period, accumulated wealth and experienced income trajectories far beyond those of their counterparts. This provides support for the relevance of entrepreneurship in understanding and tackling growing economic inequality. The findings further give indication

¹ See for example Sanders, 2026, The Guardian, and Mamdani, 2025, NBC News

² See Institute for Policy Studies, n.d.

³ See Holmes et al. 2025, The Guardian

⁴ See for example Tamny, 2013, Forbes

⁵ See for example Moynihan, 2026, New York Post

⁶ See for example Aaron, 2015, Brookings. See also section 1.2

⁷ See section 1.2

that a relationship between taxation and entrepreneurial activity may exist, although its strength and consistency vary across specifications. While the pooled household-level regressions provide limited evidence of a robust relationship between aggregate tax rates and entrepreneurship, the first-difference specifications indicate that changes in specific tax components, particularly corporate and individual income taxes, are negatively associated with entrepreneurship gaps between top- and bottom-income households.

The policy-event study examining U.S. federal tax reforms provides limited empirical evidence that tax reforms substantially alter entrepreneurial behavior among top-income households. Likewise, the synthetic household-panel analysis does not provide strong support for tax changes significantly affecting transitions into entrepreneurship. The aggregate cell-level analysis fails to provide any additional insight into entrepreneurs' relationship with either effective or marginal tax rates. Finally, the link from entrepreneurship to measured income inequality is suggestive but not empirically conclusive in our data. Despite entrepreneurs being overrepresented in top income- and wealth groups, empirical evidence supporting a direct relationship is weak and requires further research.

1.2. Literature Review

This thesis utilizes a range of current literature on the topic and primarily highlights the link between three relationships: (1) The link between taxation and entrepreneurship, (2) the link between taxation and inequality, and (3) the link between entrepreneurship and inequality.

Firstly, Boar & Midrigan (2023) examine how capital income taxes, compared to wealth taxes, impact entrepreneurial decisions and the distribution of wealth. The study attempts to answer the question of whether governments should rely on either capital income taxes or wealth taxes in an economy where private businesses play a major role in both production and inequality, such as in the U.S. In their economic model, private business owners hold a disproportionate share of both wealth and income but, at the same time, face financial constraints that generate heterogeneous returns to saving. The paper considers private businesses' contribution to production and redistribution, on one hand, and their impact on growing inequality, on the other. Their main finding is that capital income taxation dominates wealth taxation from a welfare perspective, primarily because capital income taxes have more prominent redistributive effects. An important note is that, since entrepreneurial profits are primarily concentrated among wealthy households, capital income taxation allows

the government to transfer resources effectively toward non-entrepreneurs and lower-wealth households through lump-sum transfers.

Secondly, Guvenen et al. (2023) examine the link between taxation and inequality, also by analyzing the efficiency and redistributive effects of wealth taxation relative to capital income taxation, but in an economy with heterogeneous returns to wealth. The central research question is whether replacing capital income taxes with wealth taxes improves welfare after accounting for allocative efficiency and redistribution, using a model closely replicating key features of U.S. data.⁸ There is a growing amount of empirical evidence that individuals earn persistently different rates of return on their wealth and that these differences are especially present at the very top of the wealth distribution, where most entrepreneurs are present; individuals face idiosyncratic income risk and can save in risky entrepreneurial capital that yields heterogeneous but persistent returns only if they have enough wealth accumulation, given that financial frictions limit external financing. Guvenen et al. (2023)'s model reflects the mechanism of the “use-it-or-lose-it” effect of wealth taxation; a wealth tax discourages low-return or idle uses of capital and reallocates resources toward higher-return entrepreneurs, improving the aggregate efficiency through a small share of high-performers.

Beyond their main results, addressing optimal wealth taxation, they show that introducing progressivity into the wealth tax through an exemption threshold increases welfare (even if only slightly) and strengthens the redistributive effect. They provide a framework for understanding how wealth taxes influence inequality through changes in savings, investment, and entrepreneurial behavior and, together with Boar & Midrigan (2023), help us understand how tax regimes may shape entrepreneurship-driven inequality.

Thirdly, Cagetti & De Nardi (2006) focus on the role of entrepreneurship, borrowing constraints, and risk in shaping inequality. Their work highlights how financial frictions restrict entrepreneurial entry and scaling, leading to a small but extremely wealthy entrepreneurial elite. The paper is motivated by the empirical observation already mentioned; that entrepreneurs constitute a small fraction of the population yet hold a disproportionately large share of total wealth. Following this, the authors examine how the constraints affect occupational choice and whether entrepreneurship actually manages to help explain the extreme right tail of the wealth distribution observed in U.S. data.

The research found that a model disregarding entrepreneurship fails to generate the observed degree of wealth concentration, particularly at the top of the distribution. In

⁸ Particularly the highly skewed wealth distribution, the thickness of the top tail, and the presence of extremely wealthy households

contrast, a model that includes entrepreneurial choice and borrowing constraints closely matches both the aggregate wealth distribution and the distinct wealth discrepancies between entrepreneurs and non-entrepreneurs. Over time, successful entrepreneurs accumulate large fortunes which can persist across generations, generating a fat upper tail of accumulated wealth.

Quadrini (2000) also demonstrates how entrepreneurship generates wealth concentration and emphasizes the role entrepreneurs hold in shaping wealth accumulation, wealth inequality, and social mobility in the U.S. The paper was motivated by two robust empirical facts: That wealth generally is more concentrated than income, and that entrepreneurs hold a disproportionately large share of total wealth while exhibiting higher saving rates than workers, also when conditioned on income. Similar to Cagetti & De Nardi (2006), Quadrini (2000) shows that entrepreneurs in particular have distinct wealth accumulation dynamics. The substantially higher wealth-to-income ratio displayed by entrepreneurs, compared to non-entrepreneurs, indicates that their wealth accumulation is not simply driven by higher earnings. However, since entrepreneurs face higher income risk and costly external finance, their marginal return to savings is higher, generating persistent differences in savings behavior and allowing extensive accumulation of wealth by a small fraction of households over long horizons.

The role of financial frictions is interesting and crucial for their results; they limit entrepreneurial entry for low-wealth households and reduce mobility from the bottom of the wealth distribution. Thus, the study highlights the tension between entrepreneurship as a source of opportunity and entrepreneurship as a force that amplifies inequality by favoring already-wealthy households.

Both Cagetti & De Nardi (2006) and Quadrini (2000) demonstrate the impact entrepreneurs have on shaping overall inequality and the dynamics of primarily already-wealthy households being able to pursue entrepreneurial activity. In turn, the question of whether and, if so, how, massive wealth accumulation by entrepreneurs should be managed from a tax perspective has sparked debate. Heterogeneous returns to wealth break the equivalence between capital income and wealth taxation. Some, such as Guvenen et al. (2023), conclude that wealth taxation dominates, because the large efficiency gains from reallocating capital toward high-return entrepreneurs is preferred regardless of its impact on inequality. While some, such as Boar and Midrigan (2023), conclude that capital income taxation dominates because the redistribution towards non-entrepreneurs is more important than such efficiency gains.

1.3. Our contribution

Scholars have spent substantial research examining entrepreneurs' role in driving economic divide, as well as taxation's role in redistributing resources and improving overall welfare. The literature is relatively unanimous on two main points: (1) that entrepreneurs, relative to non-entrepreneurs, make up a disproportionate fraction of the very top of the wealth and income distribution, and (2) that increased taxation, with some disagreement on the specific type of taxation, generally reduces inequality. This two-step dynamic opens up the need to understand the drivers of entrepreneurial activity in order to tackle growing economic inequality. The impact of taxation on entrepreneurs in particular, and whether that could have subsequent implications for inequality, is yet to be assessed in one coherent study, which is the primary motivation for this thesis. We provide empirical evidence on entrepreneurs' contribution to wealth and income inequality, and on how households' decisions to become or remain entrepreneurs in turn are shaped by taxation.

1.4. Research questions

This thesis aims to examine the effects of taxation on entrepreneurial activity to deepen the understanding of the drivers for this small, but impactful, fraction of the population. By assessing the impact of changes in the U.S. federal tax rates on the fraction of entrepreneurs, and discussing how such changes in composition in turn affect overall inequality, we aim to answer the following research questions:

- (1) Are entrepreneurial households disproportionately overrepresented at the top of the U.S. income and wealth distributions?
- (2) Do changes in tax rates impact the fraction of households choosing to become or remain entrepreneurs?
- (3) Does the empirical evidence suggest that entrepreneurship is a relevant channel linking taxation to observed economic inequality?

1.5. Thesis outline

The thesis is outlined as follows: Chapter 2 will be dedicated to the methodology, which will provide a breakdown of the empirical method, a thorough explanation of the datasets used and the variable construction process, as well as present summary statistics. The model identifications follow thereafter, providing estimation specifications for all methods used.

The chapter will end with some important statistical remarks on the limitations related to data availability and the findings' causal strength. Chapters 3 and 4 consist of findings, including descriptive statistics, all regression models, policy-event study results, and synthetic panel and aggregate cells analyses. This also includes a discussion on the impact on inequality and its relationship with both entrepreneurs and taxation. Chapter 5 highlights the main conclusions drawn from these findings. References and Appendices are found in Chapter 6 and 7-9, respectively.

2. Methodology

2.1. Empirical method

The relationship between taxation, entrepreneurship, and inequality in the U.S. will be studied using a combination of descriptive distributional analysis, household-level regressions, first-difference estimations, event-study specifications, and matched synthetic panel and aggregate cell-level analyses. The empirical strategy uses several complementary approaches rather than relying on a single identification design, primarily to consider the Survey of Consumer Finances data (SCF)⁹ being a repeated cross-section, rather than a true household panel, and limited in its number of survey waves.¹⁰

First, we estimate household-level regressions to examine income-group tax rates relationship to household entrepreneurship status, allowing tax coefficients to vary across income groups. This includes lagged taxation rates on income group level and a first-difference model, capturing the decomposition of different federal taxes. Second, we estimate the effect of tax progressivity on the differences in entrepreneurship rate between the highest and lowest income groups to fully understand the emphasized role of the top 1 percent income group. Third, we examine four major federal tax reforms using an event-study framework that compares the top 1 percent to the 96th–99th percentiles (henceforth referred to as the top 5 percent). Fourth, we use synthetic panels to address the fact that households are not followed directly over time and an aggregate cell-level analysis to more accurately proximate entrepreneurial incentives using households' own tax rates. Lastly, we document where entrepreneurs are located in the income and wealth distributions and assess the links between entrepreneurship and taxation on economic inequality using Gini coefficients in both descriptive and regression analyses.

⁹ Details are available at <https://www.federalreserve.gov/econres/scfindex.htm>

¹⁰ See section 2.3

Since the empirical analysis uses several levels of aggregation, each specification should be interpreted at the level at which the tax variation is measured: household-level regressions assign income-group tax rates to households, progressivity regressions use income-group gaps, event studies compare top-income groups around reforms, synthetic panel regressions use matched transition cells, and simulated tax rate specifications use cohort-by-education cells.

2.2. A note on the U.S. tax system

This thesis will focus solely on U.S. federal taxes, which contain multiple components with individual income tax and corporate tax constituting the largest shares, representing 50.7 percent and 9 percent of total federal revenue in 2025, respectively.¹¹ Although we ignore state and local taxes in this study for data limitation reasons, federal taxes represent more than half of all U.S. taxes,¹² and thus remain interesting to consider on its own. The impact of local and state taxes on the overall progressivity of the U.S. tax system is debated. In this work, the approach taken by Pechman (1985), showing that state and local taxes are close to proportional to income across income groups, is adopted. Thus, ignoring state and local taxes is assumed to be of minimal consequence when assessing and discussing the overall tax progressivity.

Government transfers are not included in the federal tax rates used in this thesis. However, the main tax measures include refundable tax credits, meaning that lower-income groups may display negative average tax rates when refundable credits exceed their tax liabilities.¹³ It should be noted that, as they accrue disproportionately to the bottom of the income distribution, they also reduce economic inequality in disposable income across income groups. However, as our focus is primarily on the top of the income distribution and governmental transfers represent a very small fraction of middle- and high-income earners' incomes, ignoring them has little effect on our results and including them would risk dampening the true inequality occurring in the sample.¹⁴

¹¹ See U.S. Department of the Treasury, n.d.

¹² See US Bureau of Economic Analysis, 2023

¹³ See Figure B2 in Appendix B

¹⁴ See a similar reasoning in Piketty & Saez (2007)

2.3. Data

The analysis combines household-level SCF data with tax measures from the Congressional Budget Office (CBO) and the National Bureau of Economic Research (NBER). The SCF contains triennial cross-sectional survey information on U.S. households' income, wealth, business ownership, and demographic characteristics, conducted by the Federal Reserve. Each household is allocated a survey weight indicating how large a portion of the U.S. population it represents and statistics are then computed using these weights to ensure proper representation. Additionally, the SCF oversamples high-wealth households using replicate weights, making it particularly well suited for studying inequality and entrepreneurial activity at the very top of the income and / or wealth distribution, which otherwise would be constrained in sample size. The analysis uses the survey waves from 1989 to 2022, although the sample period varies somewhat across specifications depending on tax data availability,¹⁵ and utilizes both the SCF summary files and selected variables from the main survey files. The summary files provide harmonized variables across survey waves, which facilitates consistent analysis over time, while the main files allow construction of a stricter measure for identifying entrepreneurial households. The combined use thus allows for both comparability across years and flexibility in variable construction.

The main tax measure comes from the CBO; The Distribution of Household Income, 2022,¹⁶ and provides annual income group level measures of income and federal taxes for the period 1979 to 2022. The average federal tax rates are sorted into eight income groups based on quintiles: bottom quintile, second quintile, middle quintile, fourth quintile, 81st–90th percentiles, 91st–95th percentiles, 96th–99th percentiles,¹⁷ and the top 1 percent. These average federal tax rates in turn consist of four components: individual income tax, payroll tax, corporate income tax, and excise tax.

The NBER TAXSIM supplement to the SCF provides household-level simulated effective tax rates through 2019 and should be interpreted as approximations of household taxes, rather than observed tax payments.¹⁸ The total effective tax rate in this measure consists of two components: individual income tax and FICA payroll tax.¹⁹ Additionally,

¹⁵ Specifically, the TAXSIM ETR sample used for effective tax rates includes data up until 2019, while the TAXSIM v35 MTR sample used for marginal tax rates includes data up until 2022

¹⁶ Details are available at <https://www.cbo.gov/publication/61911>

¹⁷ I.e. what is referred to as the top 5 percent

¹⁸ Details are available at <https://taxsim.nber.org/to-taxsim/scf27-32/> and <https://taxsim.nber.org/to-taxsim/scf-2022/>

¹⁹ The FICA (Federal Insurance Contributions Act) payroll tax is used to fund both Social Security and Medicare

TAXSIM v35 provides simulated marginal tax rates for the SCF sample through the 2022 wave and is used as a complementary measure of the marginal tax incentives faced by households. This dataset is available only at the primary tax unit level,²⁰ rather than household level.

2.4. Variable construction

2.4.1. Entrepreneurs

The main outcome variable is an indicator of whether a household is classified as entrepreneurial. The definition uses the SCF dataset to identify households with (1) reported business ownership, (2) reported active involvement in managing that business, and (3) reported the business being started by the household.²¹ This definition is intended to separate active entrepreneurship, inclusive of decision-making and day-to-day business activity, from passive ownership of business assets or non-founding stakes in businesses arising from employment, i.e. partnerships.

Let $Entrepreneur_{it}$ be an indicator equal to one if household i in year t is classified as entrepreneurial, and zero otherwise. For descriptive statistics and group-level analysis, entrepreneurship rates are calculated as weighted shares using SCF survey weights, w . For income group g in year t , the weighted entrepreneurship rate becomes:

$$E_{gt} = \frac{\sum_{i \in g,t} w_{it} Entrepreneur_{it}}{\sum_{i \in g,t} w_{it}}$$

The same structure is used for synthetic panel cells. For cell c in year t , the weighted entrepreneurship rate becomes:

$$E_{ct} = \frac{\sum_{i \in c,t} w_{it} Entrepreneur_{it}}{\sum_{i \in c,t} w_{it}}$$

where cells are defined by birth cohort and education group in the synthetic panel analysis (henceforth referred to as cohort-by-education cells), allowing for examination over time despite the repeated cross-sectional structure of the SCF. A similar variable construction is

²⁰ A primary tax unit is the main tax filing unit within a household, typically the household head and spouse, and excludes any other household members who file separately

²¹ This definition is in all essence equivalent to the one used in Quadrini (2000)

made for income and wealth,²² respectively, for each year and each income group. The full specification is provided in Appendix 7.A.

2.4.2. Gini coefficients

To place entrepreneurship trends in a broader distributional context, we calculate weighted Gini coefficients for income and wealth within each SCF survey year. Let y_{it} denote household income or wealth, with i and j indexing all household pairs observed in year t , and let w_{it} denote the SCF household weight. The weighted mean becomes:

$$\mu_t = \frac{\sum_i w_{it} y_{it}}{\sum_i w_{it}}$$

The SCF weighted Gini coefficient can then be written as:

$$G_t^{SCF} = \frac{\sum_{i,j} w_{it} w_{jt} |y_{it} - y_{jt}|}{2\mu_t \left(\sum_i w_{it} \right)^2}$$

In the implementation, households are ordered from lowest to highest income or wealth, cumulative population and income or wealth shares are calculated using SCF household weights, and the Gini is obtained from the resulting weighted Lorenz curve.²³

For the CBO analysis, we construct a grouped Gini coefficient from CBO income-group data. Let g index income groups ordered from lowest to highest income, with k indexing the income groups and G denoting the total number of groups. Let s_g denote the population share and $y_{g,t}^-$ the average income of group g in year t . The cumulative population share then becomes:

$$P_{g,t} = \sum_{k=1}^g s_k$$

And the cumulative income share becomes:

$$L_{g,t} = \frac{\sum_{k=1}^g s_k y_{k,t}^-}{\sum_{k=1}^G s_k y_{k,t}^-}$$

The grouped CBO Gini is then calculated from the area under the grouped Lorenz curve:

²² Note that wealth is consistently referring to net wealth throughout this study

²³ Note that this Gini coefficient allows for negative wealth values, similarly to what is argued in Lee & Suh (2025)

$$G_t^{CBO,grouped} = 1 - 2 \sum_{g=1}^G \frac{L_{g,t} + L_{g-1,t}}{2} (P_{g,t} - P_{g-1,t})$$

This measure approximates the Lorenz curve using CBO income groups rather than household-level observations. By construction, it captures inequality across the available CBO income groups, but not within-group inequality.

2.4.3. TAXSIM tax rates

For the TAXSIM effective tax rate analysis, both federal and total effective tax rates are constructed. At the household-level, obtained through SCF, the federal effective tax rate consists of simulated federal tax liability divided by TAXSIM income:

$$ETR_{it}^{federal} = \frac{FederalTaxLiability_{it}}{TAXSIMIncome_{it}}$$

And total effective tax rate consists of federal tax liability and FICA payroll taxes:

$$ETR_{it}^{total} = \frac{FederalTaxLiability_{it} + FICA_{it}}{TAXSIMIncome_{it}}$$

Cell-level effective tax rates are constructed analogously as weighted liability-to-income ratios within each cohort-by-education cell. Further specification is provided in Appendix 7.B.

2.5. Summary statistics

A summary of the final dataset, together with main variables across the SCF survey years, is provided in Figure 1. It can be observed that there are large differences in the number of households per survey and that the mean number of kids has decreased over time, from 0.93 to 0.71. Notable is the very large increase in mean wealth, 143%, and mean income, 50%, in comparison to median wealth, 78%, and median income, 17%, during the 33 years time span. This shows how wealth and income become further concentrated at the top over time. Considering that the U.S. inflation over the same period was 136%,²⁴ it demonstrates that the median household has become poorer in real terms while the wealthy have accumulated.

²⁴ Calculated based on the U.S.'s official inflation between 1989 and 2022 using data from the Federal Reserve Bank of St. Louis

Summary Statistics by Survey Year

Year	Households	Weighted pop.	Entrepr. rate (%)	Mean income	Median income	Mean wealth	Median wealth	Mean age	Married (%)	Mean kids
1989	3,143	93,020,101	7.73	94,575	60,130	436,604	108,501	47.93	58.19	0.93
1992	3,906	95,917,619	9.82	82,967	55,213	387,602	102,977	48.44	57.58	0.85
1995	4,299	99,010,458	8.91	86,454	59,464	412,483	111,868	48.46	58.48	0.83
1998	4,305	102,548,842	8.86	96,719	60,986	517,171	130,733	48.73	58.52	0.82
2001	4,442	106,495,809	9.56	115,782	67,150	665,540	145,208	48.97	60.31	0.81
2004	4,519	112,108,936	9.76	111,001	67,755	705,714	146,286	49.56	57.97	0.81
2007	4,417	116,107,641	9.26	120,423	67,696	796,757	173,151	50.02	58.83	0.83
2010	6,482	117,609,217	9.70	106,954	62,457	675,727	105,166	50.53	58.07	0.84
2013	6,015	122,530,070	8.58	110,288	59,437	672,592	103,609	51.16	57.15	0.80
2016	6,248	125,981,701	9.43	126,150	64,964	850,415	119,995	51.68	56.72	0.76
2019	5,777	128,642,431	9.89	123,171	68,454	865,704	141,145	51.73	56.05	0.72
2022	4,595	131,306,389	10.44	141,390	70,259	1,059,457	192,700	52.45	57.46	0.71

Notes: Weighted estimates from the Survey of Consumer Finances, 1989-2022 survey waves. Households is the unweighted number of SCF household records; weighted population uses SCF survey weights. Income is household income and wealth is household net worth; percentages are weighted household shares. Entrepreneur denotes active business ownership with an active management role.

Figure 1

An overview of the income composition and the corresponding wealth and income Gini coefficients over the sample time period is provided in Figure 2. A clear upward trajectory in the Gini coefficients is observable, albeit fluctuations in the income Gini coefficient over time.

SCF Inequality Summary

Year	Wealth Gini	Income Gini	Top 1% wealth	Top 5% wealth	Top 10% wealth	Top 1% income	Top 5% income	Top 10% income
1989	0.790	0.540	29.9%	54.1%	67.0%	17.0%	31.6%	42.3%
1992	0.786	0.501	30.0%	54.2%	66.8%	11.7%	26.4%	37.6%
1995	0.791	0.515	34.7%	55.9%	67.7%	14.3%	28.7%	39.3%
1998	0.800	0.530	33.7%	57.0%	68.3%	16.6%	31.0%	41.2%
2001	0.805	0.562	32.2%	57.3%	69.5%	20.0%	35.3%	45.4%
2004	0.809	0.540	33.2%	57.3%	69.3%	17.0%	31.8%	42.6%
2007	0.816	0.574	33.5%	60.2%	71.2%	21.3%	37.1%	47.1%
2010	0.846	0.549	33.8%	60.4%	73.9%	17.2%	33.7%	44.4%
2013	0.850	0.574	35.3%	62.5%	74.5%	19.7%	36.2%	46.9%
2016	0.860	0.598	38.4%	64.8%	76.7%	23.4%	39.6%	49.9%
2019	0.852	0.572	37.0%	64.6%	76.0%	19.1%	35.8%	46.6%
2022	0.830	0.607	35.0%	60.8%	73.2%	22.2%	40.1%	51.0%

Notes: The table reports weighted SCF inequality measures by survey year. Wealth Gini is calculated from household net worth, and Income Gini is calculated from household income. Top shares report the weighted share of total income or wealth held cumulatively by the top 1%, top 5%, and top 10% of households in each survey year.

Figure 2

As seen by the summary statistics in Figure 3, the characteristics between entrepreneurs and non-entrepreneurs differ. Most importantly, entrepreneurs hold significantly more income and wealth, compared to non-entrepreneurs. They also have a higher probability of being married and having a higher education, as well as a more centred age distribution, with more than half of entrepreneurs being between 40-59 years old.

Summary Statistics by Entrepreneurial Status

Statistic	Non-entrepreneurs	Entrepreneurs	Difference
Household characteristics			
Weighted share of sample	90.6%	9.4%	-81.3 pp
Average income	99,126	228,976	129,850
Median income	59,887	113,234	53,347
Average wealth	521,771	2,327,554	1,805,783
Median wealth	110,053	495,204	385,151
Share married	55.7%	78.6%	22.8 pp
Education distribution			
Less than high school	15.4%	6.0%	-9.3 pp
High school diploma	30.4%	22.6%	-7.8 pp
Some college	25.3%	23.2%	-2.1 pp
College degree	29.0%	48.2%	19.2 pp
Age distribution			
Less than 20	0.5%	0.0%	-0.5 pp
20-29	13.1%	6.3%	-6.8 pp
30-39	18.9%	19.6%	0.7 pp
40-49	19.0%	25.8%	6.8 pp
50-59	16.9%	24.7%	7.8 pp
60+	31.6%	23.7%	-8.0 pp
Children distribution			
No children	58.1%	50.2%	-7.9 pp
1 child	18.2%	18.9%	0.8 pp
2 children	14.6%	19.3%	4.7 pp
3+ children	9.1%	11.5%	2.4 pp

Notes: Weighted Survey of Consumer Finances summary statistics pooled across survey years. Entrepreneurs are households reporting active business ownership with an active management role; non-entrepreneurs are all other households. Dollar values are household income or net worth, percentages are weighted household shares, and the difference column is entrepreneurs minus non-entrepreneurs.

Figure 3

Figure 4 indicates that the entrepreneurship rate differs vastly across the different income groups, being between 30 and 40 percent among the top 1 percent of income earners for most of the sample period, and consistently below 5 percent for the bottom quintile.

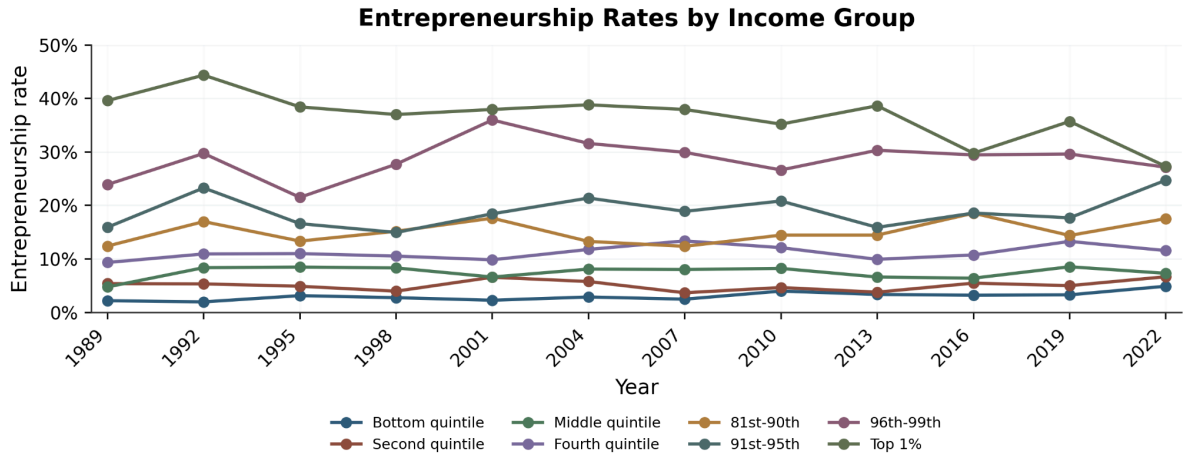


Figure 4

Additional complementary statistics are provided in Appendix B.

2.6. Model identifications

2.6.1. Pooled income-group regressions

To examine whether taxation is associated with households' entrepreneurial activity, pooled household-level regressions linking CBO income-group tax rates to entrepreneurship status is conducted. This provides a baseline test of whether the taxation-entrepreneurship relationship differ across the income distribution before moving to specifications that focus more directly on tax progressivity, policy events, and synthetic transitions.

Households are assigned to income groups within each SCF year, the tax rate corresponding to the household's income group and year is then merged onto each household observation. The model allows the tax coefficient to differ across income groups by interacting current and lagged tax rates with income-group indicators, such that:

$$Entrepreneur_{igt} = \alpha + \sum_g \beta_g Tax_{gt} 1\{G_i = g\} + \sum_g \theta_g Tax_{g,t-1} 1\{G_i = g\} + X'_{it} \gamma + \lambda_g + \delta_t + \varepsilon_{igt}$$

where $Entrepreneur_{igt}$ is the entrepreneurship indicator for household i in income group g in year t and Tax_{gt} is the CBO average federal tax rate for income group g in year t . $Tax_{g,t-1}$ is the corresponding lagged tax rate, allowing for delayed responses as entrepreneurship decisions may not adjust immediately after changes in the tax environment. It also reduces simultaneity concerns between current tax rates and current entrepreneurship

outcomes.²⁵ The vector X'_{it} includes household-level controls: age, age squared, marital status, children, and education, and γ is the corresponding vector of coefficients. The household controls are chosen to account for observable household characteristics that are associated with both their entrepreneurial decisions and tax exposure. The terms λ_g and δ_t denote income-group and year fixed effects, respectively. An alternative version of this model, used as a robustness check, additionally includes the households' inverse-hyperbolic-sine wealth:

$$Entrepreneur_{igt} = \alpha + \sum_g \beta_g Tax_{gt} 1\{G_i = g\} + \sum_g \theta_g Tax_{g,t-1} 1\{G_i = g\} + X'_{it} \gamma + \rho(Wealth_{it}) + \lambda_g + \delta_t + \varepsilon_{igt}$$

To examine whether changes in specific federal tax components are associated with changes in entrepreneurship, income-group-level first-difference regressions are estimated. First-differencing removes trends and unobserved slow-moving factors that otherwise could cause bias through spurious correlation arising when both tax rates and entrepreneurial choices tend to vary with the business cycle. As it also removes time-invariant differences across groups, these specifications do not require time fixed effects adjustment. For a single tax component c , the model is:

$$\Delta E_{gt} = \alpha + \beta \Delta Tax_{gt}^c + \delta_\tau + \varepsilon_{gt}$$

where ΔE_{gt} is the change in the weighted entrepreneurship rate for income group g between adjacent SCF waves, ΔTax_{gt}^c is the change in CBO tax component c , and δ_τ are transition fixed effects. Transition fixed effects are included to absorb common shocks specific to each adjacent SCF wave pair. Instead of controlling for the survey year, the model controls for period-wide differences in entrepreneurship rates, tax changes, macroeconomic conditions, and survey-wave composition that affect all observations within the same transition (i.e. between each SCF wave-to-wave transition). As a result, the estimated tax coefficient is identified from variation across households or groups within the same wave-to-wave transition, rather than from average differences across periods.

We also estimate specifications including all federal tax components jointly:

$$\Delta E_{gt} = \alpha + \beta_1 \Delta Tax_{gt}^{individual} + \beta_2 \Delta Tax_{gt}^{corporate} + \beta_3 \Delta Tax_{gt}^{payroll} + \beta_4 \Delta Tax_{gt}^{excise} + \delta_\tau + \varepsilon_{gt}$$

This decomposition allows assessment of whether any association between federal taxation and entrepreneurship is primarily driven by individual income taxes, corporate income taxes, payroll taxes, and / or excise taxes.

²⁵ See section 2.7

2.6.2. Tax progressivity regressions

To study tax progressivity directly, differences between high- and low-income groups are constructed, comparing the top 1 percent to the bottom 20 percent.²⁶ The entrepreneurship gap and tax gap are defined as the percentage point difference between these two income groups in entrepreneurship rates and CBO average federal tax rates, respectively:

$$GapE_t = E_{Top1,t} - E_{Bottom20,t}, \quad GapTax_t = Tax_{Top1,t} - Tax_{Bottom20,t}$$

The level specification relates the entrepreneurship gap to both the contemporaneous and one-year-lagged tax gap, such that:

$$GapE_t = \alpha + \beta_0 GapTax_t + \beta_1 GapTax_{t-1} + X_t' \gamma + \varepsilon_t$$

where X_t' denotes the additional control variables included in the extended specifications. The baseline model sets X_t' equal to zero. The following specifications then add, sequentially, a linear time trend, the log income gap, and the inverse-hyperbolic-sine wealth gap. The linear time trend controls for gradual long-run changes over the sample period that occur independently of the main relationship being estimated (i.e. irrespective of any relationship between the taxation gap and the entrepreneurship gap).

To examine whether changes in tax progressivity are associated with changes in the entrepreneurship gap, first-difference specifications are also estimated:

$$\Delta GapE_t = \alpha + \beta \Delta GapTax_t + \varepsilon_t$$

To provide insight into how the different tax components contribute, the federal tax gap is decomposed into individual income tax, corporate income tax, payroll tax, and excise tax, such that:

$$\Delta GapE_t = \alpha + \sum_c \beta_c \Delta GapTax_t^c + \varepsilon_t$$

where c indexes the federal tax component. Since the gap is constructed between the same two income groups within each transition, common transition-level shocks are partly differenced out by construction. Transition fixed effects are therefore not included in these specifications.

²⁶ Tax progressivity estimates based on a comparison between the top 1 percent and the second quintile is provided as well, in Appendix 9.A. That model specification is the same as here

2.6.3. Policy-event study

In addition to the above, we examine entrepreneurship outcomes around four major federal tax reforms, primarily targeting high-income households: Bill Clinton's 1993 Omnibus Budget Reconciliation Act, which increased both federal income taxes and corporate taxes, George W. Bush's 2003 Tax Cuts, which reduced federal income taxes but increased corporate federal taxes, Donald Trump's 2018 Tax Cuts and Jobs Act, which reduced federal corporate tax, and Barack Obama's 2012 American Taxpayer Relief Act,²⁷ which raised federal income taxes.

The event study analysis compares changes in entrepreneurship outcomes for the affected income group (the top 1 percent) with a comparison income group (the top 5 percent). Using this comparison group risks generating some spillover effects but minimizes the differences in fundamental characteristics between the groups.²⁸ Due to the triennial nature of the SCF data, the event window is relatively narrow and includes event times -1 , 0 , and $+1$, where -1 is the omitted pre-reform SCF wave, 0 is the first SCF wave post-reform, and $+1$ is the following SCF wave. The event-study model looks as follows:

$$Entrepreneur_{irt} = \alpha + \sum_{\tau \neq -1} \beta_{\tau} (Top1_i \times 1\{EventTime_{rt} = \tau\}) + X'_{it} \gamma + \lambda_g + \delta_r + \eta_{\tau} + \varepsilon_{irt}$$

where $Top1_i$ indicates membership in the top 1 percent, $1\{EventTime_{rt} = \tau\}$ indicates event time relative to reform r , X'_{it} includes household-level controls, and λ_g are income-group fixed effects. Since multiple reforms are pooled, δ_r constitute reform fixed effects, adjusting for level differences across different reform-events, and η_{τ} are event-time fixed effects, adjusting for common shocks between pre- and post reform event-time. The coefficients β_{τ} measure how entrepreneurship among top 1 percent households changes around the reform, relative to the top 5 percent.

For the reform-level model, the same specification, excluding the reform fixed effects, is estimated separately for each policy-event.

2.6.4. Matched synthetic household panel

As the SCF does not follow the same households over time, we construct synthetic household transitions by matching each household in one SCF wave to an observationally similar

²⁷ The Act was signed in 2013 but effective from 2012

²⁸ Further elaboration on the appropriateness of this control group is provided in section 3.3

household in the following wave. The matched household from the following wave is used as a synthetic next-period counterpart, not to be considered the same household observed over time. Matching is performed separately for each adjacent survey-wave pair and within each implicate.²⁹ For each baseline household, candidate matches are drawn from the next SCF wave and the household with the lowest distance score is selected.³⁰ Future-wave matches are selected with replacement, meaning that the same household from the following SCF wave can be matched to multiple baseline households. The resulting transitions are therefore synthetic and should not be interpreted as true household panel observations.

The regression analysis focuses on synthetic entry into entrepreneurship with persistence presented descriptively by income group. Let $m(i)$ denote the matched household in the next SCF wave. Synthetic entry into entrepreneurship is defined as:

$$Entry_{i,t,t+1}^{syn} = 1\{Entrepreneur_{i,t} = 0\} \times 1\{Entrepreneur_{m(i),t+1} = 1\}$$

Analogously, synthetic persistence is defined as:

$$Persistence_{i,t,t+1}^{syn} = 1\{Entrepreneur_{i,t} = 1\} \times 1\{Entrepreneur_{m(i),t+1} = 1\}$$

These transition outcomes are aggregated by baseline income group and adjacent SCF wave pair, and then linked to CBO tax changes. For each income group and transition, the tax change is defined as the change in the relevant CBO average tax rate between the two SCF waves. The resulting specification relates synthetic entry rates to income-group-level tax changes:

$$Entry_{g,t,t+1}^{syn} = \alpha + \beta \Delta Tax_{g,t,t+1} + \lambda_g + \delta_\tau + \varepsilon_{g,t,t+1}$$

where $Entry_{g,t,t+1}^{syn}$ is the synthetic entry rate among baseline non-entrepreneurs in income group g between SCF waves t and $t + 1$. The variable $\Delta Tax_{g,t,t+1}$ denotes the change in the CBO tax rate for income group g between the two waves. Income-group fixed effects, λ_g , absorb persistent differences in synthetic entry rates across income groups, while transition fixed effects, δ_τ , absorb shocks common to all income groups in a given adjacent-wave transition. In specifications decomposing the federal tax rates, the corresponding model becomes:

$$Entry_{g,t,t+1}^{syn} = \alpha + \beta_1 \Delta Tax_{g,t,t+1}^{individual} + \beta_2 \Delta Tax_{g,t,t+1}^{corporate} + \beta_3 \Delta Tax_{g,t,t+1}^{payroll} + \beta_4 \Delta Tax_{g,t,t+1}^{excise} + \lambda_g + \delta_\tau + \varepsilon_{i,t,t+1}$$

²⁹ More information on the SCF implicates and how they are accounted for in the estimations' standard errors are provided in Appendix 7.D

³⁰ A full walkthrough of how the matching algorithm operates is provided in Appendix 7.C

2.6.5. Effective tax rate cohort-by-education cells

The next analysis groups households into cohort-by-education cells. Cohorts are observed across SCF survey waves, and education groups are based on SCF's education categories.³¹ For each cell and year, we calculate weighted entrepreneurship rates, weighted average income, weighted average wealth, and cell-level TAXSIM effective tax rates. The levels specification is:

$$E_{ct} = \alpha + \beta ETR_{ct} + X'_{ct}\gamma + \mu_c + \delta_t + \varepsilon_{ct}$$

where E_{ct} is the entrepreneurship rate in cell c and year t , ETR_{ct} is the cell-level simulated tax rate: federal effective tax and total effective tax, respectively. X'_{ct} consists of control variables and includes log average income and inverse-hyperbolic-sine average wealth, μ_c are cell fixed effects, and δ_t are year fixed effects. A lagged version is also conducted:

$$E_{ct} = \alpha + \beta_0 ETR_{ct} + \beta_1 ETR_{c,t-3} + X'_{ct}\gamma + \mu_c + \delta_t + \varepsilon_{ct}$$

where $ETR_{c,t-3}$ denotes the lagged cell-level effective tax rate. Since the SCF is observed every three years, the lagged tax measure refers to the prior survey wave rather than the previous calendar year.

The first-difference specification relates changes in entrepreneurship rates within a cell to changes in cell-level effective tax rates. The specification looks as follows:

$$\Delta E_{ct} = \alpha + \beta \Delta ETR_{ct} + \Delta X'_{ct}\gamma + \delta_t + \varepsilon_{ct}$$

2.6.6. Marginal tax rate cohort-by-education cells

An analysis using TAXSIM v35 federal marginal tax rates is also conducted to capture parts of the tax system that could be more directly related to the incentives relevant for entrepreneurial decisions. The levels and first-difference marginal tax rate (MTR) specifications are analogous to the effective tax rate (ETR) specification above:

$$E_{ct} = \alpha + \beta MTR_{ct} + X'_{ct}\gamma + \mu_c + \delta_t + \varepsilon_{ct}$$

Respectively:

$$\Delta E_{ct} = \alpha + \beta \Delta MTR_{ct} + \Delta X'_{ct}\gamma + \delta_t + \varepsilon_{ct}$$

³¹ The SCF education categories consist of: (1) No high school diploma, (2) High school diploma, (3) Some college (includes associate degrees), and (4) College degree (bachelor's degree or higher)

Finally, since entrepreneurs are concentrated in high-income groups, an additional estimation is conducted using a high-income heterogeneity specification:

$$\Delta E_{ct} = \alpha + \beta \Delta MTR_{ct} + \theta (\Delta MTR_{ct} \times HighIncome_{c,t-1}) + \Delta X'_{ct} \gamma + \delta_t + \varepsilon_{ct}$$

where $HighIncome_{c,t-1}$ equals one if cohort-education cell c belongs to the top 20 percent of the previous wave's income distribution. Although this high-income measure differs from the narrower top-income group used in the CBO-based specifications, the within-cell heterogeneity indicator used here is not possible with a stricter top-tail cutoff, as it would leave too few cohort-by-education cells for meaningful estimation. In this model, β captures the association between MTR changes and changes in entrepreneurship rate for cells outside the high-income group, while θ captures the additional association for high-income cells. Thus, the total high-income effect becomes:

$$\beta^{HighIncome} = \beta + \theta$$

2.6.7. Tax progressivity and inequality regressions

To further examine the link between tax progressivity and inequality, we relate Gini coefficients to the tax gap between the top 1 percent and the bottom 20 percent. The grouped CBO Gini coefficient is used for income inequality while the SCF wealth Gini coefficient is used for wealth inequality, as described in section 2.4.2. The specifications are intended to assess whether years with larger tax progressivity are associated with higher or lower measured inequality. The baseline specification relates the Gini coefficient to both the contemporaneous and lagged tax gap:

$$Gini_t = \alpha + \beta_0 TaxGap_t + \beta_1 TaxGap_{t-1} + X'_t \gamma + \varepsilon_t$$

where $Gini_t$ denotes either the grouped CBO income Gini or the SCF wealth Gini in year t , $TaxGap_t$ is the contemporaneous federal tax gap between the top 1 percent and the bottom 20 percent, $TaxGap_{t-1}$ is the corresponding one-year-lagged tax gap, and X'_t contains macroeconomic controls where indicated. The macro controls include real GDP growth and unemployment rate,³² and are included to account for business-cycle fluctuations that may affect both tax rates and overall economic inequality. The federal tax gap is decomposed by including the federal tax components jointly, such that:

³² Both are obtained from the Federal Reserve Bank of St Louis and aggregated on an annual basis from monthly data for unemployment rate and from quarterly data for real GDP growth

$$Gini_t = \alpha + \sum_{k \in K} (\beta_{0k} TaxGap_{k,t} + \beta_{1k} TaxGap_{k,t-1}) + X_t' \gamma + \varepsilon_t$$

where k indexes the four federal tax components: individual income tax, corporate income tax, payroll tax, and excise tax.

The first-difference specification looks as follows:

$$\Delta Gini_t = \alpha + \beta \Delta TaxGap_t + \varepsilon_t$$

And the corresponding specification using the separate federal tax components is:

$$\Delta Gini_t = \alpha + \sum_{k \in K} \beta_k \Delta TaxGap_{k,t} + \varepsilon_t$$

For robustness, we add macroeconomic controls in additional specifications:

$$\Delta Gini_t = \alpha + \beta \Delta TaxGap_t + X_t' \gamma + \varepsilon_t$$

Analogously:

$$\Delta Gini_t = \alpha + \sum_{k \in K} \beta_k \Delta TaxGap_{k,t} + X_t' \gamma + \varepsilon_t$$

For the SCF wealth Gini regressions, first differences are calculated between adjacent SCF survey years, while the CBO income Gini regressions use annual changes.

Finally, we report an instrumental variable specification in Appendix C, where the SCF overall entrepreneurship rate is instrumented using the CBO top 1 percent less bottom 20 percent taxation gap. The purpose is to examine whether the entrepreneurship-inequality relationship changes when entrepreneurship is isolated using tax progressivity variation. The IV specification looks as follows:

$$Ent_t = \pi_0 + \sum_{k \in Z} \pi_k TaxGap_{k,t} + X_t' \delta + u_t$$

$$WealthGini_t = \alpha + \theta \hat{Ent}_t + X_t' \gamma + \varepsilon_t$$

2.7. Statistical remarks

Several limitations are important to highlight before proceeding to the results sections. First, the SCF is a repeated cross-section rather than a true household panel. As a result, households cannot be followed over time and household-level transitions into and out of entrepreneurship are not directly observed. Given these data challenges, the methodological approach applies multiple different specifications and model identifications and the main analysis is supported by the matched synthetic household panel and cohort-education cells. The panel should,

however, be interpreted as an approximation of transition patterns among observationally similar households, rather than as evidence on actual household-level mobility.

Second, tax measures risk being endogenous to income, household composition, and entrepreneurial behavior. Although the TAXSIM measures are useful proxies for tax incentives, they are not externally assigned tax shocks. Similarly, CBO tax rates are assigned by income group, and income-group membership may itself be affected by entrepreneurship. This creates a potential mechanical relationship between income, tax exposure, and the entrepreneurial outcome. Although lagged tax specifications and variations from changes rather than levels reduce some of the potential simultaneity issues between current entrepreneurship outcomes and current tax rates, they do not fully eliminate broader endogeneity concerns.

Third, the number of SCF survey waves is limited. This reduces the degrees of freedom and thus restricts statistical power in those model specifications relying on variation across the survey waves. The event-study analysis is further limited by the triennial frequency of the SCF, only permitting a coarse before-and-after comparison around each reform event. In addition, the design requires that entrepreneurship behavior in the pre-reform SCF wave was not already affected by anticipation of the reform, an assumption that cannot be tested directly with the available data. However, it is deemed unlikely that people would act on the tax change already up to three years before the change was in place, even if it was anticipated in advance. Furthermore, using the top 5 percent of income earners as a control group for tax reforms primarily targeting the top 1 percent risks generating spillover effects and is not a perfect specification. Therefore, rather than a pure difference-in-differences design, the event-study analysis should be considered more of a differential exposure design.

Overall, the findings should be viewed as evidence on robust descriptive and conditional associations, rather than as definitive causal estimates. The use of SCF survey weights make the descriptive statistics representative of the U.S. household population, but the institutional setting, in terms of taxation policies and financial constraints and bureaucratic hindrance to pursue entrepreneurial activity, should be deemed specific to the U.S. Thus, these results should not be mechanically generalized to other jurisdictions with significantly different business climates and (or) taxation systems.

3. Entrepreneurs and taxation

3.1. Pooled income-group regressions

As a starting point, examining the association of increased tax rates for each income group gives the results seen in Figure 5. Contemporaneous changes in taxation show no significant association with entrepreneurship rates, across all income groups, in either specification. Changes in the lagged taxation rates show some association, although only significant at the 10 percent level, for the middle quintile once wealth is added as control. The only income group with significance across the two specifications is the top 5 percent, suggesting that a one percentage point change in taxation rate in the previous year is associated with a 2.451 percentage point increase in entrepreneurship rate in the current year, once controlling for wealth. Interestingly, this is the opposite relationship than expected. However, the statistical power of these estimates should still be considered with caution, given that no estimate is significant at the 5 percent level.

Income-Group Tax Regressions with Wealth-Control Robustness

Income group	Specification	Current tax coef.	Current SE	Lagged tax coef.	Lagged SE
Bottom quintile	Baseline	-0.031	0.168	0.017	0.064
Bottom quintile	+ wealth control	-0.022	0.167	0.017	0.063
Second quintile	Baseline	0.111	0.321	0.098	0.191
Second quintile	+ wealth control	0.061	0.321	0.092	0.190
Middle quintile	Baseline	-0.350	0.393	0.386	0.266
Middle quintile	+ wealth control	-0.489	0.389	0.445*	0.265
Fourth quintile	Baseline	-0.838	0.611	0.583	0.431
Fourth quintile	+ wealth control	-0.923	0.611	0.597	0.430
81st-90th percentiles	Baseline	0.095	1.043	0.172	0.768
81st-90th percentiles	+ wealth control	0.027	1.036	0.175	0.764
91st-95th percentiles	Baseline	-0.523	1.752	0.139	1.214
91st-95th percentiles	+ wealth control	-0.521	1.761	0.148	1.222
96th-99th percentiles	Baseline	-2.626	1.817	2.409*	1.369
96th-99th percentiles	+ wealth control	-2.645	1.809	2.451*	1.367
Top 1%	Baseline	-0.405	0.925	-0.121	0.870
Top 1%	+ wealth control	-0.449	0.924	-0.071	0.868

Notes: Household-level SCF regressions of an entrepreneurship indicator on current and lagged CBO average federal tax rates assigned by weighted within-year income group and survey year. Coefficients are percentage-point changes in entrepreneurship for a one percentage-point change in the tax rate. Baseline controls include age, age squared, marital status, children, education fixed effects, year fixed effects, and income-group fixed effects; the wealth-control specification adds inverse-hyperbolic-sine household net worth. Standard errors are SCF-corrected for replicate-weight sampling uncertainty and multiple-imputation uncertainty; stars denote * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Figure 5

Given that both tax rates and entrepreneurial choices tend to vary with the business cycle,³³ we also estimate a first difference model specification, shown in Figure 6, providing a more credible comparison than simple cross-sectional relationships. The model is a pooled estimation across all income groups and breaks down the different components of the federal tax rate in the CBO dataset, as depicted in Figure B1 in Appendix B.

Despite overall changes in the federal tax rate having no significant association with overall changes in the fraction of entrepreneurs, the federal corporate tax component shows a negative effect on the rate of entrepreneurs, statistically significant at the 5 percent level. Thus, corporate tax seems to have a stronger association with changes in entrepreneurship rates than the remaining tax components. It is not surprising given Figures B1 and B2 in Appendix B, and intuitively straightforward, that the domination by the top 1 percent in both entrepreneurial activity and corporate income taxes coincide. Additionally, the excise tax

³³ See Yu et al. (2014)

component shows a significant, negative association with entrepreneurship rates. Despite the magnitude of this estimate being greater than that for the corporate income tax rate, it constitutes a very small fraction of the overall taxation faced by households,³⁴ and is insignificant at the 5 percent level. Thus, the impact of the excise tax estimate should not be overly emphasized.

First-Difference Regressions Using CBO Federal Tax Components

Tax variable / statistic	Federal tax	Individual tax	Corporate tax	All components
Change in federal tax rate	-0.460 (0.287)			
Change in individual income tax rate		-0.227 (0.336)		-0.532 (0.374)
Change in corporate income tax rate			-1.627* (0.841)	-1.629** (0.779)
Change in payroll tax rate				0.142 (0.961)
Change in excise tax rate				-5.281* (2.997)
Transition fixed effects	Yes	Yes	Yes	Yes
Income-group fixed effects	No	No	No	No
Observations	88	88	88	88
R-squared	0.241	0.225	0.267	0.299

Notes: Pooled first-difference regressions using income-group transitions between adjacent SCF survey waves. The dependent variable is the change in the weighted entrepreneurship rate for an income group; tax variables are changes in CBO average federal tax rates and components for the same income group, measured in percentage points. Specifications include transition fixed effects to absorb shocks common to each wave-to-wave transition. Coefficients are percentage-point changes in entrepreneurship for a one percentage-point tax-rate change; SCF-robust standard errors are in parentheses, with stars denoting * p<0.10, ** p<0.05, *** p<0.01.

Figure 6

3.2. Tax progressivity regressions

Previous research emphasizes the importance of tax progressivity in combating economic inequality,³⁵ making the difference in entrepreneurial tendency between the top and bottom of the income and wealth distributions especially interesting. Generally, a tax system can be defined as progressive if the after-tax income is more equally distributed than the before-tax income, and regressive if vice versa.³⁶ In this setting, a progressive tax system would imply that the top income earners pay a higher effective tax rate than the lower income earners. This effect should be the most prevalent between the very top and the very bottom of the income distribution. As seen in Figure B3 in Appendix B, an increase in tax progressivity seems to occur, with the top 1 percent's tax rate increasing from 27.9% in 1989 to 30.0% in 2019 and the bottom 20 percent's tax rate decreasing from 9.6% in 1989 to 0.6% in 2019.

³⁴ See Figure B1 in Appendix B for visualization

³⁵ See World Bank Group (2022). See also Guvenen et al. (2023) and Jalles & Karras (2024)

³⁶ See Piketty & Saez (2007)

Figure 7 shows that, in the baseline model without controls, there seems to exist an association between the taxation gap and the entrepreneurial gap between the two income groups, significant at the 10 percent level. The coefficient for the lagged tax shows that a one percentage point increase in the tax gap between the top 1 percent and bottom 20 percent (i.e. increased tax progressivity) is associated with a 0.415 percentage point decrease in their entrepreneurship gap. However, this estimate becomes insignificant once we add control variables, and the statistical power of the model is limited by the small number of observations available.

Top 1% - Bottom 20% Entrepreneurship Gap: Current and Lagged Tax-Rate Gaps

Model	Current coef.	Current SE	Current p-value	Lagged coef.	Lagged SE	Lagged p-value	Years
Baseline	-0.285	(0.487)	0.558	-0.415*	(0.237)	0.080	12
Baseline + trend	-0.248	(0.593)	0.676	-0.401	(0.298)	0.178	12
Baseline + trend + income gap	-0.264	(0.608)	0.664	-0.380	(0.301)	0.208	12
Baseline + trend + income gap + wealth gap	-0.271	(0.614)	0.659	-0.393	(0.323)	0.224	12

Notes: Regressions relate the Top 1% minus Bottom 20% entrepreneurship-rate gap to current and lagged CBO federal tax-rate gaps. The observation unit is an SCF survey year. Tax gaps are Top 1% minus Bottom 20% average federal tax rates, measured in percentage points. Controls are added sequentially as shown in the model labels. Coefficients are percentage-point changes in the entrepreneurship-rate gap for a one percentage-point tax-rate-gap change. SCF-corrected standard errors are in parentheses; stars denote * p<0.10, ** p<0.05, *** p<0.01.

Figure 7

Using a first difference model improves the model specification for the tax gap's impact on the entrepreneurship gap between the same two income groups, and includes a breakdown of the different federal taxes. Figure 8 shows that the aggregate federal tax gap has a significant impact on the entrepreneurship gap, such that a one percentage point increase in the tax gap is associated with a 0.77 percentage point decrease in the entrepreneurship gap. This effect seems to be driven by a combination of individual income tax, corporate income tax, and excise tax. Despite the effect of the excise tax having the largest magnitude, it again constitutes a very small fraction of the overall taxation gap, and its impact should not be overly emphasized. The corporate and individual income taxes, however, make up the vast majority of the overall gap.³⁷ These are associated with 1.991 and 2.141 percentage points decrease in the entrepreneurship gap given a one percentage point increase in the taxation gap, respectively, both of which are significant at the 5 percent level.

These results support the hypothesis that tax progressivity plays a role in the large differences in entrepreneurial activity seen across the income groups. As mentioned, it is not surprising that the corporate income tax presents as a key driver, given entrepreneurs'

³⁷ See Figure B1 in Appendix B

exposure to corporate income. Additionally, individual income tax is the largest federal tax bucket, as shown by Figure B1 in Appendix B, making it reasonable that some of the aggregate effect would be channeled through this component. Again, it should be acknowledged that the statistical power of the model is limited by the small number of observations available, reducing the degrees of freedom available for estimation.

First-Difference CBO Tax-Rate Gap Regressions: Top 1% minus Bottom 20%

Tax variable / statistic	Aggregate federal tax gap	Individual component gap	Corporate component gap	Payroll component gap	Excise component gap	All federal tax comp.
Δ Federal tax-rate gap	-0.770*** (0.256)					
Δ Individual income tax-rate gap		-0.465 (0.435)				-2.141*** (0.537)
Δ Corporate income tax-rate gap			-1.860* (1.108)			-1.991** (0.939)
Δ Payroll tax-rate gap				-1.898 (2.262)		1.291 (1.971)
Δ Excise tax-rate gap					-5.353 (5.675)	-19.613*** (5.122)
Transition fixed effects	No	No	No	No	No	No
Observations	11	11	11	11	11	11
R-squared	0.195	0.053	0.202	0.029	0.082	0.740

Notes: First-difference regressions of the change in the Top 1% minus Bottom 20% entrepreneurship-rate gap on changes in CBO average federal tax-rate gaps and component gaps. The observation unit is an adjacent SCF-wave transition; variables are in percentage points. Component specifications use individual income, corporate income, payroll, and excise tax gaps. SCF-corrected standard errors are in parentheses; stars denote * p<0.10, ** p<0.05, *** p<0.01.

Figure 8

As shown by figure 4 in section 2.5, the fraction of entrepreneurs in the bottom 20 percent is particularly low, making significant movement in the entrepreneurship rate minimal and hard to detect, risking imprecise statistical estimation. Additionally, some scholars choose to exclude the bottom 20 percent due to many low-income earners having zero market income,³⁸ and experience an emphasized effect of receiving tax credits, occasionally pushing their annual effective income tax rate below zero.³⁹ Therefore, an additional tax progressivity analysis comparing the top 1 percent to the second quintile is conducted in Appendix 9.A. In essence, these results mirror the findings in this section.

3.3. Policy-event study

Previous research has consistently shown that the dynamics relating taxation to inequality almost exclusively take place through changes to the federal tax systems for the top 1 percent, with relatively small changes occurring below the top percentile.⁴⁰ Considering this narrows

³⁸ See for example Piketty & Saez (2007)

³⁹ See Figure B1 in Appendix B for visualization

⁴⁰ See Piketty & Saez (2007)

down the analysis to more thoroughly assess the dynamics of the top 1 percent's entrepreneurial reaction to sudden changes in taxation. To examine this, an event study was conducted based on four main tax reforms in the past 30 years: Bill Clinton's 1993 Omnibus Budget Reconciliation Act, George W. Bush's 2003 Tax Cuts, Donald Trump's 2018 Tax Cuts and Jobs Act, and Barack Obama's 2012 American Taxpayer Relief Act. The top 5 percent of income earners are used as a control group.

Figure 9 presents the results of this study in the pooled sample for tax increases and tax cuts separately. Event time -1 is the pre-reform omitted reference period, while event time 0 denotes the first SCF wave post-reform and event time +1 denotes the following SCF wave. The model provides no support for the federal tax reforms having a significant effect on the top 1 percent households' decisions to become or remain entrepreneurs, compared to the top 5 percent households. Despite the coefficients for tax cut reforms having positive signs, consistent with the idea that reduced taxes may be associated with relatively higher entrepreneurship rates among top-income households, the estimates are too imprecise to support any conclusion. The lack of evidence remains across the two post-reform survey waves. Since the estimation is pooled, multiple effects could potentially counteract each other if several federal tax components are changed simultaneously. Similarly, effects over the full three-years event window become averaged, potentially causing noise and contributing to inconclusive estimation.

Pooled Event-Study Estimates: Top 1% vs P96-P99

Event time	Tax increases	Tax cuts
0	0.172 (5.050)	5.645 (4.788)
+1	-7.972 (5.584)	3.800 (4.436)
N	6,282	6,449

Notes: Pooled event-study estimates compare Top 1% households with P96-P99 households around federal tax reforms. Tax increases are OBRA 1993 and Obama high-income tax changes 2012; tax cuts are Bush tax cuts 2003 and TCJA 2018. The dependent variable is the entrepreneurship indicator, reported in percentage points; event time -1 is the omitted pre-reform SCF wave, event time 0 is the first post-reform wave, and event time +1 is the following wave. Estimates use SCF weights and SCF-corrected standard errors accounting for multiple imputates and replicate weights. Stars denote * p<0.10, ** p<0.05, *** p<0.01.

Figure 9

As mentioned in section 2.7, the top 5 percent does not constitute a perfect control group and spillover effects across income groups is an apparent risk. To rule out that the lack of supporting evidence is due to incorrect assumptions about a suitable control group, different income groups are compared in Figure 10. Using another control group indicates no significant difference and seems to not have any fundamental impact on the estimation, except for when using the 81st–90th percentiles. Using this control group indicates a 7.378

percentage point increase in the top 1 percent's entrepreneurship rate, compared to the 81st–90th percentiles, in the first post-reform survey wave, relative to the pre-reform survey wave, when experiencing any tax cut.

However, the estimate is only marginally significant, at the 10 percent level, and fails to remain robust for tax-increase reforms. This could be due to certain entrepreneurial effects only arising from tax cuts, but the weak statistical power rather speaks towards an imprecise estimate. Most importantly, this income group is considerably different from the top 1 percent and does not share the underlying characteristics required to deem it an appropriate control group. Although spillover effects from tax regulation could be argued to be reduced by using a comparison group further away from the top income earners, the top 5 percent remains the most relevant. It constitutes the group most closely resembling behaviors and traits of the top 1 percent, thus minimizing differences in fundamental characteristics and tax responses irrespective of income.

Robustness of Pooled Event-Study Estimates across Comparison Groups

Comparison group	Tax increases, t = 0	Tax increases, t = +1	Tax cuts, t = 0	Tax cuts, t = +1
Top 1% vs P96-P99	0.172 (5.050)	-7.972 (5.584)	5.645 (4.788)	3.800 (4.436)
Top 1% vs P91-P95	5.277 (4.943)	-1.790 (5.071)	2.503 (4.394)	-5.067 (4.233)
Top 1% vs P81-P90	1.570 (4.542)	-7.449 (4.977)	7.378* (4.233)	1.263 (4.171)
Top 1% vs P81-P99	2.323 (4.480)	-5.881 (4.886)	5.865 (4.140)	0.017 (3.881)
Top 1% vs all other households	0.733 (4.347)	-5.733 (4.687)	3.481 (4.053)	-1.510 (3.791)

Notes: Pooled event-study robustness estimates compare Top 1% households with alternative comparison groups around federal tax reforms. Tax increases are OBRA 1993 and Obama high-income tax changes 2012; tax cuts are Bush tax cuts 2003 and TCJA 2018. The dependent variable is the entrepreneurship indicator, reported in percentage points; event time -1 is the omitted reference period, event time 0 is the first post-reform SCF wave, and event time +1 is the following wave. Estimates use SCF weights and SCF-corrected standard errors accounting for multiple imputates and replicate weights. Stars denote * p<0.10, ** p<0.05, *** p<0.01.

Figure 10

Breaking down the pooled sample into each individual tax reform provides reform-level event-study estimates, as depicted in Figure 11. Again, the top 5 percent serves as control group, event time -1 is the pre-reform omitted reference period, event time 0 denotes the first SCF wave post-reform, and event time +1 denotes the following SCF wave. Clinton's 1993 OBRA and Obama's 2012 high-income tax changes show negative estimates in the following SCF wave, while Bush's 2003 Tax Cuts and Trump's 2018 TCJA show positive ones, all in line with what would be expected if increased taxation has a negative impact on high-income households' entrepreneurial decisions. However, similar to the pooled event-study, the reform-level estimates are generally imprecise with large standard errors and fail to show any consistent statistically significant patterns.

Reform-Level Event-Study Estimates: Top 1% vs P96-P99

Reform	Direction	0	+1	N
OBRA 1993	Tax increase	2.083 (7.896)	-6.192 (8.469)	4,798
Bush tax cuts 2003	Tax cut	2.974 (7.425)	4.501 (6.688)	5,402
Obama high-income tax changes 2012	Tax increase	-1.214 (7.233)	-9.192 (6.726)	5,518
TCJA 2018	Tax cut	6.692 (6.805)	1.759 (5.576)	4,221

Notes: Reform-level event-study estimates compare Top 1% households with P96-P99 households separately by federal tax reform. The dependent variable is the entrepreneurship indicator, reported in percentage points; event time -1 is the omitted pre-reform SCF wave, event time 0 is the first post-reform wave, and event time +1 is the following wave. Estimates use SCF weights and SCF-corrected standard errors accounting for multiple imputates and replicate weights. Stars denote * p<0.10, ** p<0.05, *** p<0.01.

Figure 11

3.4. Matched synthetic household panel

To complement the above analysis, a synthetic panel is constructed to address the limitation that the SCF is a repeated cross-section rather than a true household panel. By matching households to observationally similar households in the next survey wave, thus allowing for approximations of transitions, it also helps distinguish whether any changes in entrepreneurship rates are primarily related to entry into entrepreneurship or the persistence among existing entrepreneurs. Figure 12 assigns CBO tax rate changes to the matched synthetic transitions and estimates whether tax changes are associated with synthetic entry into entrepreneurship. The estimates are negative across the aggregate federal tax measure as well as all its components, except for payroll tax. However, all of them are statistically insignificant, thereby providing no support for the hypothesis that changes in tax rates are associated with changes in non-entrepreneurs' decisions to become entrepreneurs. It should be noted that the matching process used to create a synthetic panel from cross-sectional data is complex and mismatches may disproportionately affect the estimation, making it less reliable than a true panel.

Matched Synthetic-Panel Tax-Change Regressions: Entrepreneurship Entry

Tax variable / statistic	Federal tax	Individual tax	Corporate tax	All components
Δ Federal tax rate	-0.039 (0.162)			
Δ Individual income tax rate		-0.064 (0.155)		-0.269 (0.223)
Δ Corporate income tax rate			-0.300 (0.841)	-0.215 (0.855)
Δ Payroll tax rate				0.350 (0.460)
Δ Excise tax rate				-2.396 (1.762)
Income-group FE	Yes	Yes	Yes	Yes
Transition/year FE	Yes	Yes	Yes	Yes
Controls	No	No	No	No
Observations	88	88	88	88
R-squared	0.944	0.944	0.944	0.945

Notes: Regressions use matched synthetic-panel transitions between adjacent SCF waves. The dependent variable is entry into entrepreneurship among baseline non-entrepreneurs. Tax variables are CBO tax-rate changes for the household's baseline income group, measured in percentage points. Specifications include income-group and transition fixed effects. Coefficients are percentage-point changes in entry for a one percentage-point tax-rate change. SCF-corrected standard errors are in parentheses; stars denote * p<0.10, ** p<0.05, *** p<0.01.

Figure 12

The synthetic transitions into entrepreneurship and persistence among existing entrepreneurs for separate income groups are provided in Figure 13. The descriptive statistics indicate that both entry into entrepreneurship and persistence among existing entrepreneurs are strongly concentrated among high-income households. The synthetic entry rate to entrepreneurship ranges from 2.9 percent in the lowest quintile to 31.2 percent in the top 5 percent, and 38.1 percent in the top 1 percent. This pattern is consistent with the previous evidence that entrepreneurs are disproportionately located in the upper part of the income distribution.

The synthetic persistence of entrepreneurs follows a similar pattern, but the relatively low persistence rates among lower-income groups should be interpreted with caution. Although it may reflect greater instability in entrepreneurial status among lower-income households (i.e. low-income households being more likely than high-income households to terminate their entrepreneurial activity), it may also indicate that the matched synthetic panel is indeed an imperfect substitute for a true household panel. The large differences in persistence speak towards the latter and could be a mechanical reaction to the fact that entrepreneurs dominate in high-income households. Given that the matching process includes an income and wealth component, such that households are expected to stay within the same income and wealth bin across two survey waves,⁴¹ the probability is low for a low-income

⁴¹ A full walkthrough of how the matching algorithm operates is provided in Appendix 7.C

household to be matched with an entrepreneurial household (which almost exclusively exists in the high-income groups). This drives the large discrepancies in persistence rates seen in Figure 13 and further explains the inconclusive results seen in Figure 12.

Synthetic Entrepreneurship Transitions by Baseline Income Group

Baseline income group	Baseline entrepreneur share	Entry rate	Persistence rate
Lowest quintile	2.9%	2.9%	8.3%
Second quintile	5.0%	4.6%	11.7%
Middle quintile	7.5%	7.4%	13.7%
Fourth quintile	11.3%	10.6%	14.5%
P81-90	14.8%	15.0%	19.7%
P91-95	18.4%	19.1%	21.9%
P96-P99	28.9%	31.2%	36.8%
Top 1%	37.4%	38.1%	38.7%

Notes: The table reports matched synthetic-panel entrepreneurship patterns by baseline SCF income group. Baseline entrepreneur share, entry rate, and persistence rate are weighted percentages. Entry is measured among baseline non-entrepreneurs, and persistence is measured among baseline entrepreneurs. Synthetic transitions use matched donor outcomes across adjacent SCF survey waves and are descriptive rather than observed household histories.

Figure 13

3.5. Effective tax rate cohort-by-education cells

To complement earlier estimations where CBO tax data have been used as tax measure and assigned to each household based on income group, the following regressions instead implement simulated effective tax rates on cohort-education cells using TAXSIM data. This increases the number of available observations and is used as an in-between alternative to the cross-sectional SCF survey data and the matched synthetic panel. Figure 14 presents the estimates using such TAXSIM simulated effective tax rates with data available until 2019. The levels specifications (columns 1 and 2) examine whether changes in tax rates within a cohort-education cell is associated with changes in entrepreneurship rates within the same cell, while the first-difference specifications (columns 3 and 4) examine whether such tax rate changes in between adjacent SCF waves are associated with within-cell changes in entrepreneurship rates. The estimates are positive across all specifications, opposite to the expected relationship, and statistically insignificant. The insufficient statistical power makes one unable to draw any concrete conclusions and fails to provide supporting evidence that changes in simulated effective tax rates are associated with changes in entrepreneurship at the cohort-education cell level.

Here, lagged taxation is not included because the simulated tax rates are limited by the triennial observations. Any lagged effects would therefore come from an extended time

period before the entrepreneurship observation. A specification including lagged tax rates for the first two models (columns 1 and 2) is provided in Figure C3 in Appendix C for reference, but does not indicate any substantial differences to the results observed here.

Cohort-Education Estimates Using TAXSIM Effective Tax Rates

Variable / statistic	(1) Levels: Fed. ETR	(2) Levels: Total ETR	(3) FD: Fed. ETR	(4) FD: Total ETR
Effective tax rate	0.023 (0.123)	0.069 (0.088)	0.022 (0.155)	0.105 (0.152)
Income control	Yes	Yes	Yes	Yes
Wealth control	Yes	Yes	Yes	Yes
Cell fixed effects	Yes	Yes	No	No
Year fixed effects	Yes	Yes	Yes	Yes
First differences	No	No	Yes	Yes
Observations	212	212	188	188

Notes: Cohort-by-education cell regressions using SCF-based synthetic panel cells and NBER TAXSIM effective tax rates. The dependent variable is the weighted entrepreneurship share in the cell, reported in percentage points; entrepreneur denotes active business ownership with an active management role. Federal ETR is simulated federal income-tax liability divided by TAXSIM income, and total ETR adds FICA/payroll tax. Specifications report the indicated combinations of year fixed effects, cell fixed effects, log average income controls, inverse-hyperbolic-sine average wealth controls, and first differences. SCF-corrected standard errors combining replicate-weight sampling variance and multiple-imputation variance are in parentheses; stars denote * p<0.10, ** p<0.05, *** p<0.01.

Figure 14

3.6. Marginal tax rate cohort-by-education cells

Figure 15 repeats the above analysis using simulated marginal tax from the newer TAXSIM v35 with data available until 2022. Marginal tax rates provide an additional test as they are more directly related to the incentives of earning an additional dollar of income. The levels specification (column 1) indicates that a one percentage point increase in a cohort-education cell's marginal tax rate is associated with a 0.228 percentage point higher entrepreneurship rate in the same cell, significant at the 5 percent level. Similar to what is observed in section 3.5, this contrasts the expected relationship between entrepreneurial incentives by increased taxation. However, this fails to remain robust across specifications, with all remaining coefficients (columns 2, 3, and 4) indicating negative associations between marginal taxation and entrepreneurship rate.

The high-income interaction term (column 3), i.e. the top 20 percent income cohort-education cells within a survey, has a coefficient of -0.389, significant at the 10 percent level. This implies that a one percentage point increase in the marginal tax rate is associated with a 0.389 percentage point decrease in the entrepreneurship rate in high-income cohort-education cells, relative to lower-income cells, suggesting a stronger negative association with entrepreneurship amongst high-income earners than lower ones. This is not a

surprising finding, given entrepreneurs overrepresentation in the top income groups. The total high-income effect is similarly estimated to -0.387, but statistically insignificant due to slightly larger standard errors. Remaining cautious given the marginal significance level of the high-income interaction term, the analysis provides some indication of heterogeneity across income groups, but lacks sufficient statistical power to support any definite conclusion.

Robustness Using TAXSIM Marginal Tax Rates

Variable / statistic	(1) Levels: Fed. MTR	(2) FD: Fed. MTR	(3) High- income interaction	(4) High- income total effect
Federal MTR	0.228** (0.096)			
Δ Federal MTR		-0.062 (0.140)		
Δ Federal MTR x high-income cell			-0.389* (0.234)	
High-income total effect				-0.387 (0.263)
Income control	Yes	Yes	Yes	Yes
Wealth control	Yes	Yes	Yes	Yes
Cell fixed effects	Yes	No	No	No
Year fixed effects	Yes	Yes	Yes	Yes
First differences	No	Yes	Yes	Yes
Observations	235	211	211	211

Notes: Cohort-by-education cell robustness regressions using TAXSIM v35 simulated total marginal tax rates from the primary-tax-unit MTR file. The dependent variable is the weighted entrepreneurship share in the cell, reported in percentage points; high-income cells are cohort-education cells in the top quintile of the within-year average-income distribution. Specifications report the indicated combinations of year fixed effects, cell fixed effects, controls for log average income, inverse-hyperbolic-sine average wealth controls, and first differences. SCF-corrected standard errors are in parentheses; stars denote * p<0.10, ** p<0.05, *** p<0.01.

Figure 15

4. The link to inequality

4.1. Entrepreneurs and inequality

Given the summary statistics presented in section 2.5, it is evident, consistent with previous research,⁴² that the upper tail of the income distribution contains the largest share of entrepreneurs. Supporting this, Figures 16 and 17 show that the entrepreneurs are consistently more prevalent in the higher income and wealth groups. Despite the entrepreneurs constituting a considerably low share of the overall sample, they hold a disproportionate fraction of the income and wealth, as seen by Figure 18. Thus, although the empirical evidence directly linking taxation and entrepreneurship is inconsistent, the descriptive links

⁴² See section 1.2

between inequality and entrepreneurs remain strong and spur interest in entrepreneurship as an inequality-relevant outcome.

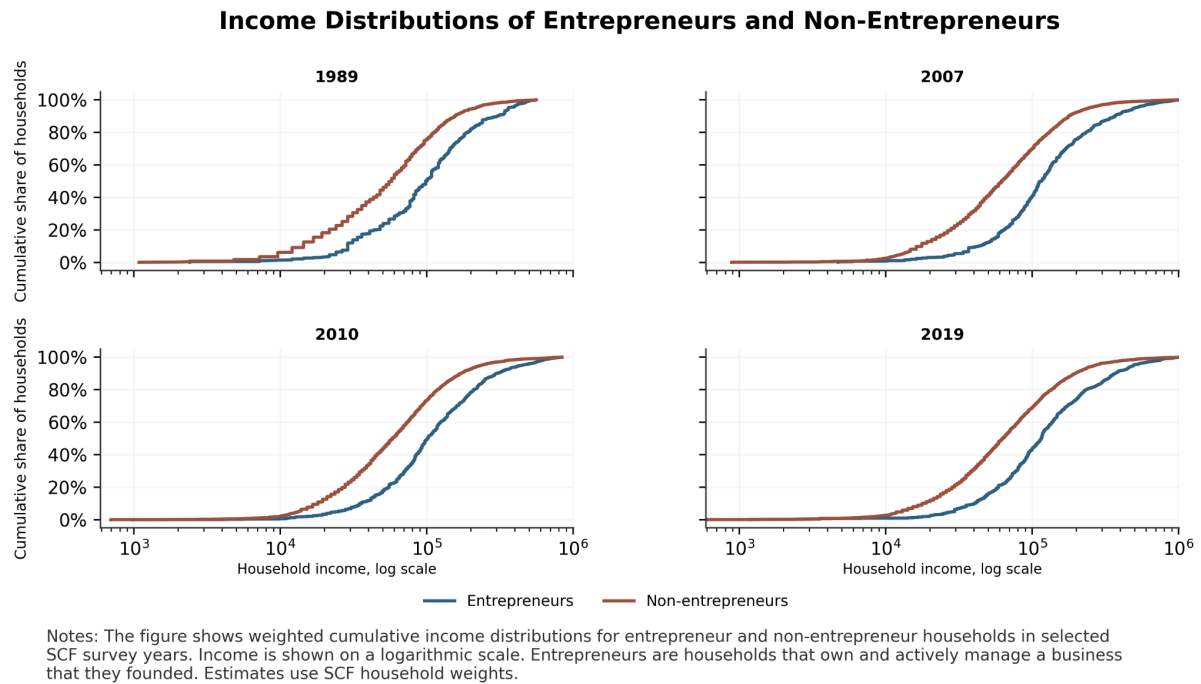


Figure 16

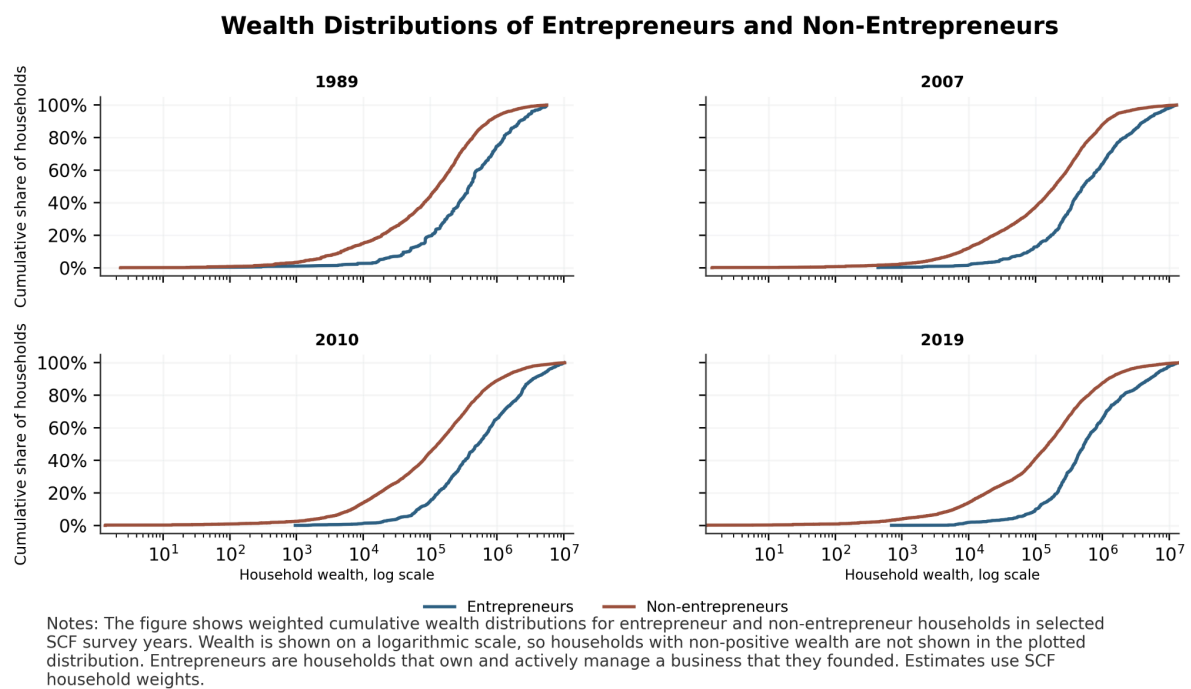


Figure 17

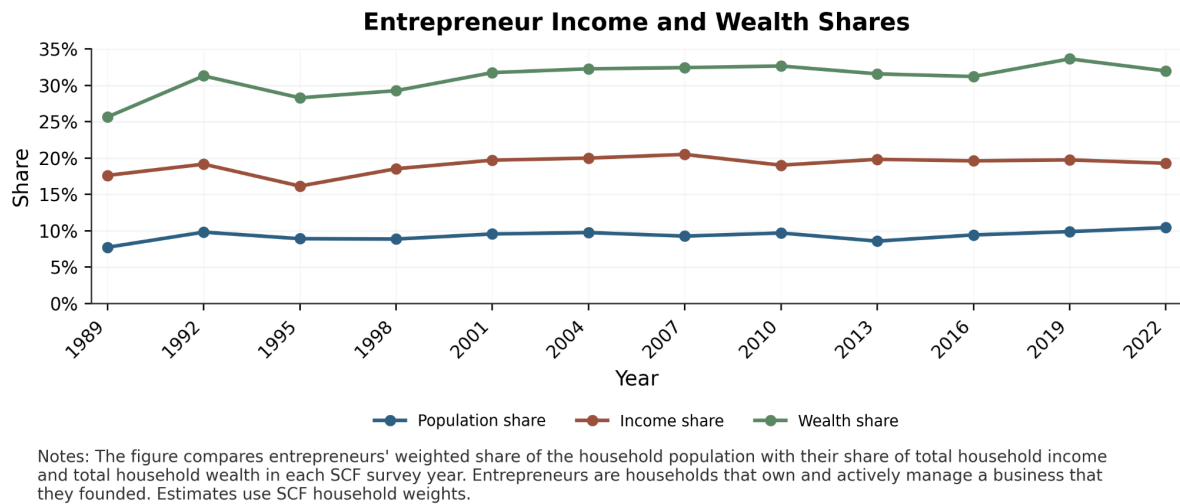


Figure 18

Understanding the dynamics and drivers of inequality during the sample period is essential to find a viable solution, or at least an improved status quo, to the widening divide in wealth accumulation and income earnings seen during this period.⁴³ Examining the development of the CBO Grouped Gini coefficient in Figure 19, a steady increase in measured economic inequality is evident. Inequality seems to systematically peak, and then fall rapidly, during crises. Visually, this is most apparent during the Dot-com Bubble in 2000, the Great Financial Crisis in 2007-2008 and most recently, Covid-19 in 2020.

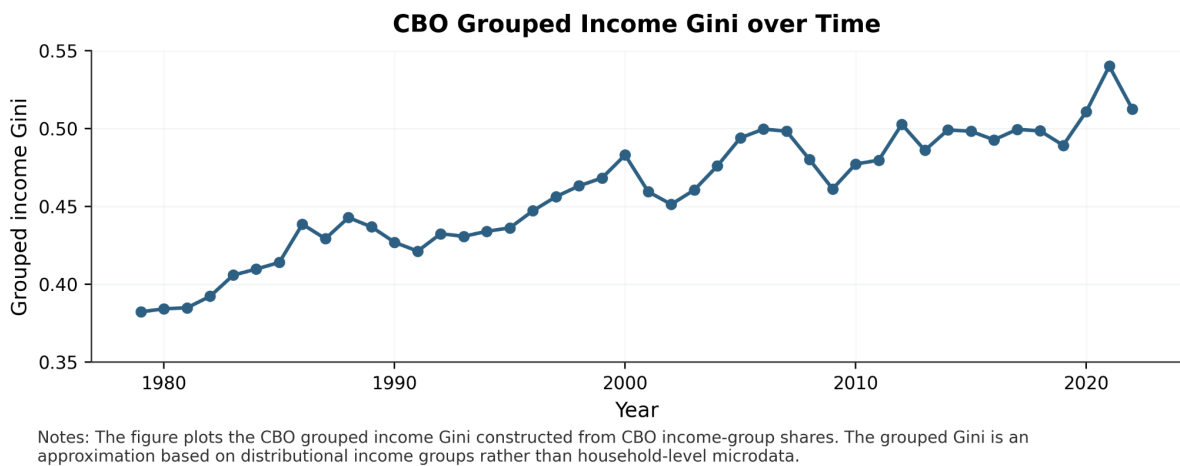


Figure 19

Figure 20 compares the SCF income Gini coefficient to the entrepreneurship rates during the same period showing no indication of systematic patterns between the levels or movements of the two measures. Comparing the SCF income Gini coefficient to only the top 1 percent's rate of entrepreneurs, depicted in Figure 21, shows even less correlation, and rather points to a negative relationship between the two. It should be noted that the SCF

⁴³ See section 2.5 and Figures 16-18

income Gini is limited in its number of observations, due to the infrequent survey waves, potentially failing to capture important movements in the measure. However, even with stronger data, a discrepancy showing a decreasing entrepreneurship rate amongst the top 1 percent simultaneously with an increasing Gini coefficient does not necessarily imply that entrepreneurs are not contributing to inequality, only that it is not through the rate of entrepreneurs specifically.⁴⁴

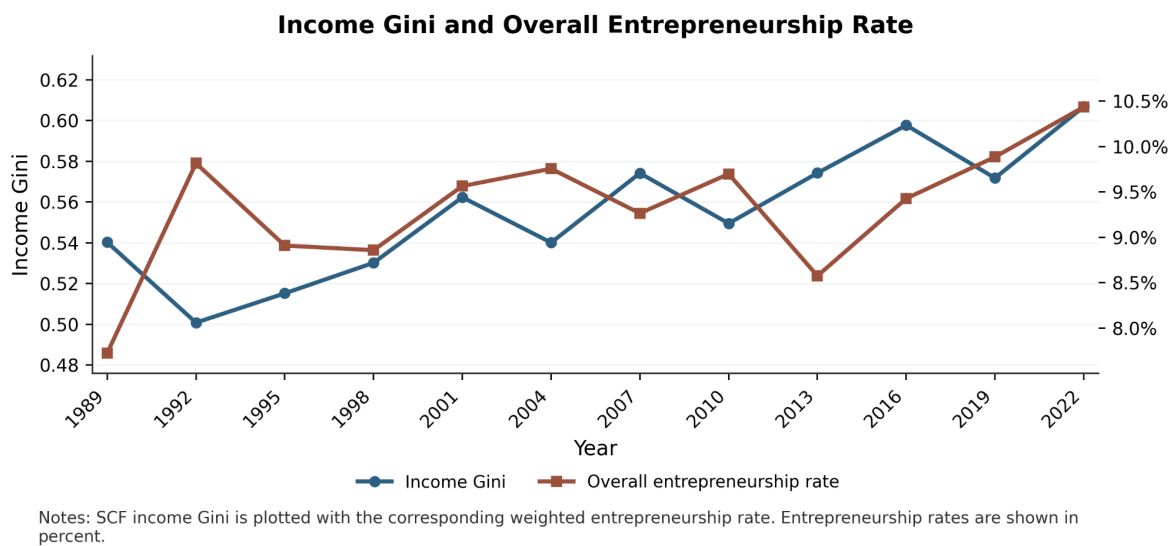


Figure 20

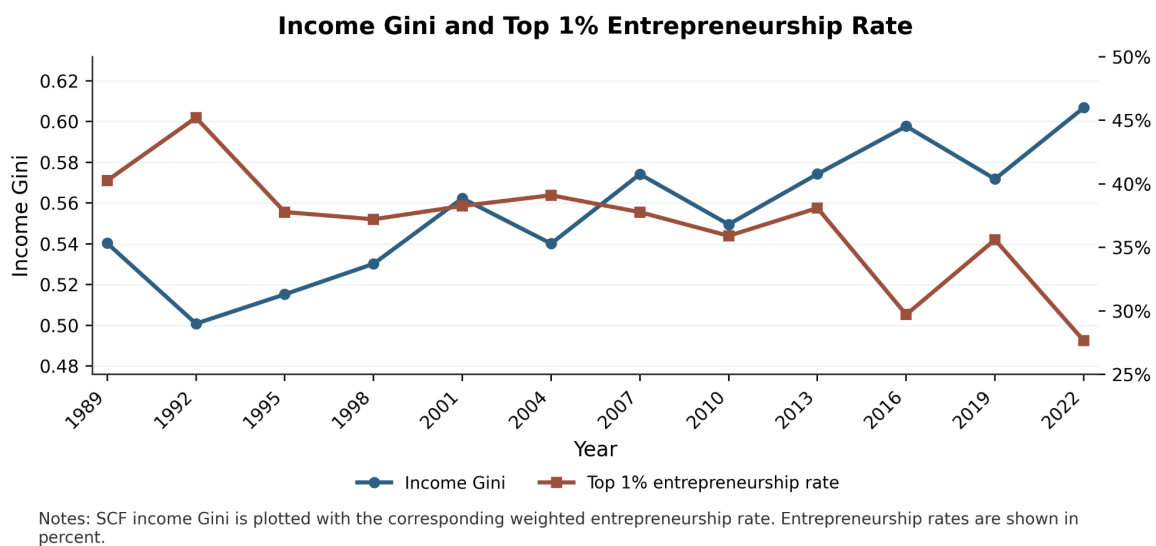


Figure 21

⁴⁴ An analysis on whether hindrance in top-tail mobility could provide further insights is provided in Appendix 9.C.

4.2. Taxation and inequality

Since the relationship between inequality and entrepreneurship is visually weak in this data, although supported through the summary statistics provided in section 2.5 and the descriptive statistics shown in Figures 16-18, linking the inequality measure to taxation grows in importance. Given the emphasis on the importance of tax progressivity to tackle growing inequality,⁴⁵ this specifically is empirically examined in Figures 22 and 23. The levels specification in Figure 22 indicates that a one percentage point increase in the current effective tax gap between the top 1 percent and the bottom 20 percent is associated with a 0.003 increase in the income Gini coefficient. The lagged effective tax gap shows a similar relationship with a coefficient of 0.002, both of which are significant at the 1 percent level. The current tax gap effect is robust once controlling for macro variables in Panel B of Figure 22, while the lagged tax gap loses some statistical power and gets a reduced estimate. Interestingly, the signs are opposite to those demonstrated by the literature, suggesting that an increased tax progressivity implies an increase in economic inequality.

When decomposing the tax rates, the current individual income tax gap component shows a similar pattern, with a coefficient of 0.002, significant at the 1 percent level, and remains robust when controlling for macro variables. Interestingly, the current payroll tax gap seems to be the other main driver of the overall tax gap effect, but suggests a negative association with a considerably higher coefficient than the individual income tax gap. Also this estimate is significant at the 5 percent level and remains robust across specifications. The two drivers thus seem to offset each other to some extent, operating in opposite directions in relation to the Gini coefficient. As mentioned before, the individual income tax constitutes the largest fraction of the total federal tax, causing this force to counteract the more intuitive effect seen from increased tax progressivity in federal payroll taxes.

⁴⁵ See World Bank Group (2022). See also Guvenen et al. (2023) and Jalles & Karras (2024) and section 3.2

CBO Tax-Rate Gap and Income Gini Levels: Current and Lagged Models

Tax variable / statistic	(1) Aggregate federal tax	(2) Individual income tax	(3) Corporate income tax	(4) All comp. jointly
Panel A: No macro controls				
Total federal tax gap (Top 1% - Bottom 20%)	0.003*** (0.001)			
Lagged total federal tax gap (Top 1% - Bottom 20%)	0.002*** (0.000)			
Individual income tax gap (Top 1% - Bottom 20%)		0.002*** (0.001)		0.002*** (<0.001)
Lagged individual income tax gap (Top 1% - Bottom 20%)		0.002*** (0.000)		0.001* (<0.001)
Corporate income tax gap (Top 1% - Bottom 20%)			0.006 (0.009)	0.003 (0.003)
Lagged corporate income tax gap (Top 1% - Bottom 20%)			-0.008 (0.008)	-0.002 (0.003)
Payroll tax gap (Top 1% - Bottom 20%)				-0.024*** (0.006)
Lagged payroll tax gap (Top 1% - Bottom 20%)				-0.006 (0.007)
Excise tax gap (Top 1% - Bottom 20%)				0.012 (0.011)
Lagged excise tax gap (Top 1% - Bottom 20%)				-0.000 (0.012)
Macro controls	No	No	No	No
Observations	43	43	43	43
R-squared	0.603	0.672	0.033	0.865
Panel B: Macro controls				
Total federal tax gap (Top 1% - Bottom 20%)	0.003*** (0.001)			
Lagged total federal tax gap (Top 1% - Bottom 20%)	0.001* (0.000)			
Individual income tax gap (Top 1% - Bottom 20%)		0.003*** (0.001)		0.002*** (<0.001)
Lagged individual income tax gap (Top 1% - Bottom 20%)		0.001*** (0.000)		0.001 (0.001)
Corporate income tax gap (Top 1% - Bottom 20%)			0.009 (0.008)	0.003 (0.003)
Lagged corporate income tax gap (Top 1% - Bottom 20%)			-0.012* (0.007)	-0.001 (0.003)
Payroll tax gap (Top 1% - Bottom 20%)				-0.020*** (0.007)
Lagged payroll tax gap (Top 1% - Bottom 20%)				-0.011 (0.008)
Excise tax gap (Top 1% - Bottom 20%)				0.015 (0.012)
Lagged excise tax gap (Top 1% - Bottom 20%)				0.003 (0.013)
Macro controls	Yes	Yes	Yes	Yes
Observations	43	43	43	43
R-squared	0.673	0.718	0.273	0.872

Notes: Annual CBO regressions relate the grouped CBO income Gini to current and lagged Top 1% minus Bottom 20% average federal tax-rate gaps and component gaps. The unit is a CBO calendar year; tax gaps are percentage-point differences between income groups for all households. Level specifications use the grouped CBO Gini, which captures between-group but not within-group inequality. Coefficients and HC1 standard errors are displayed in raw Gini units. Standard errors are reported in parentheses; stars denote * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Figure 22

The first-difference specification in Figure 23 shows that changes in the effective tax rate gap between the top 1 percent and the bottom 20 percent of income earners, again, have a significant, albeit small, positive association with changes in the income Gini coefficient.

Once controlling for macro variables in Panel B of Figure 23, this effect remains robust and strengthens in statistical power. Once again, it seems primarily driven by changes in the individual income tax component, such that a one percentage point increase in the effective individual income tax gap between the two income groups is associated with a 0.001 increase in the income Gini coefficient, statistically significant at the 5 percent level.

**Grouped CBO Income Gini as Dependent Variable: First-Difference
CBO Tax-Rate Gap Regressions**

Tax variable / statistic	(1) Aggregate federal tax	(2) Individual income tax	(3) Corporate income tax	(4) All comp. jointly
Panel A: No macro controls				
Δ Total federal tax gap (Top 1% - Bottom 20%)	0.001* (0.000)			
Δ Individual income tax gap (Top 1% - Bottom 20%)		0.001** (0.000)		0.001*** (0.000)
Δ Corporate income tax gap (Top 1% - Bottom 20%)			-0.001 (0.002)	-0.000 (0.002)
Δ Payroll tax gap (Top 1% - Bottom 20%)				-0.013** (0.006)
Δ Excise tax gap (Top 1% - Bottom 20%)				-0.006 (0.006)
Macro controls	No	No	No	No
Observations	43	43	43	43
R-squared	0.083	0.101	0.003	0.273
Panel B: Macro controls				
Δ Total federal tax gap (Top 1% - Bottom 20%)	0.001** (0.000)			
Δ Individual income tax gap (Top 1% - Bottom 20%)		0.001*** (0.000)		0.001** (0.000)
Δ Corporate income tax gap (Top 1% - Bottom 20%)			-0.004* (0.002)	-0.002 (0.002)
Δ Payroll tax gap (Top 1% - Bottom 20%)				-0.007 (0.005)
Δ Excise tax gap (Top 1% - Bottom 20%)				0.005 (0.010)
Macro controls	Yes	Yes	Yes	Yes
Observations	43	43	43	43
R-squared	0.335	0.361	0.304	0.433

Notes: Annual first-difference CBO regressions of changes in the grouped CBO income Gini on changes in Top 1% minus Bottom 20% average federal tax-rate gaps and component gaps. Tax gaps are percentage-point differences between CBO income groups for all households; component gaps use individual income, corporate income, payroll, and excise taxes. Coefficients and standard errors are displayed in Gini units. Panels without and with macro controls are reported where shown. HC1 and Newey-West standard errors are reported where shown; stars denote * p<0.10, ** p<0.05, *** p<0.01.

Figure 23

As a complement to the above analysis, we repeat the aggregate inequality regressions using the SCF wealth Gini as the dependent variable, presented in Figure 24. Similar to the effect seen on the income Gini coefficient, the aggregate federal tax gap shows a positive association with wealth inequality in the levels specification. The magnitude is slightly larger for wealth inequality, compared to income inequality, but shows in contrast no significant effect from the lagged tax gap. The association is robust to adding macro controls in Panel B

of Figure 24, although resulting in slightly larger standard errors. The individual income tax gap presents itself, once again, as the main driver amongst the different tax components. However, the estimates indicate that an increase in the lagged tax gap (higher tax progressivity) is associated with a reduced wealth Gini coefficient (improved equality), whilst the current tax gap keeps its positive sign, which is similar to its relationship with income inequality. Neither estimate remains robust once macro controls are added, indicating sensitive, and potentially biased, estimates in the first panel.

First-difference specifications and an exploratory IV specification are estimated in Appendix 9.D, neither of which provides any additional empirical evidence. In contrast to income inequality, wealth inequality shows no consistent association with tax progressivity, regardless of macro controls or lagged taxation gap, and across tax components. Considering the small number of data points for the wealth Gini coefficient, together with the observation that wealth and income Gini coefficients tend to move together,⁴⁶ suggest that the estimation is most likely affected by insufficient empirical power to examine this relationship accurately.

⁴⁶ See Figure 2

CBO Tax-Rate Gap and Wealth Gini Levels: Current and Lagged Models

Tax variable / statistic	(1) Aggregate federal tax	(2) Individual income tax	(3) Corporate income tax	(4) All comp. jointly
Panel A: No macro controls				
Total federal tax gap (Top 1% - Bottom 20%)	0.005*** (0.001)			
Lagged total federal tax gap (Top 1% - Bottom 20%)	-0.000 (0.000)			
Individual income tax gap (Top 1% - Bottom 20%)		0.006*** (0.001)		0.006*** (0.001)
Lagged individual income tax gap (Top 1% - Bottom 20%)		-0.000 (0.000)		-0.002*** (0.000)
Corporate income tax gap (Top 1% - Bottom 20%)			0.012 (0.011)	-0.002 (0.003)
Lagged corporate income tax gap (Top 1% - Bottom 20%)			-0.010* (0.005)	0.015 (0.007)
Payroll tax gap (Top 1% - Bottom 20%)				0.015 (0.013)
Lagged payroll tax gap (Top 1% - Bottom 20%)				0.006 (0.009)
Excise tax gap (Top 1% - Bottom 20%)				0.002 (0.013)
Lagged excise tax gap (Top 1% - Bottom 20%)				0.090** (0.028)
Macro controls	No	No	No	No
Observations	12	12	12	12
R-squared	0.700	0.703	0.094	0.989
Panel B: Macro controls				
Total federal tax gap (Top 1% - Bottom 20%)	0.005** (0.002)			
Lagged total federal tax gap (Top 1% - Bottom 20%)	-0.000 (0.001)			
Individual income tax gap (Top 1% - Bottom 20%)		0.005** (0.002)		0.006 (0.001)
Lagged individual income tax gap (Top 1% - Bottom 20%)		-0.000 (0.001)		-0.000 (0.002)
Corporate income tax gap (Top 1% - Bottom 20%)			0.017 (0.014)	-0.019 (0.025)
Lagged corporate income tax gap (Top 1% - Bottom 20%)			-0.013* (0.006)	0.030 (0.024)
Payroll tax gap (Top 1% - Bottom 20%)				0.033 (0.026)
Lagged payroll tax gap (Top 1% - Bottom 20%)				0.005 (0.010)
Excise tax gap (Top 1% - Bottom 20%)				0.029 (0.030)
Lagged excise tax gap (Top 1% - Bottom 20%)				0.067 (0.045)
Macro controls	Yes	Yes	Yes	Yes
Observations	12	12	12	12
R-squared	0.705	0.711	0.348	0.993

Notes: The dependent variable is the SCF wealth Gini by survey year, calculated from household net worth. Tax gaps are average tax-rate differences between the Top 1% and the Bottom 20%. Current and one-year-lagged tax gaps are matched to SCF survey years. Column (1) uses the aggregate federal tax gap. Columns (2) and (3) use the individual income-tax and corporate income-tax gaps, respectively. Column (4) includes the individual income, corporate income, payroll, and excise tax gaps jointly. Panel B adds macro controls. Coefficients and HC1 standard errors are displayed in Gini units. Standard errors are reported in parentheses. Significance stars denote * p<0.10, ** p<0.05, *** p<0.01.

Figure 24

Despite the statistical significance of these results, the magnitudes are generally small and the signs in overall tax rate gap and individual income tax gap are mostly opposite to that

argued by scholars, indicating that an increased tax progressivity implies an increase in economic inequality. It should be noted that the progressivity of the U.S. tax system has historically been primarily driven by taxation of the top earners,⁴⁷ while it during this sample period, as seen by Figure B3 in Appendix B is predominantly driven by increased tax credits to the bottom quintiles.⁴⁸ This dynamic could potentially alter the effects seen by tax progressivity in these estimates when compared to previous literature's findings. It should be acknowledged that the data used suffers from limitations and is insufficient in its ability to support a consistent conclusion. However, whether tax progressivity's role in shaping overall economic inequality has changed over time is something future research should consider.

5. Conclusions

Research has over time shown how taxation policy is key in managing rising economic inequality. Descriptive statistics from the SCF confirms that entrepreneurs are disproportionately dominant in terms of wealth accumulation and income development, consistently over time. However, across the multiple empirical specifications, the distinct relationship between taxation and entrepreneurship appears mixed and sensitive to the choice of tax measure, sample structure, and level of aggregation.

Changes in overall federal tax rates do not seem to have any impact on the average American's entrepreneurial activity, although specific tax components appear to have a significant negative association to entrepreneurship rate. Across income groups, this is predominantly driven by corporate income tax and excise tax, albeit the latter being of limited relevance given its small share of overall federal tax. Focusing on the top 1 percent, which contains the vast majority of entrepreneurs, indicates that an increased tax progressivity between this group and the bottom quintile have a positive impact on reducing their difference in entrepreneurship rate. This aligns with previous research and stresses the potential of tackling inequality through steeper tax progressivity. Interestingly, individual tax policy reforms, some providing tax reliefs and some placing additional tax burden on the top 1 percent, do not seem to show any impact on this group's entrepreneurship rate, when compared to the top 5 percent.

It should be emphasized that the data has considerable limitations. The entrepreneurial data relies on infrequent SCF survey waves, reducing the degrees of freedom and restricting statistical power in the model specifications. Additionally, the survey is cross-sectional, rather

⁴⁷ See for example Piketty & Saez (2007)

⁴⁸ See Figure B2 in Appendix B for visualization

than panel data, making us unable to follow the same households over time. To provide robustness to the above findings, we therefore construct a matched synthetic household panel, allowing for approximations of transitions into becoming and remaining an entrepreneur. However, large differences in persistence rates across income groups give rise to suspicion of whether the synthetic panel is free from considerable data limitations and provides reliable estimates.

Changing the tax measures from CBO data to TAXSIM data and implementing simulated effective tax rates on cohort-education cells do not provide any additional insights on the impact of effective tax rates. However, within-cell marginal tax rates are shown to have a positive relationship with the same cell's entrepreneurship rate, contradicting previous analysis. The same model also indicates a negative association for high-income groups, providing some support for heterogeneity of marginal tax rates' effect on entrepreneurial activity, although this estimate lacks sufficient statistical precision to support any definite conclusion.

Puzzling findings regarding inequality's link to both entrepreneurship and taxation give rise to further concerns about substantial data limitations and its inability to support a coherent interpretation. On the one hand, the analysis provides no robust evidence of a relationship between either the income or wealth Gini coefficients and entrepreneurship rates. On the other hand, an underlying relationship seems to exist between entrepreneurs and inequality, but the lack of empirically strong evidence in this thesis, much attributed to data limitations, calls for future research to access and apply more sophisticated data. Similarly, the inconclusive findings regarding the relationship between inequality and taxation make it unlikely that the empirical strength of these results is sufficient to contradict both current literature as well as overall intuition.

This analysis uses several model specifications, methodological approaches, and tax measures. Isolated statistically significant coefficients should therefore not be interpreted causally when they fail to remain robust across specifications. Further research should put effort into empirically strengthening the link between inequality, taxation, and entrepreneurs, preferably using long-term frequently-observed panel data. Researchers should build on the wealth accumulation and high-income presence of entrepreneurs, as well as the impact of tax progressivity, emphasized in previous research and confirmed in this work. Even though no empirically robust conclusion can be drawn about these effects based on this work itself, it should serve as a starting point for further examination and methodological inspiration.

6. References

- Aaron, H., J., 2015, Can taxing the rich reduce inequality? You bet it can!, Brookings, October 27
- Boar, C. and Midrigan, V., June 2023, Should We Tax Capital Income or Wealth?, *American Economic Review: Insights*, 5 (2), 259–74
- Cagetti, M. and De Nardi, M., 2006, Entrepreneurship, Frictions, and Wealth, *Journal of Political Economy*, 114 (5), 835–870
- Guvenen, F., Kambourov, G., Kuruscu, B., Ocampo, S., and Chen, D., 01 2023, Use It or Lose It: Efficiency and Redistributive Effects of Wealth Taxation*, *The Quarterly Journal of Economics*, 138 (2), 835–894
- Holmes, O., Swan, L., Kirk, A., & Henley, J., 2025, ‘The 0.001%’: a quick visual breakdown of the world’s wealthiest people, *The Guardian*, December 10
- Institute for Policy Studies, n.d., Inequality and Covid-19, *Inequality.org*
- Jalles, J. T. and Karras, G., 2024, Tax progressivity and income inequality in the US, *Economics Letters*, Volume 238, 111715
- Lee, D., and Suh, S., 2025, Measuring income and wealth inequality: A note on the Gini coefficient for samples with negative values, *Soc Indic Res* 176, 947–965
- Mamdani, Z., 2025, Zohran Mamdani says, ‘I don’t think that we should have billionaires’: Full interview, *Meet the Press*, NBC News, June 29
- Moynihan, L., 2026, Billionaires aren’t villains — they’re the engine of American innovation, *New York Post*, May 14
- Pechman, J. A., 1985, Who paid the taxes: 1966–85? (Brookings, Washington, DC)
- Piketty, T., and Saez, E., 2007, How progressive is the U.S. federal tax system? A historical and international perspective, *Journal of Economic Perspectives*, 21(1), 3–24
- Quadrini, V., 2000, Entrepreneurship, saving, and social mobility, *Review of economic dynamics*, 3 (1), 1–40
- Sanders, B., 2026, Yes, the rich must start paying their fair share of taxes, *The Guardian*, April 1
- Tamny, J., 2013, Sorry Henry Blodget, but the rich do create nearly every job, *Forbes*, December 11
- US Bureau of Economic Analysis, March 2023, National GDP & Personal Income, *Interactive Data Tables*, Section 3

U.S. Department of the Treasury, n.d., How much revenue has the U.S. government collected this year?, Fiscal Data: America's Finance Guide, <https://fiscaldata.treasury.gov/americas-finance-guide/government-revenue/>

World Bank Group, 2022, Poverty and Shared Prosperity 2022: Correcting Course, Overview booklet, World Bank, Washington, DC, License: Creative Commons Attribution CC BY 3.0 IGO

Yu, L., Orazem, P. F., Jolly, R. W., 2014, Entrepreneurship over the business cycle, Economics Letters, Volume 122, Issue 2, Pages 105-110

7. Appendix A - Model specifications

7.A. Income and wealth

Income is measured using SCF household income, while wealth is measured as household net worth. Income and wealth groups are constructed separately within each survey year using SCF household weights. This ensures that percentile groups reflect the weighted U.S. household population rather than the unweighted survey sample. For each synthetic panel cell, average income is calculated as the weighted cell mean:

$$\overline{Income}_{ct} = \frac{\sum_{i \in c,t} w_{it} Income_{it}}{\sum_{i \in c,t} w_{it}}$$

Average wealth is calculated analogously:

$$\overline{Wealth}_{ct} = \frac{\sum_{i \in c,t} w_{it} Wealth_{it}}{\sum_{i \in c,t} w_{it}}$$

In the regression analysis, average income is included in logarithmic form while average wealth is transformed using the inverse-hyperbolic-sine transformation, since wealth can be zero or negative.

7.B. TAXSIM tax rates

The cell-level effective tax rates are constructed as weighted liability-to-income ratios within each cohort-by-education cell. The federal cell-level effective tax rate is:

$$ETR_{ct}^{federal} = \frac{\sum_{i \in c,t} w_{it} FederalTaxLiability_{it}}{\sum_{i \in c,t} w_{it} TAXSIMIncome_{it}}$$

And total cell-level effective tax rate is:

$$ETR_{ct}^{total} = \frac{\sum_{i \in c,t} w_{it} (FederalTaxLiability_{it} + FICA_{it})}{\sum_{i \in c,t} w_{it} TAXSIMIncome_{it}}$$

7.C. Matching algorithm

For each baseline household i in year t , candidate future-wave households j are drawn from the next SCF wave, in $t + 3$. Candidate matches are restricted to have the same sex and race

classification as the baseline household. Among the remaining candidates, the household with the lowest distance score is selected as the future-wave match. The matching score is defined as:

$$\begin{aligned}
S_{ij} = & 2.0|Age_{j,t+3} - \min(Age_{i,t} + 3, 95)| + 1.5|Edu_{j,t+3} - Edu_{i,t}| \\
& + 1.0I(Married_{j,t+3} \neq Married_{i,t}) + 0.5\min(|Kids_{j,t+3} - Kids_{i,t}|, 4) \\
& + 1.5|IncBin_{j,t+3} - IncBin_{i,t}| + 1.5|WealthBin_{j,t+3} - WealthBin_{i,t}|
\end{aligned}$$

The score penalizes differences in expected age, education, marital status, number of children, broad weighted income group, and broad weighted wealth group. Age is matched to expected future age, so a household aged $Age_{i,t}$ in year t is compared to households close to age $\min(Age_{i,t} + 3, 95)$ in year $t + 3$. The age cap avoids requiring exact extrapolation at very high ages. The number of children enters the score with the absolute difference capped at four.

Income and wealth are not matched exactly. Households are first assigned to weighted income and wealth groups within each survey year and implicate. Income groups are then collapsed into four broad matching bins: bottom 40 percent, middle 40 percent, 80th–95th percentiles, and top 5 percent. Wealth groups are collapsed into three broad matching bins: bottom 50 percent, 50th–90th percentiles, and top 10 percent. The matching algorithm allows at most a one-bin difference in both income and wealth. This keeps matched households economically comparable without requiring exact equality in income or wealth. The candidate set is defined as the following:

$$\begin{aligned}
C_{i,t+3} = & \{j: Sex_{j,t+3} = Sex_{i,t}, Race_{j,t+3} = Race_{i,t}, \\
& |Age_{j,t+3} - \min(Age_{i,t} + 3, 95)| \leq 2, Edu_{j,t+3} \geq Edu_{i,t}, \\
& |IncBin_{j,t+3} - IncBin_{i,t}| \leq 1, |WealthBin_{j,t+3} - WealthBin_{i,t}| \leq 1, S_{ij} \leq 11.5\}
\end{aligned}$$

The algorithm requires the future-wave household to fall within the age limit, to have education at least as high as the baseline household, to be within one income and wealth bin, and to have a total match score no greater than 11.5. For baseline households aged 90 or older, the age caliper is relaxed from two to five years:

$$|Age_{j,t+3} - (Age_{i,t} + 3, 95)| \leq 5 \quad \text{if } Age_{i,t} \geq 90$$

The selected future-wave match is the candidate with the lowest score:

$$j(i, t + 3) = \arg \min S_{ij}, \text{ subject to } j \in C_{i, t+3}$$

If several candidates have the same lowest score, the candidate with the largest SCF survey weight is selected.

7.D. Standard errors correction

The SCF provides five implicates to address multiple imputation of missing values and 999 replicate weights to account for its complex survey design. Ignoring these risks understating the uncertainty in the estimates and might lead to incorrect standard errors. Thus, for regression inference, we combine replicate-weight sampling uncertainty with multiple-imputation uncertainty across the five implicates, producing SCF-corrected standard errors. Let $\hat{\beta}_m$ denote the coefficient estimate from implicate m . The final point estimate is the average across implicates:

$$\bar{\beta} = \frac{1}{M} \sum_{m=1}^M \hat{\beta}_m$$

Sampling uncertainty is estimated using the 999 SCF replicate weights on the first implicate, with each replicate weight then multiplied by the corresponding SCF multiplicity factor. The factor, assigned to each household, scales the replicate weights to reflect the number of similar households it represents in the population. Let $\hat{\beta}_r^{rep}$ denote the coefficient estimated using replicate weight r , and let $\hat{\beta}_1$ denote the full-sample estimate for the first implicate. The sampling variance becomes:

$$V_{sampling} = \frac{1}{R} \sum_{r=1}^R \left(\hat{\beta}_r^{rep} - \hat{\beta}_1 \right)^2$$

Imputation uncertainty is calculated from the dispersion of coefficient estimates across implicates:

$$V_{imputation} = \frac{1}{M-1} \sum_{m=1}^M \left(\hat{\beta}_m - \bar{\beta} \right)^2$$

The total variance combines sampling and imputation uncertainty:

$$V_{total} = V_{sampling} + \left(1 + \frac{1}{M} \right) V_{imputation}$$

And the SCF-corrected standard error becomes:

$$SE(\bar{\beta}) = \sqrt{V_{total}}$$

This approach accounts for both the SCF's complex sampling design and uncertainty from multiple imputation.

8. Appendix B - Complementary statistics

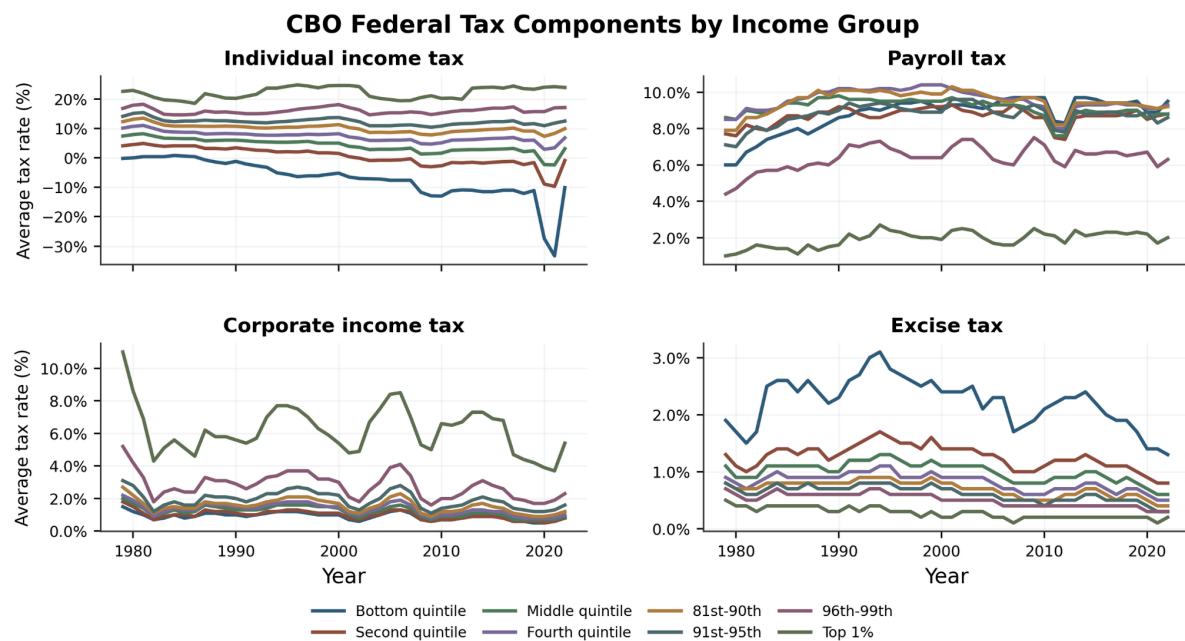


Figure B1

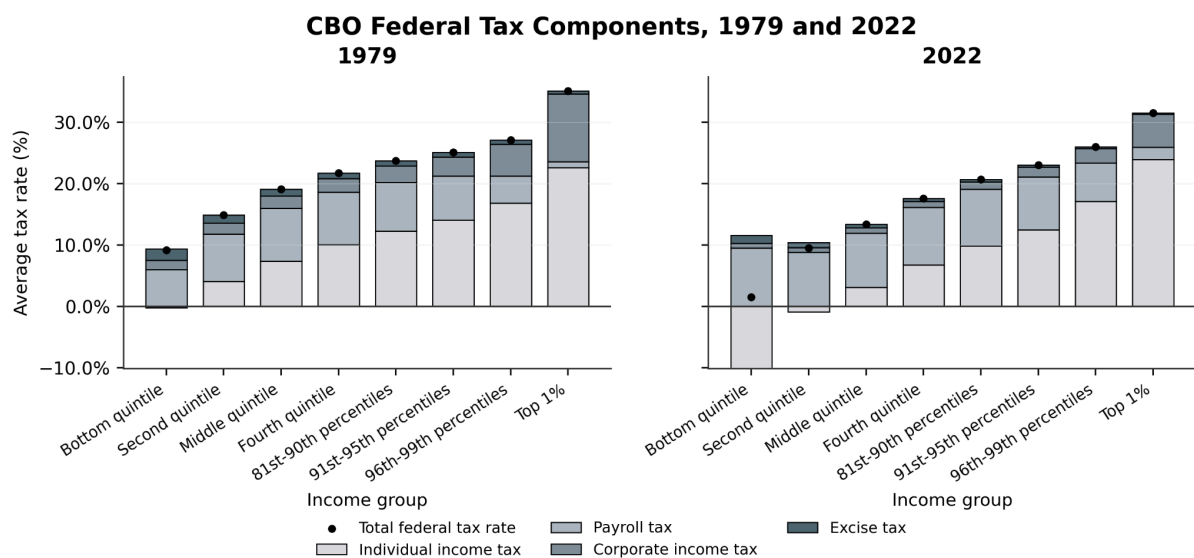


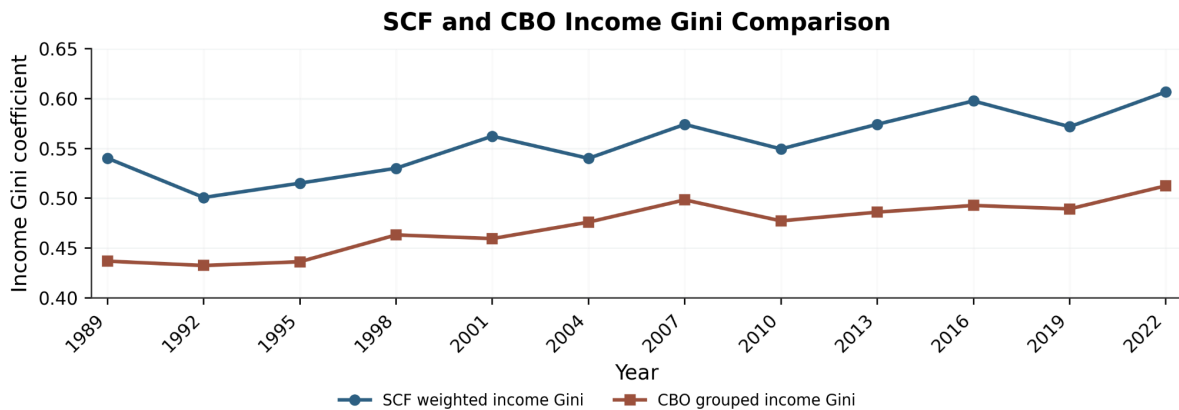
Figure B2

Tax Rates by Income Group in Selected Years

Year	Income group	CBO federal tax rate	TAXSIM total ETR	TAXSIM total MTR
1989	Bottom 20%	9.6%	5.2%	0.9%
1989	Middle 20%	18.1%	19.0%	16.6%
1989	P96-P99	25.2%	26.2%	30.3%
1989	Top 1%	27.9%	25.4%	27.8%
2004	Bottom 20%	5.0%	2.4%	-3.2%
2004	Middle 20%	14.7%	17.1%	16.7%
2004	P96-P99	25.4%	27.5%	32.0%
2004	Top 1%	29.9%	29.1%	34.2%
2019	Bottom 20%	0.6%	2.5%	-3.4%
2019	Middle 20%	13.0%	15.1%	15.5%
2019	P96-P99	24.4%	26.7%	31.3%
2019	Top 1%	30.0%	30.6%	35.3%

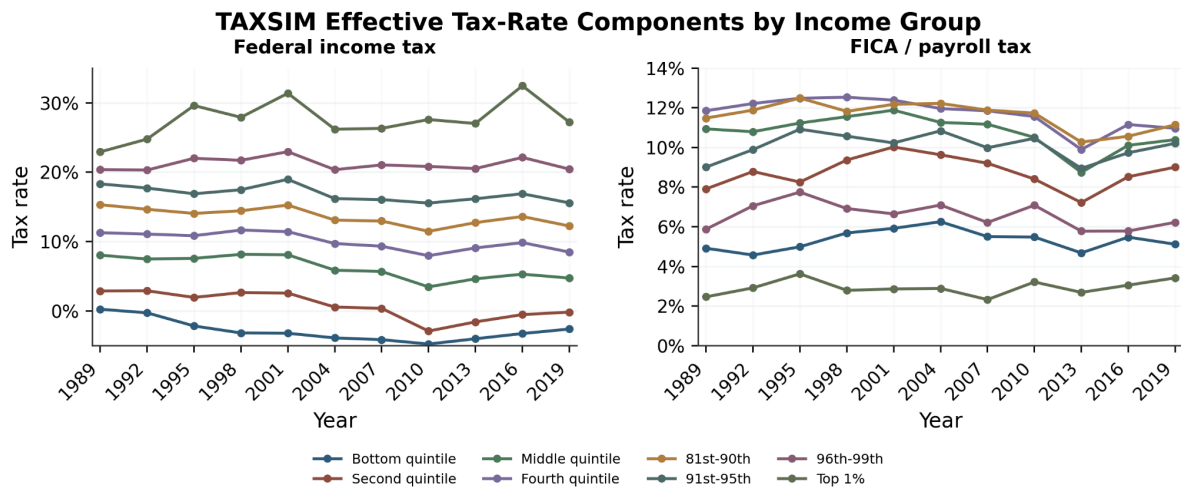
Notes: Selected-year tax-rate comparison by income group. CBO federal tax rates are average federal tax rates from CBO distributional data for all households. TAXSIM total ETR is the weighted mean of simulated household effective tax rates, constructed from federal income-tax plus FICA/payroll liabilities divided by positive TAXSIM income. TAXSIM total MTR is the weighted mean simulated total marginal tax rate from TAXSIM v35 primary tax units. Rates are percentage points; TAXSIM groups are based on weighted SCF household income ranks.

Figure B3



Notes: The figure compares the weighted SCF household income Gini with the CBO grouped income Gini in SCF survey years. The CBO measure is grouped and therefore not directly identical to the SCF microdata Gini.

Figure B4



Notes: The figure shows weighted average simulated TAXSIM effective tax rates for SCF households by income group and survey year. Federal income tax and FICA/payroll taxes are divided by TAXSIM income. Income groups are based on weighted SCF household income ranks.

Figure B5

9. Appendix C - Robustness specifications

9.A. Tax progressivity

Top 1% minus P20-P40 Entrepreneurship Gap: Current and Lagged Tax-Rate Gaps

Model	Current coef.	Current SE	Current p-value	Lagged coef.	Lagged SE	Lagged p-value	Years
Baseline	0.454	(0.768)	0.554	-1.248**	(0.611)	0.041	12
Baseline + trend	0.484	(0.787)	0.539	-1.181	(0.802)	0.141	12
Baseline + trend + income gap	0.391	(0.805)	0.627	-1.168	(0.806)	0.147	12
Baseline + trend + income gap + wealth gap	0.411	(0.803)	0.609	-1.268	(0.865)	0.142	12

Notes: Regressions relate the Top 1% minus P20-P40 entrepreneurship-rate gap to current and lagged CBO federal tax-rate gaps. The observation unit is an SCF survey year. Tax gaps are Top 1% minus P20-P40 average federal tax rates, measured in percentage points. Controls are added sequentially as shown in the model labels. Coefficients are percentage-point changes in the entrepreneurship-rate gap for a one percentage-point tax-rate-gap change. SCF-corrected standard errors are in parentheses; stars denote * p<0.10, ** p<0.05, *** p<0.01.

Figure C1

First-Difference CBO Tax Component Gap Regressions: Top 1% minus P20-P40

Tax variable / statistic	Federal tax gap	Individual income gap	Corporate gap	Payroll gap	Excise gap	All components
Δ Federal tax-rate gap	-0.592 (0.479)					
Δ Individual income tax-rate gap		-0.322 (0.629)				-2.759*** (1.066)
Δ Corporate tax-rate gap			-1.388 (1.195)			-2.184** (1.099)
Δ Payroll tax-rate gap				-1.183 (1.660)		2.409 (2.326)
Δ Excise tax-rate gap					-13.060 (10.539)	-34.446*** (10.905)
Transition fixed effects	No	No	No	No	No	No
Macro controls	No	No	No	No	No	No
Observations	11	11	11	11	11	11
R-squared	0.076	0.012	0.110	0.016	0.110	0.530

Notes: Each observation is one SCF transition. The dependent variable is the change in the Top 1% entrepreneurship share minus the change in the P20-P40 / second quintile entrepreneurship share. Tax variables are defined analogously as Top 1% minus P20-P40 first-difference gaps. The All components specification includes the individual income, corporate, payroll, and excise tax-rate gaps jointly. Transition fixed effects and macro controls are excluded. Robust standard errors are shown in parentheses. * p<0.10, ** p<0.05, *** p<0.01.

Figure C2

9.B. Lagged TAXSIM tax

Estimates Using TAXSIM Effective Tax Rates

Variable / statistic	(1) Levels: Federal ETR current + lagged	(2) Levels: Total ETR current + lagged	(3) FD: Federal ETR	(4) FD: Total ETR
Current ETR	0.005 (0.121)	0.162 (0.118)		
Lagged ETR	0.045 (0.088)	-0.085 (0.114)		
Δ ETR			0.022 (0.155)	0.105 (0.152)
Income control	Yes	Yes	Yes	Yes
Wealth control	Yes	Yes	Yes	Yes
Cell fixed effects	Yes	Yes	No	No
Year fixed effects	Yes	Yes	Yes	Yes
First differences	No	No	Yes	Yes
Observations	188	188	188	188

Notes: Cohort-by-education cell regressions using SCF-based synthetic panel cells and NBER TAXSIM effective tax rates. The dependent variable is the weighted entrepreneurship share in the cell, reported in percentage points. Columns (1) and (2) include current and lagged effective tax rates in levels; columns (3) and (4) include first differences in effective tax rates. Specifications include the controls and fixed effects shown in the table. SCF-corrected standard errors combining replicate-weight sampling variance and multiple-imputation variance are in parentheses; stars denote * p<0.10, ** p<0.05, *** p<0.01.

Figure C3

9.C. Top-tail mobility

Another angle of inequality is the mobility across income and wealth groups. The question is whether households have opportunities to move upwards or if the top wealth and income groups are occupied by the households over time. Examining whether entrepreneurship is associated with future top-tail location in the constructed matches enables assessment of whether the matched transition method preserves the strong top-tail sorting of entrepreneurs documented in the cross-sectional analysis. This should not be interpreted as true household mobility, since the SCF does not follow households over time and the matched future-wave household is therefore not the same household observed at baseline. Moreover, broad income and wealth bins are part of the matching algorithm, meaning that top-tail location among matched future-wave households is partly shaped by the matching design itself.

The results provided in Figure C4 show that those being entrepreneurs at baseline are substantially more often matched to future-wave households in the upper tails of both the income and wealth distributions. Baseline entrepreneurs are matched to future-wave households in the top 10 percent of income 25.7 percent of the time, compared to only 8.1 percent of the time for baseline non-entrepreneurs. The difference is also visible at the very top: baseline entrepreneurs are matched to top 1 percent income households 6 percent of the time, compared to 1.3 percent for non-entrepreneurs. Similar patterns are seen across the wealth distribution.

This exercise reinforces the descriptive link between entrepreneurship and inequality. Entrepreneurs are not only more concentrated at the top in the cross-sectional SCF data, but are also more strongly associated with top-tail future-wave matches in the synthetic transition dataset. However, because the results partly reflect the matching design and do not track actual households, they should be interpreted as descriptive evidence of top-tail sorting, rather than as causal evidence of entrepreneurship-driven mobility.

Matched Future Top-Tail Location by Baseline Entrepreneurship

Baseline status	N	Top 10% income	Top 5% income	Top 1% income	Top 10% wealth	Top 5% wealth	Top 1% wealth
Non-Entrepreneurs	43,389	8.1%	3.7%	1.3%	7.8%	4.8%	1.5%
Entrepreneurs	9,988	25.7%	16.1%	6.0%	27.7%	18.7%	6.5%

Notes: The table reports matched future-wave location in the top tails of the income and wealth distributions by baseline entrepreneurship status. Estimates use the preferred synthetic matching specification, are weighted within each SCF implicate using baseline household weights, and are averaged across implicates. Matched observations are not the same actual households, so the table should be interpreted as descriptive evidence of top-tail sorting in the synthetic transition data.

Figure C4

9.D. Wealth inequality

First-Difference Regressions Using CBO Top 1%-Bottom 20% Tax Progressivity Gaps

Tax variable / statistic	(1) Aggregate federal tax	(2) Individual income tax	(3) Corporate income tax	(4) All components jointly
Panel A: No macro controls				
Δ total federal tax gap (Top 1% - Bottom 20%)	0.002 (0.001)			
Δ individual income tax gap (Top 1% - Bottom 20%)		0.002 (0.002)		0.001 (0.002)
Δ corporate income tax gap (Top 1% - Bottom 20%)			-0.000 (0.002)	-0.001 (0.003)
Δ payroll tax gap (Top 1% - Bottom 20%)				0.004 (0.008)
Δ excise tax gap (Top 1% - Bottom 20%)				-0.004 (0.024)
Macro controls	No	No	No	No
Observations	11	11	11	11
R-squared	0.151	0.213	0.002	0.235
Panel B: Macro controls				
Δ total federal tax gap (Top 1% - Bottom 20%)	-0.000 (0.001)			
Δ individual income tax gap (Top 1% - Bottom 20%)		0.000 (0.001)		0.002 (0.002)
Δ corporate income tax gap (Top 1% - Bottom 20%)			-0.002 (0.002)	-0.003 (0.003)
Δ payroll tax gap (Top 1% - Bottom 20%)				-0.003 (0.009)
Δ excise tax gap (Top 1% - Bottom 20%)				0.042* (0.018)
Macro controls	Yes	Yes	Yes	Yes
Observations	11	11	11	11
R-squared	0.409	0.413	0.450	0.654

Notes: The dependent variable is the change in the SCF wealth Gini between adjacent survey years, calculated from household net worth. Tax gaps are average tax-rate differences between the Top 1% and the Bottom 20%, differenced over the same intervals. Column (1) uses the aggregate federal tax gap. Columns (2) and (3) use the individual income-tax and corporate income-tax gaps, respectively. Column (4) includes the individual income, corporate income, payroll, and excise tax gaps jointly. Panel B adds macro controls measured in the current survey year. Coefficients and HC1 standard errors are displayed in Gini units. Standard errors are reported in parentheses. Significance stars denote * p<0.10, ** p<0.05, *** p<0.01.

Figure C5

An exploratory IV specification is also estimated in which aggregate entrepreneurship is instrumented using CBO top 1 percent less bottom 20 percent tax progressivity gaps. These estimates should be viewed only as a diagnostic exercise and an exploratory extension. Although the entrepreneurship coefficient is positive in most specifications, the first-stage F-statistics are all below conventional thresholds, indicating weak instruments. In addition, the exclusion restriction is unlikely to hold strictly, since tax progressivity might affect wealth inequality directly rather than solely through entrepreneurship rates.

Exploratory IV: Wealth Gini and Entrepreneurship

Model	Instrument(s)	Entrepreneurship coefficient	Observations	First-stage F	Controls
IV: Total federal tax gap	Total federal tax gap	8.783 (12.775)	12	0.82	Macro
IV: Individual income tax gap	Individual income tax gap	7.393 (8.724)	12	1.22	Macro
IV: Corporate income tax gap	Corporate income tax gap	-1.428 (6.570)	12	0.47	Macro
IV: Payroll tax gap	Payroll tax gap	3.975 (4.030)	12	0.64	Macro
IV: Excise tax gap	Excise tax gap	4.896 (3.525)	12	1.76	Macro
IV: All component gaps jointly	All component gaps	4.377* (2.232)	12	0.58	Macro

Notes: Manual 2SLS estimates. The dependent variable is the SCF wealth Gini, calculated from household net worth, and the endogenous variable is the SCF overall entrepreneurship rate. Instruments are CBO Top 1%-Bottom 20% tax progressivity gaps. Controls are real GDP growth and unemployment. HC1-style robust standard errors are reported in parentheses. Significance stars denote * p<0.10, ** p<0.05, *** p<0.01.

Figure C6