NBER WORKING PAPER SERIES

REAL-TIME INEQUALITY

Thomas Blanchet
Emmanuel Saez
Gabriel Zucman

Working Paper 30229
http://www.nber.org/papers/w30229

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
July 2022, Revised November 2022

We thank Akcan Balkir, Anand Bharadwaj, and James Feng for outstanding research assistance, and Heather Boushey, Dennis Fixler, Marina Gindelsky, Damon Jones, Robert Kornfeld, Greg Leiserson, Danny Yagan, and numerous conference participants for helpful comments and reactions. We acknowledge financial support from the Center for Equitable Growth at UC Berkeley, the Carnegie Foundation, NSF grant SES-1559014, the Stone Foundation, and the European Research Council. This paper is supplemented by a website, https://realtimeinequality.org, with timely inequality and distributional growth estimates updated every quarter and detailed visualizations. All our data and programs are also posted online at this address. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Real-Time Inequality
Thomas Blanchet, Emmanuel Saez, and Gabriel Zucman
NBER Working Paper No. 30229
July 2022, Revised November 2022
JEL No. E01,H2,H5,J3

ABSTRACT

This paper constructs high-frequency and timely income distributions for the United States. We develop a methodology to combine the information contained in high-frequency public data sources—including monthly household and employment surveys, quarterly censuses of employment and wages, and monthly and quarterly national accounts statistics—in a unified framework. This allows us to estimate economic growth by income groups, race, and gender consistent with quarterly releases of macroeconomic growth, and to track the distributional impacts of government policies during and in the aftermath of recessions in real time. We test and successfully validate our methodology by implementing it retrospectively back to 1976. Analyzing the Covid-19 pandemic, we find that all income groups recovered their pre-crisis pretax income level within 20 months of the beginning of the recession. Although the recovery was primarily driven by jobs rather than wage growth, real wages experienced significant gains at the bottom of the distribution in 2021 and 2022, highlighting the equalizing effects of tight labor markets. After accounting for taxes and cash transfers, real disposable income for the bottom 50% was nearly 20%higher in 2021 than in 2019, but fell in 2022 as the expansion of the welfare state during the pandemic was rolled back. All estimates are available at https://realtimeinequality.org and are updated with each quarterly release of the national accounts, within a few hours.

Thomas Blanchet
University of California
Evans Hall #3880
Berkeley, CA 94720
United States
thomas.blanchet@icloud.com

Emmanuel Saez
Department of Economics
University of California, Berkeley
530 Evans Hall #3880
Berkeley, CA 94720
and NBER
saez@econ.berkeley.edu

Gabriel Zucman
Department of Economics
University of California, Berkeley
530 Evans Hall, #3880
Berkeley, CA 94720
and NBER
zucman@berkeley.edu

Real-Time Inequality website is available at https://realtimeinequality.org
Code repository is available at https://github.com/thomasblanchet/real-time-inequality

1 Introduction

A major gap in global economic statistics is the lack of timely information on the distribution of income. Thanks to a sophisticated system of national accounts and labor market statistics, detailed macroeconomic data are published almost in real time. In the United States, estimates of quarterly gross domestic product are released less than a month after the end of each quarter; monthly personal income, jobs, and unemployment statistics within a month; unemployment claims data weekly. These figures, scrutinized by the business community, are a vital input for the analysis of the business cycle and the conduct of monetary and fiscal policy. But they are not disaggregated by income level. While we know how GDP evolves quarterly, we do not know which social groups benefit from this growth, or which are most affected by economic crises as they unfold. This gap limits the ability of governments and central banks to design effective policies in crisis situation and in the aftermath of recessions. Moreover, because of this gap, macroeconomic statistics get a lot more attention than inequality in the press and the public debate, which naturally focuses on the most recent data.

Our paper attempts to address this gap by creating high-frequency and timely distributions of income for the United States, thus putting distributional statistics on an equal footing with macroeconomic statistics. We propose a methodology to combine the information contained in high-frequency public data sources, including monthly household and employment surveys, quarterly censuses of employment and wages, and monthly and quarterly national accounts series. The result of this combination is a set of harmonized monthly micro-files in which an observation is a synthetic adult (obtained by statistically matching public micro-data) and variables include income and its components. These variables add up to their respective national accounts totals and their distributions are consistent with those observed in the raw input data. Using these files, we can estimate quarterly economic growth by social group as soon as official macroeconomic growth is released. Following a recession, it becomes possible to estimate "distributional output gaps," that is the extent to which income remains below its pre-recession level or trend for the bottom 50% of the distribution, the next 40%, and the top 10%. Since our files incorporate tax and government transfer variables, they can be used to monitor how losses for different social groups during a crisis are mitigated by stabilization policies as they are implemented. Looking forward, these files could be used to estimate parameters of macroeconomic models of the business cycle with heterogeneous agents and to calibrate such models. Our files and distributional growth statistics, available at `https://realtimeinequality.org`, are updated with each release of the national accounts, within a few hours. We automated the code and website with a view to being able to provide high-frequency updates sustainably. This should allow us to analyze future business cycles in real time, maximizing the usefulness of this tool for economists, policymakers, and the public.

An impetus for this project is the Covid-19 pandemic. The pandemic dramatically affected the US economy and led to large-scale government intervention, with federal government deficits of around 15% of GDP in both 2020 and 2021, the greatest as a share of the economy since World War II. There is clear evidence that by the spring of 2022 the economy had on aggregate recovered from the shock, but inflation was at its highest since the beginning of the 1980s. Did government spend too much to support incomes at the height of the crisis? During the recovery, did all groups of the population benefit from running the economy hot, or was the recovery unequal? And in the aftermath of the recession, were real wages rising despite inflation, and if so for which socio-economic groups?

Addressing such questions in a timely manner is difficult, for two main reasons. First, income tax data—a crucial source to study income dynamics, especially at the top of the distribution—are only available with a lag of almost two years, due to late tax returns filings by top earners and processing times at the Internal Revenue Service. Second, the different data sources that contain high-frequency distributional information are incomplete and dispersed. The monthly Current Population Survey captures less than half of US national income, because it excludes certain forms of income (such as employment benefits and business profits) and does not capture high earners well (due to top-coding, small sample sizes, and income under-reporting). Employment and establishment surveys contain information on high earners, but only indirectly as they are not tabulated by income brackets. High-frequency national account series also indirectly contain a great deal of distributional information—since different forms of income are distributed differently across households—but are typically analyzed separately from survey data.

To uncover the dynamics of inequality in real time, we combine these and other publicly available sources systematically. A historical analogy is helpful to understand our objectives and the gist of our approach. Before the creation of the US national income and product accounts in the context of the Great Depression, several business surveys existed, each providing valuable information on aspects of the business cycles. But these surveys were not integrated in a consistent system, making it hard to capture the dynamics of the economy as a whole. The national accounts solved this issue and became a reference tool for the study of the business cycle, with numerous applications in macroeconomics and for policymaking. Today, high-frequency distributional information is available from various sources, but there is no unified framework in which comprehensive (i.e., capturing all sources of income) growth statistics could be computed for the different social groups. Our paper proposes a methodology to bridge this gap.

The main steps of our methodology can be summarized as follows. We start with the annual distributional national accounts micro-files of Piketty, Saez and Zucman (2018), which allocate 100% of annual national income, household wealth, and many components of these macroeconomic aggregates using primarily individual tax data. We then statistically match these files to Current Population Survey and Survey of Consumer Finance micro-data using optimal transport matching methods. To our knowledge this is the first time such a "one-to-one" statistical match is conducted. This allows us to bring gender, race, and education variables—which are missing in tax data—into the Piketty, Saez and Zucman (2018) distributional national accounts, thus allowing us to produce the first statistics on the distribution of national income by gender, race, and educational attainment.

We then move to higher frequency and estimate real-time statistics in three steps. First, taking moving averages of current and adjacent-year micro-data, or using the latest file (2019) from 2020 onwards, we create monthly files by rescaling each component of national income to its monthly seasonally-adjusted aggregate value. Second, we incorporate high-frequency changes in the wage distribution using monthly survey micro-data and tabulations of monthly and quarterly surveys and administrative records. Building on the important work of Lee (2020), we show that public tabulations of the Quarterly Census of Employment and Wages by 6-digits NAICS industry $\times$ county $\times$ ownership sector (public vs. private) can be used to predict changes in wage inequality remarkably well, including at the top of the distribution which is not well covered in existing household surveys. For example, the share of wages earned in the top 1% of industries $\times$ counties $\times$ ownership sector with the highest average wage (e.g., securities brokerage in New York county; Internet publishing and broadcasting in Santa Clara county) is strongly correlated with the share of wages earned by the top 1% workers. This allows us to project high-frequency changes in wages within the top 10% of the distribution reliably. For the bottom 90%, which is well covered by household surveys, we estimate real-time wage levels by averaging predictions from tabulated employment surveys and from the monthly Current Population Survey.

Third, we model changes in other components of pretax and posttax national income. For business and capital income, we account for changes in the aggregate value of each component (rental income, corporate profits, etc.) and assume that within-component distributions are unchanged in the short term. For government transfers, we model the distribution of new programs using program parameters, eligibility rules, and public sources, paying special attention to the new programs created during the Covid-19 pandemic such as the Paycheck Protection Program, for which detailed public data about beneficiaries and studies of incidence exist.

We test and successfully validate our methodology by applying it retrospectively back to 1976. Comparing predicted to observed income changes, we find that we correctly anticipate whether income is growing or falling 89% of the time for the top 1%, 93% for the next 9%, and 82% for the next 40% and for the bottom 50% (which often have negligible income growth over our sample period). We provide extensive quantification of the bias and noise of our projections, which for all income concepts (e.g., pretax vs. posttax) and income groups are found to be limited. Our methodology delivers accurate predictions during and in the immediate aftermath of recessions, when real-time estimates are most valuable from a policy perspective. Even tough it does no rely on tax data, our methodology is unbiased for the top 1%.

The intuition for why our methodology delivers reliable results is the following. About 30% of national income is capital income. Because wealth is a stock variable, the concentration of the various components of capital income is relatively slow-moving at high frequency. The impact of capital income on total income inequality is mostly driven by changes in the size of the different components of aggregate capital income—such as corporate profits and housing rents—over the business cycle, changes which are captured by our methodology. For labor income, which accounts for about 70% of national income, short-term changes in the distribution can be large, as unemployment spikes in recessions. But in contrast to capital income, for labor income we do not assume stable distributions within component: we capture high-frequency distributional changes thanks to our combination of household and employment surveys.

Because our methodology only uses public data, it can easily be replicated, tested, and extended. Looking forward, it could be enriched by combining administrative datasets within government agencies or by incorporating additional data sources, such as private sector information (Chetty et al., 2020). We view our paper as constructing a prototype of real-time distributions combining all currently publicly available data sources—a prototype that could be refined using additional data and eventually incorporated into official national account statistics.[1]

Using our monthly micro-files to examine the Covid-19 pandemic yields three main findings. First, all social groups recovered their real pre-crisis pretax income levels within 20 months of the start of the Covid recession. The recovery was much more equal than the recovery from

[1]The Federal Reserve has published Distributional Financial Accounts since 2019, distributing aggregate household wealth quarterly (Batty et al., 2019). For income, the Bureau of Economic Analysis (BEA) distributes annual personal income and has explored the feasibility of higher-frequency statistics (Fixler, Gindelsky, and Kornfeld, 2021). We have greatly benefited from discussions with the Federal Reserve and BEA teams.

the Great Recession of 2008–2009, during which it had taken nearly 10 years for the bottom 50% to recover its pre-crisis pretax income level—even though GDP per adult recovered in 4 years. The Covid recovery was also more equal across gender and racial groups. These findings illustrate the fact that a given trajectory of GDP growth is compatible with widely different market income dynamics for the various social groups, highlighting the usefulness of timely distributional growth statistics.

Second, labor earnings experienced significant gains at the bottom of the labor income distribution during the Covid recovery, in a context of loose monetary policy until the spring of 2022 and tight labor market. Between February 2020 (the eve of the recession) and September 2022—two months with equal employment rates—real average labor income for low-wage workers increased by more than 10%, faster than for all other groups of the population (except the top 1% which grew slightly faster). Gains were limited elsewhere. The Covid recovery was thus characterized by a reduction in wage inequality among the bottom 99%, a break from the trend prevailing since the early 1980s that highlights the equalizing effects of tight labor markets.

Third, government programs enacted during the pandemic led to an unprecedented—but short-lived—improvements in living standards for the working class. After accounting for taxes and cash and quasi-cash transfers, disposable income for adults in the bottom 50% was 20% higher in 2021 than in 2019. However, disposable income fell in the beginning of 2022 and then flatlined, as the expansion of the welfare state enacted during the pandemic—e.g., an expanded child tax credit and earned income tax credit—was rolled back. The only reason why disposable income for the bottom 50% was higher in 2022 than in 2019 (by about 10% in real terms) was the higher market income for this group, driven by wage gains.

The rest of this paper is organized as follows. In Section 2 we relate our work to the literature. Sections 3 and Sections 4 detail our methodology. Section 5 provides validation tests. In Section 6 we study the dynamics of income inequality during the Covid-19 pandemic and in its aftermath, and contrast it with the Great Recession of 2008–2009. We discuss racial inequality in Section 7 and conclude in Section 8.

2 Related Literature

2.1 Previous Attempts at Estimating Inequality at a High Frequency

There has been and there are ongoing efforts to provide timely estimates of inequality in the United States.

The Federal Reserve Bank of Atlanta maintains a monthly wage growth tracker, constructed using microdata from the Current Population Survey following a methodology developed in Daly et al. (2011).[2] The tracker reports the median percent change in the hourly wage of employed individuals observed 12 months apart. Breakdowns by, e.g., wage quartiles, gender, occupation, and census divisions are shown. Although a useful tool, this wage tracker has some limitations. First, it does not account for non-workers, hence the statistics do not map onto overall income inequality. During recessions, the median wage of employed workers in the bottom quartile often rises through composition effects as low-wage workers are laid off; even though bottom wages may appear to be growing relatively fast, inequality may in fact be rising. Second, the data are top-coded at $150,000 in annual wage, roughly the 95^{th} percentile of the wage distribution. They miss the dynamic of income in the top 5%, a group that earns about a quarter of all wages. In contrast to the Atlanta wage growth tracker, our statistics include non-workers, top earners, and all other forms of income beyond wage income (e.g., capital income and transfers), making it possible to distribute all of national income and to decompose its growth.

Since 2019, the Federal Reserve has published Distributional Financial Accounts (DFA), distributing aggregate household wealth at the quarterly frequency (Batty et al., 2019). Following Saez and Zucman (2016), the DFA allocate the official Federal Reserve Financial Accounts totals across the population. In contrast to Saez and Zucman (2016) who primarily rely on individual income tax data and the capitalization method for this allocation, the Federal Reserve uses the Survey of Consumer Finances, a triennal survey of about 6,000 families. In this paper, although our focus is primarily on income we also construct real-time estimates of wealth inequality. As detailed in Section 6.4, the evolution of wealth inequality we obtain is consistent with the DFA estimates. Our value-added is to capture the top of the distribution all the way to the top 0.01%, to provide longer time series (back to 1976, while the DFA starts in 1989), to have more distributional information at the annual frequency (due to the annual nature of tax data, as opposed to the triennal nature of the Survey of Consumer Finances), and to provide current-day estimates of wealth inequality (updated daily on `https://realtimeinequality.org`), based on daily changes in stock market indices.

Recently, Fixler, Gindelsky, and Kornfeld (2021) build on the annual distributional personal income statistics created by the Bureau of Economic Analysis (Fixler et al., 2017) to explore the feasibility of a quarterly distribution of personal income. The main methodological difference with our work is that Fixler, Gindelsky and Kornfeld (2021) do not attempt

[2]See `https://www.atlantafed.org/chcs/wage-growth-tracker`.

to project changes in distributions within components, but simply rescale the annual personal income totals component by component to match the corresponding quarterly totals. As they show (and as we confirm in Section 5.3 below), this methodology produces reasonable results in years of normal growth but significantly underestimates inequality during recessions. A key contribution of our work is to demonstrate that a more sophisticated methodology—projecting changes in the distribution of labor income using high-frequency household and employment surveys—overcomes this issue. There are a number of additional methodological differences between the two projects. In contrast to Fixler, Gindelsky and Kornfeld (2021) who distribute personal income, we distribute national income, the aggregate used to compute macroeconomic growth.[3] We start from annual estimates which are largely based on individual tax return data, while Fixler, Gindelsky and Kornfeld (2021) rely primarily on the Current Population Survey, making it harder to provide estimates within the top 10%. These differences notwithstanding, both projects share the same objective of creating timely inequality statistics consistent with the national accounts. Our work was inspired by discussions with staff of the Bureau of Economic Analysis and the ongoing dialogue between academics and researchers within government agencies is in our view highly valuable.

2.2 Impacts of the Covid-19 Pandemic on Inequality

Our work also relates to the literature on the impact of the Covid-19 pandemic on inequality, recently surveyed in Stantcheva (2022). The literature emphasizes the equalizing effects of government intervention in high-income countries, while suggesting several channels through which the pandemic may, once these interventions fade out, eventually widen economic disparities. In the US context, Parolin et al. (2022) use the monthly Current Population Survey and the Survey of Income and Program Participation to produce monthly poverty rates in real time. Several studies attempt to analyze inequality in real time in countries other than the United States. For example, Aspachs et al. (2020) and Bounie et al. (2020) use micro-level bank account data to construct high-frequency distributional data for Spain and France respectively.

Relative to this body of work, our main contribution is to provide a general methodology that can be applied to all business cycles and could be implemented throughout the world. The main feature of our methodology is its comprehensive character (capturing 100% of national income), timeliness (estimates are available online within a month), and granularity (with estimates available from the bottom 50% to the top 0.01% for pretax income, posttax income, disposable

[3]See Saez and Zucman (2020) for a discussion of the differences between these two concepts and their implications.

income, and wealth). Applied to the Covid-19 crisis in the US context, our methodology delivers new insights, such as the fast recovery of working class incomes even before government intervention, and the decline in wage inequality in the tight post-Covid labor market.

3 Rescaling to Match Monthly Income Aggregates

There are two main steps in our methodology. First, we rescale existing annual income distributions to match monthly macroeconomic income totals, income component by component. Second and most importantly, we incorporate information on changes in the distribution of income within key components, most notably wage income. In this Section we define and construct our monthly aggregates and explain how we rescale annual distributions to match these aggregates, before turning to changes in distributions within components in Section 4.

3.1 Definition of Income

Our goal is to estimate the monthly and quarterly distributions of the income concepts studied in Piketty, Saez and Zucman (2018) and in the distributional national accounts literature (Blanchet et al., 2021): factor, pretax, posttax, and disposable income.[4] Factor income is the income earned from labor and capital, before any tax and government spending and before the operation of the pension system. Pretax income is factor income after the operation of the pension system (public and private), disability insurance, and unemployment insurance. Contributions to pensions (including Social Security taxes) and to unemployment and disability insurance are removed, while the corresponding benefits are added. Pretax income thus in particular captures the effect of expanded unemployment insurance during the Covid-19 pandemic. Posttax income is pretax income minus all taxes (other than Social Security taxes, already subtracted from pretax income), plus all government transfers (other than Social Security and unemployment benefits, already included in pretax income) and the government deficit.

Factor, pretax, and posttax income all add up to national income. National income is the most comprehensive and harmonized notion of income: it includes all income that accrues to resident individuals, no matter the legal nature of the intermediaries through which this income is earned. In contrast to personal income, national income is not affected by business decisions to operate as corporations vs. non-corporate businesses such as partnerships, a decision influenced by the tax system. This feature of national income maximizes comparability over time. National

[4] Detailed definitions are presented Piketty, Saez and Zucman (2018) and Saez and Zucman (2020) in the US context, and in Alvaredo et al. (2016) and Blanchet et al. (2021) in the international context.

income is computed following internationally-agreed methods, maximizing comparability across countries. Last, it is closely related to GDP, the aggregate most often used to compute economic growth: National income is GDP minus capital depreciation plus net income received from abroad. Since capital depreciation and net foreign income account for a relatively small fraction of GDP, the growth of national income is conceptually close to the growth of GDP.[5] Our focus on national income is in line with recommendations made by the Commission on the Measurement of Economic Performance and Social Progress (Stiglitz, Sen and Fitoussi, 2009).

Factor income—the sum of income from labor and capital, the two factors of production—naturally lends itself to decompositions of economic growth. Pretax income and posttax income include income which is socialized through social insurance and the tax-and-transfer system. At the individual level, the growth of pretax and posttax income thus reflects both output growth and changes in transfers. Comparing the growth of posttax income to the growth of factor income provides a comprehensive view of the extent to which taxes and government spending equalize growth across the distribution.

We also consider a fourth income concept, disposable income. It is equal to pretax income minus all taxes, plus all cash and quasi-cash transfers. Disposable income captures the income individuals have at their disposal to consume private goods and to save. In contrast to posttax income, disposable income excludes in-kind transfers such as Medicare and Medicaid, collective consumption expenditures, and the government deficit. Disposable income does not add up to national income and thus cannot be used to decompose growth. It is, however, a useful concept to study the distributional impacts of stabilization policies during economic crises.[6]

3.2 Construction of Monthly Income Aggregates

To construct aggregate monthly factor, pretax, disposable, and posttax income and their components, we use the monthly and quarterly national accounts published by the Bureau of Economic Analysis. We start from the most detailed components of personal income (published monthly) and domestic product and income (published quarterly) available. All the monthly and quar-

[5]Conceptually, GDP and gross domestic income GDI (from which national income is derived by subtracting depreciation and adding net foreign income) are identical, but in practice they are estimated using largely independent sources in the United States and hence their growth can diverge; see Section 3.2 below.

[6]In periods of crisis, posttax income—which includes government spending other than cash transfers but adds back the government deficit—can be lower than disposable income. This was the case in the second quarter of 2020, due to the massive federal deficits induced by the economic response to the Covid pandemic. Disposable income has two advantages relative to posttax income in this context. First, it does not require one to make (necessarily debatable) assumptions about who bears the burden of the government deficit. Second, it is more directly informative of the consumption possibilities of households and of the extent to which government policies manage to smooth them over the business cycle.

terly aggregates used in this paper are seasonally-adjusted and expressed in real dollars using the national income price deflator.

Four remarks about the construction of our aggregates are in order. First, factor income is estimated using the income approach of the US national income and product accounts, not the product approach. As is well known, there is a statistical discrepancy between gross domestic income (GDI) and gross domestic product (GDP) in the United States (e.g., Fixler, de Francisco and Kanal, 2021). We do not allocate the statistical discrepancy. Our estimates match income growth, not product growth; whenever the statistical discrepancy is not zero, we do not exactly capture GDP growth.[7] Second, corporate profits and GDI are only available one month after the publication of the first estimate of GDP.[8] As a result, national income and our estimates of quarterly growth by social group are published one month after the initial estimate of GDP. Third, in addition to being published quarterly in the context of GDP statistics, most components of factor income—such as compensation of employees, proprietors' income, and rental income, but not corporate profits—as well as government transfers are also published monthly as part of personal income, about four weeks after the end of each month. This allows us to estimate certain monthly statistics within a month, such as factor and disposable income growth for the bottom 50% (where corporate profits are negligible) and the distribution of labor income. Fourth, for the components of income that are only available quarterly but not monthly—e.g., for factor income, corporate profits; for postttax income, collective government expenditure—we disaggregate the quarterly series when they become available using Denton's (1971) method, following the International Monetary Fund (2017) recommendations to compile high-frequency national accounts.

3.3 From Annual to Monthly Micro-Files

Just as monthly and quarterly national accounts data are seasonally adjusted and annualized (i.e., presented in levels equivalent to a full year), our monthly and quarterly distributional data are seasonally adjusted and annualized so that they are also directly comparable in level to

[7] Appendix Figure A1 compares the growth of GDI to the growth of GDP. Both track each other closely but not perfectly. The statistical discrepancy has been significant during the recovery from the Covid-19 recession, during which GDI has recovered faster than GDP. The other reason why we do not exactly match GDP growth is the fact that net foreign income (included in national income but not in GDI) and depreciation (included in GDI but not in national income) can grow at different rates than GDI.

[8] The first estimate of quarterly GDP is available near the end of the first month after each quarter. A second estimate is released about a month after, and a third and final estimate about a month after the second estimate, i.e. about three months after the end of the quarter. Relative to GDP, quarterly GDI is produced with a a lag of an additional month (2 months for the fourth quarter) as it requires estimating corporate profits, which is done using the detailed (but not quite as timely) Census Bureau Quarterly Financial Report (US Census, 2022).

annual inequality estimates. Seasonal adjustment is important because some forms of income such as executive bonuses are highly seasonal (e.g., most bonuses are paid once a year in January). Annualization means that we estimate the distribution of what annual income would be if seasonally-adjusted monthly income totals and their distributions remained stable over 12 months.[9] Concretely, seasonal adjustment means that a January bonus is spread out over the 12 months of the year; annualization means that if bonuses double from one January to the next, this doubling is spread out smoothly over 12 months.

Another way to measure inequality at a high frequency would be to estimate the inequality of actual monthly income. Because of income mobility (e.g., losing or starting a job in the middle of a year), this approach would lead to more inequality at the monthly frequency than at the annual frequency. By contrast, our procedure which annualizes income makes inequality statistics comparable at high vs. low frequency.[10]

Our starting point to estimate monthly distributions is the annual distributional national accounts synthetic micro-data of Piketty, Saez and Zucman (2018). These files combine IRS tax micro-data, surveys, and national accounts data to construct annual distributions of income and wealth consistent with national accounts aggregates. Since their first publication, a literature has developed to test assumptions, conduct robustness tests, develop improvements, and maximize comparability with other countries where similar methods are followed (Blanchet et al., 2021). The current files, updated in Saez and Zucman (2020b), incorporate the results of this body of work. A new file is created each year when the most recent tax statistics and annual national accounts become available. The last annual micro-file is for the year 2019, the year preceding the Covid-19 pandemic.[11] This file currently serves as our baseline for 2020 and onwards estimations.

To convert these annual files to the monthly frequency, we normalize the population and the distribution of each income component to one. We then create monthly versions of the annual files mixing samples from two adjacent years with unequal weights. Specifically, to create a file corresponding to month m in year y, we combine the micro-data for the year y with its weights multiplied by $m/12$ with the micro-data for the year $y-1$ with its weights multiplied by $1-m/12$. Therefore, each monthly file is a moving average of the yearly files over the last twelve months. This procedure smoothes out short-run, year-specific, mean-reverting variations,

[9] Quarterly income is computed as the average of monthly income over the three months of the quarter.

[10] There is no micro-data in the United States allowing one to track the longitudinal evolution of household income month after month or quarter after quarter (Fixler, Gindelsky and Kornfeld, 2021). Our approach that focuses on monthly and quarterly distributions of annualized incomes does not require longitudinal data and usefully by-passes this issue.

[11] The Covid pandemic caused a backlog at the Internal Revenue Service so that paper returns for 2020 incomes were not yet fully processed as of June 2022.

which are not informative of the distribution for a given month, and would otherwise introduce discontinuities in the monthly series. Like in the annual micro-files, each observation in the monthly micro-files represents an adult individual, defined as an individual aged 20 or more.

We then rescale the components of factor, pretax, posttax, and disposable income so that they add up to their seasonally-adjusted monthly total value, component by component and at the most granular level possible. Specifically, for factor income we rescale wages and salaries, supplements to wages and salaries, proprietor's income, rental income, corporate profits, interest income, production taxes, production subsidies, non-mortgage interest payments, and government interest payments to their respective monthly totals.[12] For pretax income we additionally rescale private pension contributions, Social Security taxes, contributions to unemployment insurance, private pension benefits, Social Security benefits, and unemployment insurance benefits; for disposable income, Medicare taxes, direct taxes, the estate tax, veteran benefits, and other cash benefits; and for posttax income, Medicare, Medicaid, other in-kind transfers, collective expenditures, and the government deficit. Components of household wealth are similarly rescaled to their end-of-month values, as detailed in Appendix B.

Because the various components of aggregate income do not grow at the same rate from one month to another, the mere act of rescaling to match monthly totals changes the distribution of income. From the second to the third quarter of 2020, for example, corporate profits grew by close to 25%, much faster than wages. Since profits are more concentrated than wages toward the top of the distribution, this pushes towards a higher top 1% pretax income share. Rescaling to match aggregates, however, is not sufficient to accurately capture high-frequency changes in inequality, especially during recessions. With publicly available data it is possible to do more, namely to project changes in distributions within key components—a task we now turn to.

4 Incorporating Changes Within Income Components

The most important step of our methodology involves incorporating information on the month-to-month evolution of the distribution of labor income, which accounts for about 70% of national income. We estimate both changes in the extensive margin (number of employed vs. non-employed individuals, including recipients of unemployment insurance benefits) and in the intensive margin (changes in the wage distribution). We estimate these changes by cells of race $\times$ education $\times$ gender $\times$ 5-year age group $\times$ marital status. To be able to do so, we need to

[12] Appendix A provides a detailed mapping of these National Income and Product Accounts concepts to the variables used in our micro-files.

bring in race, education, and age variables into the Piketty, Saez and Zucman (2018) annual distributional national accounts files. We do so by statistically matching these files to the March CPS and the Survey of Consumer of Finances. This Section begins by describing this statistical matching, before turning to how we incorporate changes in the extensive margin and intensive margin, and in the distribution of other income components (government transfers).

4.1 One-to-One Statistical Matching to Survey Data

Method. Consider two datasets: A (sometimes referred to as the base file, the annual distributional national accounts micro-files in our case) and B (sometimes referred to as the supplemental file, e.g., the March CPS). Assume A and B have common variables, denoted by X (e.g., income and its components). The remaining variables are denoted by Y in file A and Z in file B. The goal is to bring the Z variables (e.g., education) into file A. The optimal way to do so is to implement a constrained statistical match (e.g., Rodgers 1984), in which each observation from B is matched "one-to-one" with an observation in A, while minimizing the sum of distances over the X variables between matched observations. This constrained statistical match can be implemented using optimal transport methods.

Formally, assume the two datasets are of size n and m respectively, and that observations in each datasets have weights $u = (u_1, u_2, \ldots, u_n)$ and $v = (v_1, v_2, \ldots, v_m)$. Without loss of generality, assume weights sum to one. Denote by D_{ij} the distance between observation i in the first dataset and j in the second over the X variables (in our application we use the L^1 norm, i.e., the sum of the absolute values of the differences). The optimal transport map $\Gamma \in \mathbb{R}^{n\times m}$ which matches observations "one-to-one" while minimizing the sum of distances between matched observation is the solution of the following linear programming problem:

$$\min_{\Gamma\in\mathbb{R}^{n\times m}} \sum_{i=1}^{n}\sum_{j=1}^{m}\Gamma_{ij}D_{ij} \qquad \text{such that} \qquad \Gamma 1 = u \qquad \Gamma' 1 = v \qquad \Gamma \geq 0$$

The main appeal of constrained statistical matching is that the multivariate distributions of all the Z variables are preserved in the matched dataset. In that sense, there is no less of distributional information, in contrast to other matching procedures.[13] The main practical obstacle to implementing constrained statistical matches so far has been computational requirements. Finding the optimal Γ involves solving for a large-scale linear programming problem: if the two datasets to be matched each have 100,000 observations (which is the order of magnitude

[13]One-sided matching does not respect the distributions seen in the second dataset. One-sided matching without replacement is inefficient as match quality becomes very poor for the last observations matched.

in our case) then the matrix of pairwise distances has 10 billion entries. Thanks to recently developed implementations of optimal transport algorithms, solving this type of problem has become doable in a reasonable amount of time.

Implementation. To statistically match the distributional national accounts micro-files to survey data, we start by supplementing the March CPS with group quarter observations from the American Community Survey: individuals living in correctional facilities, nursing homes, college dormitories, etc., who are not sampled by the March CPS. This allows us to capture the entire population of US residents, as in our distributional micro-files, which is important to capture income dynamics at the bottom of the distribution. We then statistically match the augmented March CPS to our annual micro-files over the following X variables (observed in both datasets): wage income, pension income, business income, interest, dividend and rents, Social Security benefits, welfare benefits, and government transfers other than Social Security and welfare benefits. We bring in the following Z variables: race, education, age. The match is done at the household level, separately for married vs. singles, individuals aged more vs. less than 65, and employed vs. unemployed individuals.

We similarly match our micro-files to the Survey of Consumer Finances (SCF), a triennal survey of about 6,500 families that over-samples wealthy households. We annualize the SCF by taking moving averages of the two closest waves of the SCF and match it to our micro-files over wage income, pension and Social Security income, business income, interest and dividend income, capital gains, financial and business assets, housing assets, and debts. This allows us to bring socio-economic characteristics for high-income households, which are not well covered in the March CPS. We use the socio-economic variables transported from the SCF for households in the top 5% of the income or wealth distribution, and those transported from the March CPS for the bottom 95%.

Value and limitation. The key value of our one-to-one statistical match is to join together several micro-databases (including individual tax data, the CPS, and the SCF) in a single file. Variables that are common across databases exist in versions corresponding to each dataset (e.g., SCF net wealth and net wealth estimated from tax data) and are close to each other record by record thanks to the matching procedure. This makes it straightforward to switch from one database to another as needed. For example, researchers who are used to working with the CPS can primarily focus on CPS variables and replace them with tax-data variables whenever the analysis concerns the top of the distribution.

However, we emphasize that this matched database cannot provide reliable information about the joint distribution of variables that are not jointly included in at least one of the original database. For example, because neither the public-use individual tax data, nor the CPS or the SCF provides comprehensive information on the joint distribution of income and state,[14] we cannot analyze patterns in income growth by state $\times$ social group. As another example, because the only information on the joint distribution of wealth and race comes from the SCF and the SCF is noisy above the top 1% (due to small sample sizes), it is not possible to use the matched dataset to obtain reliable estimates of the racial composition of wealth above the 99^{th} percentile. The limitations of the source files carry over to the matched file and knowledge of these limitations is critical to make an informed used of the matched database. An obviously superior way to construct a unified database would be to proceed to exact matches across administrative sources and surveys, as can be done for research purposes in, e.g., Scandinavian countries. However, this is not yet fully feasible in the United States even within government agencies (see, e.g., Chetty et al., 2010) and certainly not using public data. Furthermore, any such match could not be produced quickly and thus would not help with the production real-time inequality estimates. The statistical match we implement leverages the strength of US public data, which are rich but scattered.

Test of the matching procedure. We test and validate the quality of our matching procedure by comparing the demographic composition of income and wealth by income and wealth group in the survey micro-data and the matched file. To understand the logic of the test, consider the case of gender and wage earnings. The CPS provides good information on gender composition by annual wage earnings deciles (i.e., the fraction of individuals in each decile of the earnings distribution who are women). We can check that this gender composition is preserved when using the annual wage earnings variable coming from the tax data (as opposed to the CPS) in the matched file. This will naturally be the case if the matched observations have similar earnings in the CPS and the tax data.[15]

Figure 1 reports the results of this test. The figure compares demographic composition along one key income (or wealth) variable of interest, as measured in the original survey data vs. in the matched file tax data. Panels (a) and (b) consider gender and racial composition by annual wage earnings measured in the CPS vs. tax data; in both cases, the sample includes

[14]The state variable is not available for top earners in the public-use tax files.

[15]By contrast, this test could fail if the statistical match is not excellent on the wage earnings variable, e.g., because the samples matched are too small to find close matches, or because the matching procedure puts high weight on other variables.

the full population of working-age individuals (aged 20–64), including individuals with no wage earnings. Panel (c) considers racial composition by annual total income in the CPS, the SCF, and the matched file. For the purpose of this exercise, total income is the sum of wages, pensions, Social Security benefits, business income, interest, dividends, and rents. Last, panel (d) considers racial composition by wealth in the SCF vs. the matched file. In both panels (c) and (d), we consider the entire CPS (and SCF) samples. All four panels show that demographic compositions are very close in the original survey data and the matched file. The reason for this success is that our one-to-one matching procedure preserves ranks in the income and wealth distribution well, as illustrated in Appendix Figure A2 in the case of wage earnings.

4.2 Changes in Employment

To capture high-frequency changes in the distribution of labor income, the first step of our methodology involves adjusting employment status at the micro-level each month. To do so we compute employment rates by race $\times$ education $\times$ gender $\times$ 5-year age group $\times$ marital status cells in the monthly CPS, and use these tabulations to impute employment rates by cells in our monthly distributional national accounts micro-files. We proceed in four steps.

Estimation of aggregate employment rate First, we estimate the number of workers each month using the BLS monthly release of non-farm employment at the national level. Since the number of employed people in a given year is mechanically higher than in a given month, we adjust monthly employment numbers to make them commensurable with yearly estimates from Social Security tax data. We do so using the strong and consistent linear relationship observed between the BLS and the Social Security numbers over time. Appendix Figure A3 depicts the raw monthly employment rates from the BLS, the annual employment rates from the Social Security tax data, and the monthly employment rates adjusted to match annual rates (in all cases for the working-age population, age 20–64). In recent decades, raw monthly employment rates are around 75% while adjusted monthly employment rates (that track annual levels) are 10 percentage points higher, around 85% (because of part-year workers).

Estimation of the aggregate number of UI recipients Second, we estimate the number of unemployment insurance (UI) recipients each month. To do so, we use the Department of Labor's weekly publication of unemployment claims. We aggregate this data by month and adjust it for seasonal variations using the X11 procedure (Shiskin et al., 1967). Since the number of UI recipients in a given week is mechanically lower than in a given year, we adjust the number of UI claims by a constant coefficient to match the annual levels recorded in the tax data (consistent with our goal of constructing monthly distributions of annualized income).

Estimation of labor force status by individual characteristic Third, using the monthly CPS, we estimate monthly series of labor force status by race × education × gender × 5-year age group × marital status. In each monthly CPS dataset, we run a logistic regression of (i) employment status and (ii) unemployment status against race, education, age by 5-year group, and marital status interacted with gender. We use the prediction from these regressions to estimate employment and unemployment rates by cell. We adjust this data for seasonal variations using the X11 procedure (Shiskin et al., 1967).

Adjustment of employment and UI recipients in the monthly microfiles. Last, in our monthly microfiles, we adjust at the margin whether someone (i) is employed and (ii) receives UI benefits based on the information collected in the three preceding steps. The procedure, described in detail in Appendix C.1, reproduces relative changes in labor force status by race × education × gender × 5-year age group × marital status, while also matching the aggregate levels of employment and number of UI benefit recipients.

4.3 Changes in the Distribution of Earnings

The second step of our methodology to capture high-frequency changes in the distribution of labor income involves estimating changes in the distribution of wages at the monthly frequency. We do this by combining all the available evidence on this issue: monthly and quarterly employment surveys, and the monthly CPS.

Predicting wage inequality from tabulated employment surveys. We first estimate wage inequality monthly using the timely employment censuses and establishment surveys of the Bureau of Labor Statistics. The Quarterly Census of Employment and Wages (QCEW) provides employment and wage statistics for about 95% of employees, based on state and federal unemployment insurance administrative records. At the monthly frequency, the Current Employment Statistics (CES) survey provides similar information based on a representative sample of about 144,000 businesses and government agencies. Both the QCEW and the CES are published in the form of tabulations by industries × geographical areas, up to 6-digits NAICS industry × county × type of ownership (i.e., public or private) in the case of the QCEW. Although the underlying individual-level micro-data are not publicly available, valuable information about the distribution of wages, including at the top, can be retrieved from these granular tabulations.

Building on Lee (2020), we construct quarterly wage income distributions using the QCEW data. The idea is to use the QCEW as if it were a micro-level dataset, treating each 6-digits-NAICS-industry $\times$ county $\times$ type-of-ownership cell as an observation whose weight is the employment count and whose value is the average wage. Each wave of the QCEW contains about a million such observations in recent years, much more than a typical wage survey. We remove outliers, defined as cells whose wage is less than half of a full-time minimum wage job. We then estimate the average wage by percentile, after implementing three adjustments described below. As detailed in Section 5 below, our procedure delivers remarkably accurate predictions of trends in wage inequality.

Adjustments to the QCEW data. The adjustments we apply to the QCEW data are the following. We first convert the QCEW wage data from quarterly to monthly. Employment counts are reported monthly in the QCEW, but wage earnings only quarterly. This is not a significant issue since wages are sticky, so changes in the wage distribution in the short run are driven by changes in the relative employment of low-wage and high-wage workers rather than by changes in their respective salaries. We run the wage data in the QCEW through a moving average of the last twelve months to get smooth monthly wages and to get rid of the seasonality in the wage data (due, for example, to end-of-year bonuses).[16]

Second, because the QCEW data is aggregated in cells, it understates the level of inequality between individual workers. This issue is illustrated on Figure 2. In the annual tax micro-data, the top 1% wage share (among individuals with positive wages) increases from 6% in the late 1970s to about 12% in the 2000s. In the raw but de-seasonalized QCEW data, the top 1% wage share is lower. We fix this discrepancy by implementing a simple adjustment to the monthly series. Specifically, we regress the tax data real wage against the QCEW real wage for each percentile, and use the prediction from these regressions as our monthly estimate for each percentile. This procedure works well because the relationship between the average wage of a given percentile in the QCEW data and the tax data is strongly linear. Importantly, this correction does not vary with time and thus does not weaken the predicting power of the QCEW. Figure 2 illustrates this correction in the case of the top 1% share, which is typically hard to capture with non-tax data. Adjusting the raw de-seasonalized QCEW wage series with

[16]Whenever this procedure introduces missing values, we impute them back by regressing log wages on county, time, industry, and type-of-ownership fixed effects. Remaining seasonal variations (introduced by seasonality in employment numbers) is corrected by running the average wage for each percentile through the X11 procedure.

a multiplier and level shift (constant over the period) generates the adjusted QCEW series depicted in blue. This adjusted series almost perfectly aligns with the actual top 1% wage share observed in the micro tax data in both levels and trends over the entire period.

Third, we supplement the QCEW with the Current Employment Statistics (CES) survey to get timely estimates. The QCEW is published with a lag of one to two quarters. The CES is released every month, albeit with a coarser level of aggregation: about 19,000 monthly series covering up to 300 industries and 450 areas, compared to about a million series in the QCEW. We match each QCEW cell to three CES series. The first matches the location of the QCEW cell as precisely as possible, the second matches it at the state level, and the third at the national level. Because there is a trade-off in the CES between the level of geographical and industry disaggregation, using these three series allows us to extract as much information as possible. We average the trend from these three series in each cell and use this average trend to extend the QCEW data in the most recent quarters. Appendix Figure A4 depicts the quality of this projection by comparing bottom 50% and top 10% wage shares estimated using the raw QCEW data vs. projected using the CES up to 6 months forward (in dashed lines). By definition, the projection matches the QCEW level in the last month of each quarter, and then projects forward over the next two quarters. Overall, the CES projection comes close to the QCEW, although the fit is a bit weaker for the top 10% than for the bottom 50%.

Wage inequality in the monthly CPS. In addition to estimating high-frequency changes in the wage distribution using the QCEW, we also estimate changes in average wages by percentile using the weekly earnings variable of the monthly CPS. Our final estimates of average wage by percentile in a given month is obtained as the average of the CPS and QCEW predictions. For the bottom 80% of the wage distribution, QCEW and CPS predictions are weighted equally. The weight on the CPS prediction then linearly falls to 0 as we move to the 90^{th} percentile, and is 0 above the 90^{th} percentile where the CPS is not informative because of top-coding. Once we have our final estimate of wages by percentile, the distribution within percentiles is interpolated using generalized Pareto interpolation methods (Blanchet et al., 2022).

Adjustment of wage levels in the monthly microfiles. We use the resulting marginal wage income distribution to update our monthly micro-files as follows. In the monthly CPS, we compute the average rank in the wage distribution of each race $\times$ education $\times$ gender $\times$ age $\times$ marital status cell. We also compute these average ranks in the 12 preceding months. In our monthly micro-files, we adjust the average rank by cell to replicate the evolution seen in the

CPS. Once the ranks are adjusted, we assign each individual observation the wage corresponding to his or her adjusted rank. Appendix C.2 provides complete details.

4.4 Distribution of New Government Transfers

In the last step of our methodology, we model the distribution of new government transfers. We simulate the key components of the government response to the Covid crisis: the Paycheck Protection Program, Covid relief payments, and expanded refundable tax credits.

The Paycheck Protection Program was a loan program designed to keep small businesses afloat, representing about $1,000 billion, or 5% of national income. The government forgave most of these loans, assuming companies kept their employees and wages stable. Following Autor et al. (2022), we distribute 70% of the program's expenditures to business owners and the remaining 30% to wage earners. We construct a novel estimate of the program's distributional effect for the incidence on wages. We use the publicly available data on each loan, which we match to the QCEW data based on the date of the loans, the industry, and the location of the business. We manage to match about 5,700,000 loans to 5,500,000 QCEW cells. We estimate both an extensive margin (fraction of the workforce covered) and an intensive margin (fraction of the wage bill covered) for each percentile of the labor income distribution, which we use to simulate the effect of the Paycheck Protection Program on workers.

The three waves of Covid relief payments ("economic impact payments") are allocated based on program rules using taxable income as reported in the updated Piketty, Saez and Zucman (2018) files. Finally, we allocate the expanded refundable tax credits (child tax credit and earned income tax credit) based on income and eligibility using our micro-data.[17]

4.5 Summary of All the Sources Used and Timing of Release

Table 1 summarizes all the sources used to construct our real-time estimates, including frequency of publication and timing of availability. Our approach only relies on public data sources. Appendix D describes the structure of the programs used to construct our real-time estimates. The programs are available online at `https://realtimeinequality.org`, making it possible for researchers to assess and improve any aspect of the methodology.

[17]Refundable tax credits corresponding to incomes earned in year t are generally paid out in year $t+1$ and counted as transfers in year $t+1$ in the national accounts, a convention we follow. In 2021, however, half of the expanded child tax credit was paid monthly (in the last 6 months of 2021) and is assigned to 2021 (not 2022).

5 Validation Tests

Since we apply our methodology back to 1976 (the first year of the QCEW), we have a large number of monthly micro-files that can be used to test the accuracy of our approach. This Section presents the results of these tests.

5.1 Wage Distribution Prediction

We begin by examining how well our monthly wage inequality series match the actual distribution of wage income—by far the largest component of national income—over the 1976–2019 period. Figure 3 compares the wage distribution in the updated Piketty, Saez and Zucman (2018) micro-files, which are based on public tax micro-files, to the distribution constructed in this paper using the QCEW and the CPS. Each panel depicts the share of total wage income earned by a specific group (bottom 50%, middle 40%, next 9%, and top 1%) among adult individuals with positive wage income.

Our monthly estimates track the annual tax-data-based statistics well for all groups, including for the top 1% which is not measured well in traditional household surveys. First, the long-run trend in rising wage concentration is accurately captured by our combination of QCEW and CPS data: as in the annual tax-data-based statistics, the top 1% gains 5.5 points between the late 1970s and 2019 and the next 9% gains 3 points, while the middle 40% loses 7 points and the bottom 50% loses 1.5 point. Second, our methodology accurately predicts short-run variation, the focus of this paper. Over the 44 years t from 1975 to 2018, we correctly capture whether a share has risen or fallen between t and $t+1$ 82% of the time for the top 1%. It is most instructive to focus on inequality dynamics during economic downturns, the periods when the real-time methodology is most valuable from a policy perspective. From 1976 to 2019, there are 8 recession years (1980, 1981, 1982, 1990, 1991, 2001, 2008, 2009). Focusing on these years and the year following each recession (1983, 1992, 2002, 2010), we correctly predict whether the top 1% and bottom 50% shares are falling or rising in 19 cases out of 24. These results are consistent with the analysis of Lee (2020) who first used the QCEW to create quarterly wage distributions for the macroeconomic analysis of the business cycle.

5.2 Volatility of Capital Income

Next, we provide support for our treatment of capital income by comparing the volatility of aggregate capital income components and the volatility of their distributions. Recall that for capital income (corporate profits, rental income, proprietors' income, interest), to form our next year(s) projection we assume that within-component distributions are stable. For example, if the top 10% earns 70% of aggregate corporate profits in 2019, we project that the top 10% still earns 70% of corporate profits in 2020 and 2021. If corporate profits shrink in 2020, individuals in the top 10% of the total income distribution are more affected than individuals in the bottom 90% and all else equal, inequality falls. If aggregate corporate profits soar, inequality increases.

To assess the merits of this procedure, Appendix Figure A5 compares the size of capital income components (as measured by their share of national income) and the concentration of these income components (as measured by the share going to the top 10% of the pretax income distribution). All series are normalized to 100 in 1976 to contrast volatilities. The top left panel reports results for corporate profits, the largest form of capital income, and shows that aggregate profits are highly volatile. During or just before recessions, it is common for the share of profits in national income to fall by 10%–20% (e.g., 1980, 2000, 2008). The share of profits earned by the top 10% exhibits comparatively little year-to-year variation. The same conclusion holds for other components of capital income, as reported in the other three panels. Because changes in aggregates swamp short-term changes in distributions, movements in macroeconomic aggregates capture the bulk of the contribution of capital income to changes in inequality.

5.3 Retrospective Validation

Last and most importantly, we retrospectively check whether our methodology combining prior-year annual micro-files with current-year high-frequency data sources provides accurate estimates of current-year distributions. This test incorporates all forms of income, as opposed to wages or components of capital income only.

Methodology. For each year t from 1975 to 2018 (or 2017), we start from the year t updated annual micro-files of Piketty, Saez and Zucman (2018) and implement our real-time methodology to age these data into a year $t+1$ (or year $t+2$) simulated annual micro-dataset—using the same sources (high-frequency national accounts aggregates, household and employment surveys, etc.) and assumptions as in our real-time methodology.[18] Using these simulated micro-data for year $t+1$ or $t+2$, we can compute any distributional statistics and compare them with the statistics coming out of the actual annual micro-data. The most directly relevant statistic for

[18]The only exception is that for simplicity, we do not attempt to specifically model the distribution of new government transfers created between t and $t{+}1$ (or $t{+}2$), while in our real-time methodology we carefully model the distribution of post-2019 new transfers (see Section 4.4). This means that for disposable and posttax income, our real-time methodology (post-2019) is likely to be more highly predictive than implied by our retrospective tests.

our purposes is real income growth. For each group of the population, we compute actual income growth using the annual distributional national account micro-files for both years t and $t+1$ (or $t+2$), and predicted income growth rates using the actual annual micro-files for year t but the simulated micro-data for year $t+1$ (or $t+2$). The $t+2$ simulations are relevant given that our real-time inequality estimates for year t are typically based on annual micro-files for year $t-2$, due to the nearly 2-year delay in the availability of tax data.

Results. Figure 4 compares actual and predicted real income growth rates for the bottom 50% (top panels) and the middle 40% (bottom panels) of the factor income distribution, with income equally split among married spouses. Left panels show actual vs. predicted growth from year t to $t+1$ and right panels from t to $t+2$. In all cases the dots are closely scattered around the 45-degree line, showing that our predictions are highly informative of the actual growth in bottom 50% and middle 40% incomes. For example, we correctly predict whether bottom 50% is rising or falling 82% of the time one year forward and 91% of the time 2 years forward.[19] One can condition on certain actual growth rates to visually ascertain how well our methodology performs in different contexts. For instance if one conditions on actual growth below -2.5% (or below -5% in the 2-year ahead graph), typically corresponding to recession years, the correlation between actual and predicted growth remains very high. Figure 5 repeats this analysis for top 1% incomes and next 9% incomes. The dots again align well with the 45-degree line. We correctly predict whether top 1% incomes are rising or falling 89% of the time one year forward and 93% of the time 2 years forward.

One caveat when considering the top of the distribution is that predicting short-run growth during tax reforms can be challenging. The largest errors are observed in 1987 and 1988, when predicted growth—though strongly positive—is significantly lower than observed growth. As shown by Figure 3, in those years the QCEW predicts large gains in the top 1% wage share but fails to capture the full magnitude of the rise in this top share. One possible interpretation is that the Tax Reform Act of 1986, which reduced the top marginal income tax rate from 50% in 1986 to 28% in 1988, led to an immediate and across-the-board increase in top-end wages and bonuses for executives within industries $\times$ counties (in addition to gains in specific high-paying industries $\times$ counties, by construction captured by our methodology). More broadly, it can be challenging to predict the growth of top incomes in years of significant tax reforms, which

[19] The 2-year forward prediction performs better by that metric because the bottom 50% often has little annual income growth during our sample period, so that in a few cases predicted growth is slightly negative when actual growth is barely positive (or vice-versa), a problem that attenuates when one considers growth over 2 years.

in addition to real responses can generate avoidance responses, such as inter-temporal income shifting. As shown by Figure 5, the prediction errors in top 1% growth are concentrated during tax reform years (although not all tax reform years lead to errors). In non-tax-reform years our methodology delivers accurate predictions for the top 1%.

Rescaling vs. adjustments of distributions. To better understand which aspects of our methodology matter to generate accurate predictions, Appendix Figure A6 reports similar figures of actual vs. predicted growth rates but using a simplified prediction methodology that only rescales macroeconomic aggregates (following Section 3) without incorporating any changes in the distribution of labor income. The simplified methodology performs well in years of normal growth, but delivers significantly worse results than our full methodology for the bottom 50% during recessions, when it significantly over-estimates growth. This finding echoes the results of Fixler, Gindelsky, and Kornfeld (2021) discussed in Section 2.1 and shows that adjusting the labor income distribution is critical to project inequality during recessions. Once the distribution of labor income is adjusted, our methodology accurately predicts growth during downturns.

Goodness of fit summary. To summarize the performance of our approach, Table 2 reports detailed statistics for goodness of fit and noise. For each group (bottom 50%, middle 40%, next 9%, and top 1%) and income concept (and wealth), we compute the fraction of years in which we correctly predict whether income (or wealth) is growing or falling, the mean difference between predicted and observed 1-year growth, the standard error of this mean, and the root mean square error. For reference we also report the standard deviation of actual 1-year growth rates. We consider three samples of years: all 44 years from 1976 to 2019, the 34 years which are not tax reform years, and the 12 years corresponding to recessions and their immediate aftermath. Appendix Table A1 reports the same statistics for growth over 2 years. A number of results are worth noting.

First, the good fit and limited noise obtained for factor income (Figures 4 and 5) extends to other income concepts—pretax, disposable, and posttax—and to wealth. Across concepts and samples of years, we correctly predict the sign of growth around 90% of the time.[20] Bias in annual growth is limited to a few tenth of a percentage point, which is reasonable considering the sample sizes involved. Remarkably, our methodology is unbiased for the top 1% even though

[20]As noted in footnote 19 for factor income, predictions are slightly worse for the bottom 50% because growth is often close to zero for that group in those decades. If one considers growth over 2 years, we correctly predict the sign of growth for the bottom 50% about 90% of the time just like for other groups (Appendix Table A1).

it does not rely on any tax data. Standard errors for top 1% predictions fall when excluding tax-reform years; for other groups (where avoidance possibilities and incentives to avoid are more limited), including tax reforms does not make a difference. Second, these results carry over to recessions: our methodology is highly predictive of income dynamics during downturns and the ensuing recoveries. We under-estimate disposable and posttax income growth for the bottom 50% during past recessions, but this is because we do not attempt to incorporate the creation of new government transfers during past recessions (e.g., $600 individual tax credits in 2008) in our simulated micro-files. Post-2019 monthly files, by contrast, carefully incorporate all new government programs (as detailed in Section 4.4), maximizing the accuracy of disposable and posttax income predictions at the bottom during and after the Covid-19 pandemic.

6 Inequality During the Covid-19 Pandemic

This Section uses our real-time estimates to analyze the dynamics of income and wealth during the Covid-19 pandemic and in its aftermath. We start by studying the dynamic of income and wages before government intervention, then move to disposable income, before turning to wealth inequality. Unless otherwise noted, all the statistics we report are for "equal-split adults," defined as individual adults with income and wealth equally split between married spouses. On `https://realtimeinequality.org`, we also report statistics at the household level, where a household is a tax unit as defined by the tax code, i.e., either a single person aged 20 or above or a married couple, in both cases with children dependents if any. All growth numbers are adjusted for inflation using the official national income deflator. The same deflator is used for all groups of the population.

6.1 The Dynamic of Factor Income During the Covid-19 Recession

Dynamic across the income distribution. The Covid-19 pandemic led to a dramatic collapse in average national income. Between February 2020 (the last month before the recession) and April 2020 (its trough), annualized real national income per adult fell 15%. Average income then rebounded sharply. But as Figure 6a shows, the fall and recovery were uneven. The economic downturn caused by the pandemic led to the strongest factor income decline for the working-class (-33% for the bottom 50% between February 2020 and April 2020) and to a lesser extent for the top one percent (-19%) due to the collapse of business profits, a key source of income at the top. The crisis affected the middle class and upper-middle class relatively less, because individuals in these groups were more likely to remain employed.

The groups with the largest losses in 2020 experienced the largest gains in 2021. On average, real factor income per adult grew almost 6% in 2021, but by close to 8% for the bottom half of the distribution and by 11.5% for the top one percent. Growth was lower for the middle 40% (close to 3%). According to our estimates, all income groups recovered their pre-crisis factor income level within 20 months, but not at the same pace. Top groups recover faster than the middle class and the working class. For the top 10% and the top 1%, the recovery took twelve months; but for the middle 40% and the bottom 50% it took twenty one months. Because the bottom 50% was hit the hardest and recovered last, the pandemic, had, by the end of 2021, exacerbated factor income inequality. The share of factor income earned by the top 1% was 19.5% in December 2021, its highest level in the post-World War II era.

Comparing the Covid-19 and Great Recession recessions and recoveries. It is well known that in the aggregate, the recovery from the Covid-19 crisis (18 months) was much faster than the recovery from the Great Recession (4 years and a month). Our real-time estimates allow us to move beyond aggregates and compare recovery patterns for the working-class. To do so, Figure 6b focuses on the working-age population (to control for population aging in the 2010s, during the long recovery of the Great Recession) and normalizes income to 100 in the month preceding each recession. Two main results emerge.

First, in the aftermath of the Great Recession it had taken a staggering 8 years and 1 months for the bottom 50% of the working-age population to recover its pre-crisis real factor income level. From 2008 to 2012, a period during which the economy rebounded and crossed its pre-crisis output level, the bottom 50% of working-age adults experienced virtually no growth. Income started growing in 2013 but slowly, so that it only exceeded its December 2007 level in January 2016. The slow recovery of the working class is a robust feature of the Great Recession.[21] It is not an artifact of population aging (since we restrict to the working-age population), but rather reflects the stagnation of wages at the bottom of the distribution (detailed in Section 6.2 below).

Second and by contrast, real factor income for the bottom 50% rebounded more quickly after the Covid recession. By the time average income had recovered from the Great Recession, average income for the bottom 50% was still 10% below its pre-crisis income level and still four years away from a full recovery. By the time average income had recovered from the Covid-19 crisis, by contrast, the bottom 50% was only 4% below its pre-crisis income level and was growing fast. These results vividly illustrate the fact that a given trajectory of GDP growth

[21] The top fiscal income shares of Piketty and Saez (2003), which have been used to study the fraction of growth accruing to top income groups (see Saez, 2008, and subsequent updates), also revealed it.

is compatible with widely different market income dynamics for the working class, highlighting the usefulness of timely and disaggregated growth statistics.

6.2 Wage Growth After Recessions: Covid-19 vs. Great Recession

In the aftermath of the pandemic, the unemployment rate reached historically low levels (3.6% in March 2022), in a context of loose monetary policy (with interest rates of 0% until March 2022) and expansionary fiscal policy. Who benefited most from the tight labor market? To shed light on this issue we can use our micro-files to study the month-to-month dynamics of labor income. The main finding is that the Covid recovery was characterized by a reduction in wage inequality—a break from the trend prevailing since the early 1980s—due to strong wage growth at the bottom of the distribution.

Methodology to study labor income inequality. To establish this result, we analyze changes in labor income inequality in the working-age population (including non-workers) and compute growth rates of labor incomes by percentiles of labor income.[22] Our goal is not to characterize wage growth for a given worker or to fix the composition of the workforce; rather, we want to describe the evolution of the distribution of labor income as comprehensively as possible. The growth statistics by percentile we compute capture the effect of changes in both employment and wages. Our measure of labor income includes wages, supplements to wages and salaries (such as health insurance and retirement benefits) and the labor income of self-employed individuals (defined as 70% of self-employment income), before any tax or deduction for pension contributions. Conceptually, it corresponds to the total cost, for employers, of employing a worker. Our analysis, as always in this paper, is cross-sectional in nature: we do not follow individuals over time.[23]

To provide the context required to interpret changes in the distribution of labor income, Figure 7a depicts the raw monthly employment-to-working-age population ratio from BLS from January 2019 to September 2022. This rate was 78% before the Covid-19 pandemic, fell to 68% at the trough of the recession, and by September 2022 had returned to its pre-Covid level.[24]

[22]To better connect to the labor economics literature and because we have good measures of individual wages, for the analysis of labor income inequality we focus on individualized income series, i.e., we do not split income equally between married spouses.

[23]We build on a long long tradition in labor economics studying changes in cross-sectional wage inequality; see, e.g., Katz and Murphy (1992). The main difference is that our statistics incorporate the entire working-age population including non-workers (as opposed to workers only). This approach allows one to comprehensively capture how government policies affect the labor market, including changes in the extensive margin. With our micro-files, it is also possible to focus on employed individuals only. We view both approaches as complementary.

[24]Employment rates in our micro-files are higher by about 10 percentage points throughout because our micro-

This means that although the composition of the workforce may not be the same in September 2022 as in January 2019, a comparison of the level of labor income in these two months is not confounded by changes in the level of employment, which facilitates the interpretation of labor income growth.

The decline of labor income inequality in 2019–2022. Figure 7c depicts the evolution of average labor income by labor income group from January 2019 to September 2022. Since the bottom quartile of the working-age population is mostly unemployed, we focus on the next three quartiles. We also report income for the top 1%, which earns a sizable fraction of total labor income and cannot be studied with available household surveys. Average labor income in each group is normalized to 100 in January 2019. We can see that from January 2019 to September 2022, real labor income grew significantly in the second quartile, the group with the lowest-wage workers. Because the same number of adults was employed in both months, this growth does not reflect a jobs (i.e., quantity) effect; it reflects the fact that low-paying jobs paid more in September 2022 than in January 2019, by about 10% in real terms. Income grew much less for the third and fourth quartiles: in these groups real average labor income increased by a mere 3% over this 3 years and 9 months period. The relatively strong growth at the bottom illustrates the equalizing effects of tight labor markets.

One caveat is that there is heterogeneity at the top of the labor income distribution. While the top quartile experienced little growth over 2019-2022, the top 1% grew as fast as the second quartile, even slightly more (+12% between January 2019 and September 2022). Consistent with the updated Kopczuk, Saez, and Song (2010) statistics (based on Social Security records) constructed and analyzed by Mishel and Kandra (2021), we find that the top 1% grew faster than other groups in 2020. Moreover, our real-time estimates suggest that the top 1% grew especially fast in the first half of 2021, before plateauing in the second semester and falling slightly in the first three quarters of 2022. Thus, although wage inequality fell from 2019 to September 2022 within the bottom 99%, the share of labor income earned by the top 1% rose.[25]

Analyzing month-to-month changes between January 2019 and September 2022, we can see that these dynamics largely reflect changes in employment. The rise in unemployment during Covid led to a drop in average labor income in all groups, particularly marked at the bottom.

files match annual employment rates, while Figure 7a reports raw actual monthly employment rates (which are mechanically lower); see appendix Figure A3 and Section 4.2 for a discussion.

[25]Another caveat is that, as reported in Appendix Figure A10, we do not find a significant decline of the college labor income premium: in the working-age population, individuals with at least some college education earn on average twice as much labor income as individuals without a college education since the late 2010s, with no trend. This number captures differences in both wages and employment by educational attainment.

The recovery in the second quartile between the trough of the recession and September 2022 was primarily driven by jobs gains, although wage gains played a significant additional role. By the end of 2021 real labor income was stagnating or falling for the third and fourth quartile of the labor income distribution, while the second quartile kept rising.

Labor income inequality: Covid vs. Great Recession. The dynamics of labor income inequality observed during and after the Covid pandemic contrast sharply with the one observed during and after the Great Recession, as depicted in Figure 7. Panel (b) of that figure shows that it took almost 10 years for the employment rate to get back to its pre-Great Recession level. Panel (d) shows that second quartile labor incomes fell the most during Great Recession (as observed during the Covid-19 pandemic), but—in sharp contrast to the Covid-19 recession—continued to fall in real terms until the beginning of 2012, even though the employment rate had already started to recover. Second quartiles real labor incomes did not recover their real pre-Great Recession levels until 2016.

Finally, Figure 8 depicts real labor income growth rates by vingtiles of the labor income distribution above the 25^{th} percentile, from the eve of the Covid recession (February 2020) to September 2022 (as of writing the latest available month, which turns out to have an employment rate equal to that of February 2020), and from the eve of the Great Recession (December 2007) to May 2017 (the month when the employment rate recovered its December 2007 level). In both cases we capture a full employment cycle and labor income growth statistics are not confounded by changes in aggregate employment.

The Great Recession and ensuing recovery were characterized by modest gains in the middle of the labor income distribution—and a stagnation at the bottom and at the top. The Covid cycle is the mirror image: large earnings gains at the bottom and top of the distribution—and losses in the middle. More precisely, during the Covid cycle, labor income grew at annual rates of close to 2% at the bottom. By contrast, between December 2007 and May 2017, average labor income for the second quartile of the working-age population grew only 0.2% a year. By the time the employment rate had recovered its pre-Great Recession level (in May 2017), average earnings for low-wage workers were barely higher than a decade before. Growth was stronger from the 75^{th} to the 95^{th} percentile during the Great Recession, while these groups experienced real losses during the Covid cycle. Finally, the top 1% grew fast during Covid while it stagnated from 2007—a peak year for top labor incomes, which include stock options—to 2017.

6.3 The Effects of Government Intervention

Government intervention during recessions affect the level and distribution of disposable income, sometimes massively. In 2021, average real disposable income per adult in the United States was about 10% higher than in 2019 due to large government deficits.

The bottom 50% most benefitted from the increase in government spending. After accounting for taxes and cash and quasi-cash transfers, average disposable income for the bottom 50% was nearly 20% higher in 2021 than in 2019. Figure 9 shows a step-by-step decomposition of this evolution. To facilitate the interpretation of the results, we focus on the working-age population (aged 20 to 64) and we always rank by factor income so that all figures for a given month refer to the same group.[26] The figure reveals the relative importance of the different government programs enacted during the pandemic.

In the early months of the crisis, the Paycheck Protection Program lifted incomes. But available evidence on the incidence of the program (Autor et al., 2022) implies that the effect is limited: by our estimates, the Paycheck Protection Program increased the average monthly income of the bottom 50% of working-age adults by about $100. It replaced about a fifth of the decline in factor income that occurred in the first months of the crisis for this group (from about $2,000 in February 2020 to about $1,500 in April and May 2020). Unemployment insurance, which was expanded during the crisis, had much larger effects, lifting average bottom 50% monthly income by about $800 in May, June, and July 2020, and by up to $400 a month through to the summer of 2021.

The three waves of Covid-relief payments (April 2020, January 2021, and March 2021) had massive but temporary effects on monthly income. Disposable monthly income for the bottom 50% peaked in March 2021 following the third payment, to reach $4,000—twice as much as before the pandemic ($2,000). For the bottom 50%, disposable income that month was twice as large as factor income (about $2,000). This gap between disposable and factor income was historically high: disposable income is usually close to factor income for the bottom 50% (that is, this group usually pays about as much in taxes as it receives in cash and quasi-cash transfers). By the fall of 2021, disposable monthly income for the bottom 50% had declined to $2,400. The main reason why disposable income was higher for the working class in the end of 2021 than before the pandemic was the expanded child tax credit and the expanded earned income tax credit for adults with children.[27] In the beginning of 2022, disposable income fell as the expanded

[26] Appendix Figure A7 shows disposable income in the full adult population (i.e., not restricting to working-age adults) ranking by factor income; the results are similar.

[27] As in the national accounts, refundable tax credits—i.e., cash transfers administered through the tax

tax credits expired. The only reason why it remained higher than pre-Covid (by about 10%) is that factor income was higher—driven by the real wage gains documented above. In sum, government programs enacted during the pandemic led to a dramatic and unprecedented—but short-lived—improvements in living standards for the lower half of the income distribution.[28]

6.4 Changes in Wealth Concentration

Last, we study the effect of the Covid-19 crisis on wealth inequality. We produce monthly and daily estimates of wealth levels across the distribution. Monthly estimates are obtained by rescaling quarterly Financial Accounts aggregates to their monthly value using real estate and equity indices, and daily estimates by updating equity values using daily stock indices.[29] This makes it possible to track changes in wealth inequality during periods of turmoil in asset markets, which is valuable to simulate wealth effects on consumption in real time and to improve the analysis and management of the business cycle.

Figure 10 shows the monthly dynamic of wealth across the distribution from July 2019 to September 2022, with projections to October 31^{th}, 2022 using our daily methodology. Two distinct phases can be observed. First, until the end of 2021, wealth grew strongly for all groups and wealth concentration rose. From the end of 2019 to the end of 2021, average real wealth per adult grew 26%, primarily due to the rise in asset prices, both in housing and equity markets. For the top 1% the increase was 31% and for the top 0.01% it reached 34%. The share of wealth owned by the top 0.1% adults increased 1.2 point from the end of 2019 to the end of 2021, to reach 18.8%—the highest level recorded in the post-Word War II era. Wealth gains were then partly erased by the decline in stock prices in the first three quarters of 2022. In a context of still rising real estate prices (until the summer of 2022), this decline affected the wealthiest disproportionately, leading to a fall in top wealth shares. More than two-thirds of the wealth gains made in 2020 and 2021 by the top 0.1% were erased in the first 10 months of 2022. By October 2022, the top 0.1% wealth share was below its pre-Covid level.

Our findings on the dynamics of wealth inequality are consistent with other existing evidence.

system—are categorized as cash transfers (not negative taxes); thus the child tax credit and the earned income tax credits show up as "regular cash transfers" in Figure 9.

[28]Appendix Figure A8 shows that the same conclusion holds true when looking at total posttax income, i.e., including Medicaid and Medicare, other in-kind government spending, collective consumption expenditures, and the government deficit.

[29]For housing wealth we use the quarterly Case-Schiller index, projected to the most recent month with the Zillow Home Value index. For equities, we use the Wilshire 5000 Total Market Index, the most extensive representative index of US public companies. We also adjust the wealth of the top 400 daily to match the real-time estimates published by *Forbes*. See Appendix B for complete details.

The Federal Reserve Distributional Financial Accounts (DFA) show similar patterns, both for the Covid-19 crisis and over the long run. If anything, the DFA suggest an even stronger increase in wealth concentration since 1989. From the third quarter of 1989 (the start of the DFA) to the second quarter of 2022, the top 1% wealth share grew 8.6 points according to the Federal Reserve (vs. +5.7 in our series over the same period of time); the next 9% lost -0.4 point (-0.6 in our series); the middle 40% lost -7.4 points (-4.3 in our series); the bottom 50% lost -0.6 point (-0.7 in our series). Given that these two series rely on different distributional sources (a triennal survey of about 6,000 families in the case of the DFA, annual tax data in our case), this similarity suggests that the rise of wealth concentration in the United States since the 1980s is highly robust.[30] In both the DFA and our series, the top 1% wealth share was at its highest recorded level at the end of 2021 and fell by 1–2 percentage points in the first half of 2022.

7 Racial Economic Disparities

The statistical match between survey and tax data implemented in this paper allows us to study the real-time dynamics of racial income disparities, in particular whether Black and white households recovered at the same pace after the Covid-19 recession.

7.1 A Comprehensive Measure of the Black-white Income Gap

To provide context for the analysis of the Covid-19 pandemic, we start by describing the medium-run dynamics of the Black-white income gap. Although a large literature studies racial income disparities (e.g., Bayer and Charles, 2018; Chetty et al., 2020; Derenoncourt and Montialoux, 2021), to date there is no estimate of how average national income—the broadest notion of income—differs for Black vs. white Americans. Due to data limitations (the lack of information on race in tax data and the poor coverage of capital income in household surveys, in particular), most existing statistics focus on earnings or some measure of disposable income. Our approach, by contrast, allows us for the first time to provide a comprehensive measure of the Black-white income gap. Concretely, in 2021 average national income per adult in the United States was around $79,000; with our files we can ask: How does this number differ across racial groups?

[30] As detailed in Saez and Zucman (2020), top wealth shares, although they exhibit the same trend, are lower throughout in level in the DFA than in Saez and Zucman (2016); e.g., the top 1% wealth share rises from 22.5% in 1989Q3 to 31.1% in 2022Q2 in the DFA (vs. 28.6% to 34.3% in our series). This is because in contrast to Saez and Zucman (2016) and this paper, the DFA include unfunded defined benefit pensions and vehicles in wealth, both of which are relatively equally distributed. Once the same definition of wealth is used, the level, trend, and composition of top wealth shares are nearly identical in the two projects (see, e.g., Saez and Zucman, 2020, Figure 1).

Figure 11 shows the average pretax national income of Black adults relative to white adults. On average Black Americans earn half of what white Americans do: $48,000 in 2021 vs. $95,000. This gap is significantly larger than the Black-white earnings gap that is the traditional focus of the literature. As Figure 11 shows, Black Americans (including nonworkers) aged 20 to 64 on average earn 65% of what working-age white Americans do.[31] The Black-white national income gap is even larger because racial disparities in capital income are larger than disparities in labor income: on average Black adults earn only about 20% of the average capital income of white adults. This gap itself primarily reflects the major disparities in wealth reported in Figure 11 and recently studied in, e.g., Derenoncourt et al. (2022).[32] Property ownership remains much more unequal than labor market incomes and this inequality is a key contributor to the persistent racial income disparities that characterize the United States.[33] Capital income is even more unequally distributed than wealth because of differences in yields, in turn coming from differences in asset composition. Relative to white households, a greater fraction of the wealth of Black households is in relatively low-yield assets, namely housing and pensions. Business assets and corporate equity, which have a higher yield, are more concentrated among white households.

Figure 11 also shows that there has been no reduction in the Black-white income gap since the late 1980s. If anything, racial disparities are slightly higher in 2022 than in 1989. Inequality increased from 1989 to 2013; it then started falling in the mid-2010s, in both cases driven by changes in labor income disparities. The recent decline in inequality was not enough to offset the previous increase, so that the average pretax income of Black adults relative to white adults remains lower in 2022 than in 1989.

[31] Restricting to workers, the Black-white earnings gap is lower (Black workers earn about 75% of what white workers do). As shown by Bayer and Charles (2018), taking nonworkers into account is critical to analyze the dynamics of the Black-white earnings gap, due to differential trends in employment.

[32] The Black-white wealth gap displayed in Figure 11 is close to the one in the Federal Reserve Distributional Financial Accounts, the most comparable statistics. In 2019 the average wealth of Black households is 22% of the average wealth of white households in the DFA vs. 25% in Figure 11. The difference is due to the unit of observation (households in DFA vs. adult individuals in Figure 11). Because of differences in household size, racial wealth disparities are slightly larger at the household level. Using our microfiles one can also study racial wealth disparities at the household level; results are identical to the DFA. In the Survey of Consumer Finances (used, e.g., by Derenoncourt et al., 2022), racial wealth disparities are higher than in the DFA and our series because of the different wealth totals (in particular the higher total for private business wealth, whose ownership is concentrated among white households). Trends are similar in all series.

[33] Similarly, Figure A9 show that even though Black individuals account for 12% of the entire adult population, they account for less than 8% of the top 10% of the wage distribution and 4.5% of the top 10% of the wealth distribution.

7.2 Racial Disparities Over the Business Cycle

Turning to the recent dynamic, Figure 12 contrasts income growth at the quarterly frequency during the Great Recession and its aftermath and during the Covid crisis. During the Great Recession, the average income of Black people experienced a prolonged decline of 4 years. In the third quarter of 2011, it was 8% lower than on the eve of the Great Recession, while average income for white working-age adults had already fully recovered. This year corresponds to the peak in Black-white income disparities over the last three decades. Average Black income then started recovering, first at the same pace as for whites, then faster after 2014, leading to a decline in the Black-white income gap that continued until the Covid pandemic hit. The trend for Hispanics mirrors the one seen for Blacks, except that growth was even stronger from 2014.

During Covid, racial disparities were less pronounced than during the Great Recession. The collapse in income in the second and third quarter of 2020 was similar for Black and white people; it was less marked for Hispanics. The different groups then recovered at roughly the same pace: by the first quarter of 2021 all had recovered their pre-crisis pretax income level. We thus again see an illustration of the fact that the Covid-19 recovery was much more equal than the Great Recession recovery.

Appendix Figure A11 reports similar comparisons of average income for men vs. women in the two downturns and recoveries. In both cases income dropped more for men than for women. During the Great Recession, women recovered faster than men (3 years vs. more than 4 years), while during Covid both groups recovered in less than a year. We do not find evidence that the Covid recession had, by the end of 2021, exacerbated gender income inequality.

8 Conclusion

Macroeconomic growth statistics are not necessarily informative of how income grows for most social groups. Yet government statistics currently available globally do not make it possible to know who benefits from economic growth in a timely manner. Our paper attempts to address this gap in the United States by creating monthly income distributions, available within a few hours of the publication of official high-frequency national accounts aggregates. Our methodology, which we retrospectively test and successfully validate back to 1976, combines all publicly available high-frequency data in a unified framework.

Real-time distributional growth statistics could play a critical role in guiding stabilization policies during and in the aftermath of recessions. For example, following a recession, they could be used to estimate "distributional output gaps," that is the extent to which income remains below its pre-recession level or trend for the bottom 50% of the distribution, the next 40%, and the top 10%. Since our files incorporate all taxes and government transfers, they could be used to study whether fiscal policy enacted during a crisis mitigates income losses for the working class on a month-to-month basis.

This project only uses publicly available datasets and our programs are available online at `https://realtimeinequality.org`, making it possible for interested users to examine all the aspects of our approach and refine it. Although our estimation procedure appears to deliver reliable results, our estimates could be improved by complementing the data we use with real-time private sector data (Chetty et al., 2020), by leveraging internal administrative data, or by collecting new administrative data.

A number of potential data improvements are worth highlighting. First, high-frequency national accounts totals are—like all important economic statistics—still a work in progress and could be refined. It would be valuable to develop processes to remove the statistical discrepancy between GDI and GDP, i.e., to systematically reconcile the income and expenditure approach. This would allow BEA to produce a single unified estimate of quarterly growth, as many countries do. Progress in that direction could be facilitated by a more systematic exploitation of high-frequency data on corporate profits, such as listed companies' quarterly earnings statements.[34] The national accounts could also be improved by reporting separate profits estimates for public vs. private companies (or C-corporations vs. S-corporations). Because profits of private companies tend to be more concentrated towards the top of the income distribution, this additional breakdown would improve the accuracy of distributional estimates.[35] Last, government agencies could produce additional high-frequency inequality estimates. Most importantly, the Bureau of Labor Statistics could compute a quarterly individual-level wage distribution using the administrative unemployment insurance micro-data that underly the QCEW. We view our real-time statistics as a prototype which we hope will be refined, enriched, and eventually incorporated into official national account statistics.

[34]Public companies must submit quarterly earnings to the Securities and Exchange Commission within 40 days. If this delay was reduced to less than 30 days, these data could be used by BEA as a key input to form an estimate of corporate profits within a month of the end of each quarter. In turn, this would make it possible for BEA to simultaneously estimate growth from both the income and expenditure approach, making it easier to integrate data from these two approaches into a single number.

[35]Krakower et al. (2021) present prototype NIPA estimates of profits for S-corporations, but there are no quarterly estimates to date. The Census Bureau Quarterly Financial Report—which serve as a key input for the estimation of quarterly profits by BEA—could include tabulations for public vs. private firms separately.

References

Aspachs, Oriol, Ruben Durante, Alberto Graziano, Josep Mestres, José Garcia-Montalvo and Marta Reynal-Querol. 2020. "Real-Time Inequality and the Welfare State in Motion: Evidence from COVID-19 in Spain." BSE Working Paper: No.1202.

Autor, David et al. 2022. "The $800 Billion Paycheck Protection Program: Where Did the Money Go and Why Did it Go There?," *Journal of Economic Perspectives,* 36(2), 55–80.

Alvaredo, Facundo, Anthony Atkinson, Lucas Chancel, Thomas Piketty, Emmanuel Saez, and Gabriel Zucman. 2016. "Distributional National Accounts (DINA) Guidelines: Concepts and Methods used in the World Wealth and Income Database." WID WP 2016/1.

Batty, Michael et al. 2019. "Introducing the Distributional Financial Accounts of the United States," Finance and Economics Discussion Series 2019-017. Washington: Board of Governors of the Federal Reserve System.

Bayer, Patrick and Kerwin Kofi Charles. 2018. "Divergent Paths: A New Perspective on Earnings Differences Between Black and White Men Since 1940," *Quarterly Journal of Economics,* 133(3): 1459–1501.

Blanchet, Thomas, Lucas Chancel, Ignacio Flores, Marc Morgan. 2021. "Distributional National Accounts Guidelines: Methods and Concepts used in the World Inequality Database." Technical paper, 2021.

Blanchet, Thomas, Lucas Chancel, Ignacio Flores, Marc Morgan. 2022. "Generalized Pareto Curves: Theory and Applications," *Review of Income and Wealth,* 68(1), 263–288.

Bounie, David, Youssouf Camara, Etienne Fize, John Galbraith, Camille Landais, Chloe Lavest, Tatiana Pazem, B. Savatier. 2020. "Consumption Dynamics in the COVID Crisis: Real Time Insights from French Transaction Bank Data." *Covid Economics* 59: 1-39.

Card, David, Raj Chetty, Martin Feldstein, and Emmanuel Saez. 2010 "Expanding Access to Administrative Data for Research in the United States", NSF white paper.

Chetty, Raj, John N. Friedman, Nathaniel Hendren, Michael Stepner, and the Opportunity Insights Team. 2020. "The Economic Impacts of Covid-19: Evidence from a New Public Database Built Using Private Sector Data." NBER working paper #27431

Chetty, Raj, John N. Friedman, Nathaniel Hendren, Maggie R. Jones, and Sonya R. Porter. 2020. "Race and Economic Opportunity in the United States: An Intergenerational Perspective," *Quarterly Journal of Economics* 135(2): 711–783.

Derenoncourt, Ellora, and Claire Montialoux. 2021. "Minimum Wages and Racial Inequality," *Quarterly Journal of Economics,* 136(1): 169-228.

Derenoncourt, Ellora, Chi Hyun Kim, Moritz Kuhn, and Moritz Schularick. 2022. "Wealth of Two Nations: The U.S. racial wealth gap, 1860-2020." NBER WP #30101.

Daly Mary, Bart Hobijn, and Theodore Wiles. 2011. "Dissecting Aggregate Real Wage Fluctuations: Individual Wage Growth and the Composition Effect," FRBSF WP 2011-23.

Denton, Frank T.. 1971. "Adjustment of Monthly or Quarterly Series to Annual Totals: An Approach Based on Quadratic Minimization," *JASA* 66(333): 99–102.

Fixler, Dennis, David Johnson, Andrew Craig, Kevin Furlong. 2017. "A Consistent Data Series to Evaluate Growth and Inequality in the National Accounts," *Review of Income and Wealth,* 63: 437–459.

Fixler, Dennis, Eva de Francisco, and Danit Kanal. 2021. "The Revisions to Gross

Domestic Product, Gross Domestic Income, and Their Major Components," *Survey of Current Business.* 101(1): 1–40.

Fixler, Dennis, Marina Gindelsky , and Robert Kornfeld. 2021. "The Feasibility of a Quarterly Distribution of Personal Income," BEA Working Paper Series, WP2021-8.

International Monetary Fund 2017. "Quarterly National Accounts Manual, 2017 Edition."

Katz, Lawrence F. and Kevin M. Murphy. 1992, "Changes in Relative Wages, 1963-1987: Supply and Demand Factors," *Quarterly Journal of Economics* 107(1), 35–78.

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

Krakower, Howard, Jennifer Lee, Kate Pinard, and Marlyn Rodriguez. 2021. "Prototype NIPA Estimates of Profits for S Corporations." Bureau of Economic Analysis.

Lee, Byoungchan. 2020. "Business Cycles and Earnings Inequality." working paper.

Mishel, Lawrence and Jori Kandra. 2021. "Wage inequality continued to increase in 2020," Economic Policy Institute report.

Parolin, Zachary, Megan Curran, Jordan Matsudaira, Jane Waldfogel, and Christopher Wimer. 2022. "Estimating Monthly Poverty Rates in the United States." forthcoming *Journal of Policy Analysis and Management*.

Piketty, Thomas, and Emmanuel Saez. 2003. "Income Inequality in the United States, 1913-1998," *Quarterly Journal of Economics* 118(1), 1–39.

Piketty, Thomas, Emmanuel Saez, and Gabriel Zucman. 2018. "Distributional National Accounts: Methods and Estimates for the United States," *Quarterly Journal of Economics* 133(2), 553–609.

Rodgers, Willard L. 1984. "An Evaluation of Statistical Matching," *Journal of Business & Economic Statistics* 2(1), 91–102.

Saez, Emmanuel. 2008. "Striking It Richer: The Evolution of Top Incomes in the United States," *Pathways Magazine*, Stanford Center for the Study of Poverty and Inequality, 6-7.

Saez, Emmanuel and Gabriel Zucman. 2016."Wealth Inequality in the United States since 1913: Evidence from Capitalized Income Tax Data," *Quarterly Journal of Economics* 131(2).

Saez, Emmanuel and Gabriel Zucman. 2020. "The Rise of Income and Wealth Inequality in America: Evidence from Distributional Macroeconomic Accounts," *Journal of Economic Perspectives*, 34(4): 3-26.

Saez, Emmanuel and Gabriel Zucman. 2020b. "Trends in US Income and Wealth Inequality: Revising After the Revisionists," NBER working paper #27921.

Shiskin, J., A. H. Young and J.C. Musgrave. 1967. "The X-11 Variant of the Census Method II Seasonal Adjustment Program," Technical Report 15, Bureau of the Census.

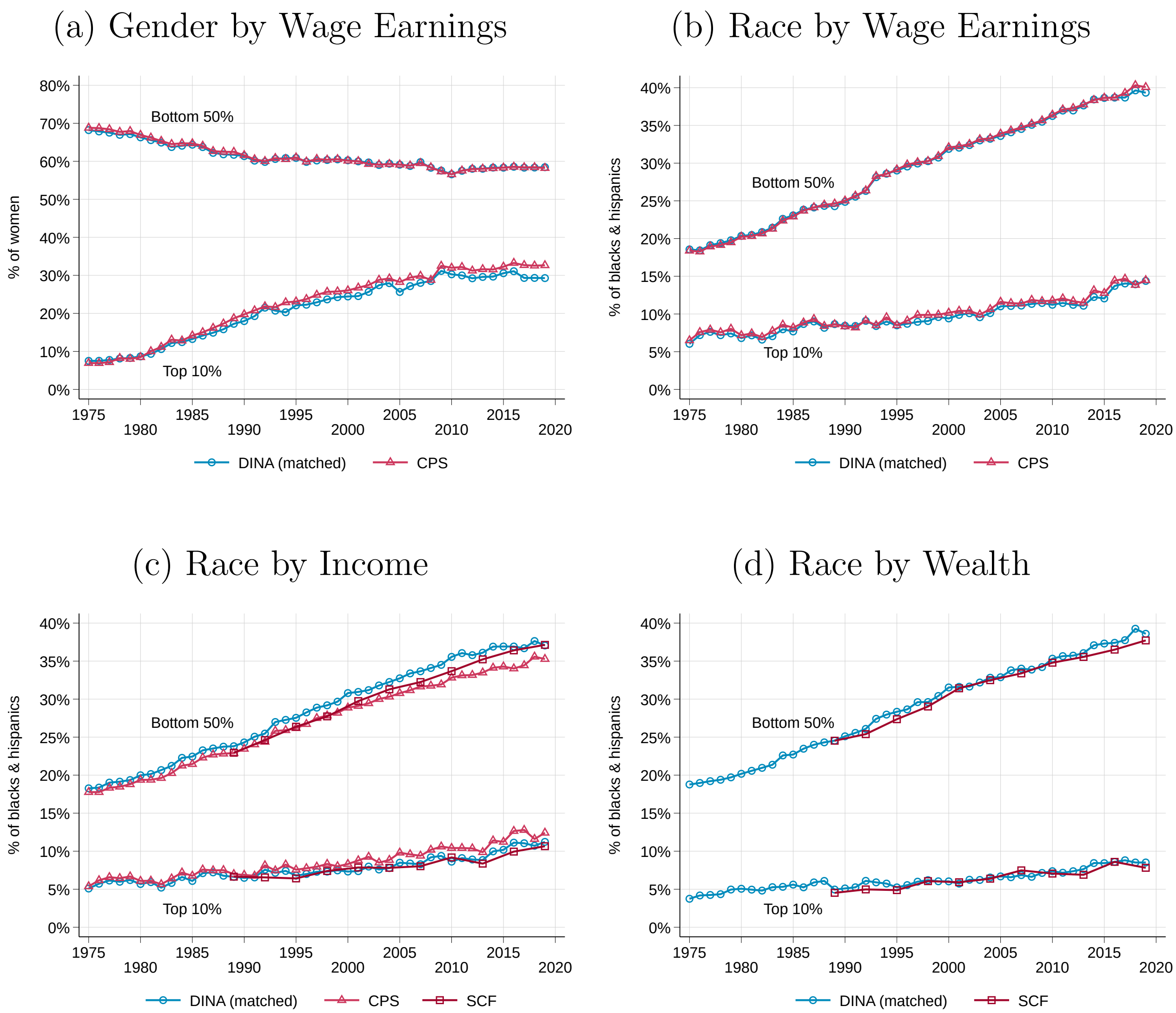
Stantcheva, Stefanie. 2022. "Inequalities in the Times of a Pandemic," NBER WP #29657.

Stiglitz Joseph, Amartya Sen, and Jean-Paul Fitoussi. 2009. *Report by the Commission on the Measurement of Economic Performance and Social Progress*, INSEE France.

United Nations. 2009. "System of National Accounts 2008," European Communities, International Monetary Fund, OECD, United Nations and World Bank.

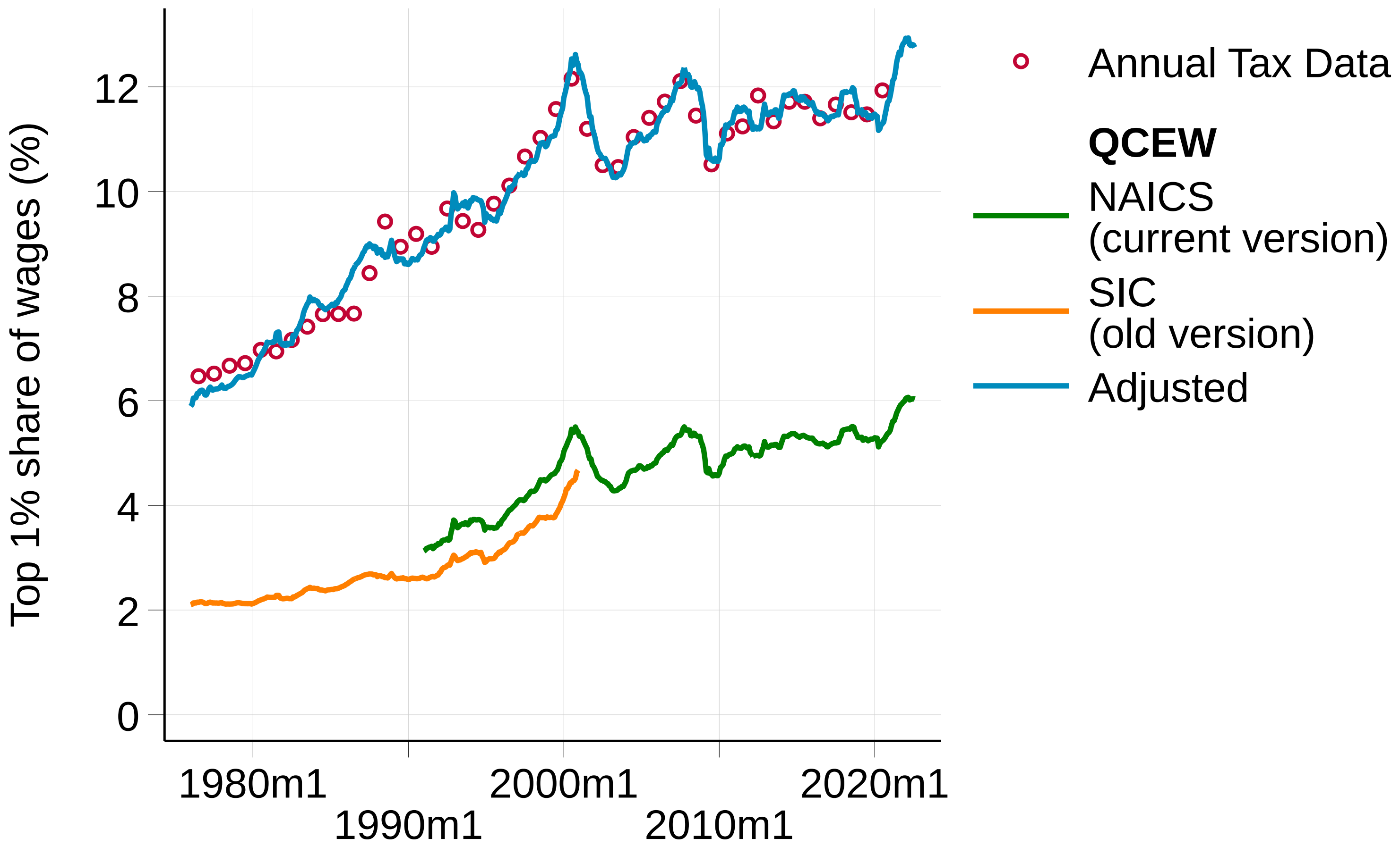
US Census Bureau. 2022. "Quarterly Financial Report: U.S. Manufacturing, Mining, Wholesale Trade, and Selected Service Industries" online at `https://www.census.gov/econ/qfr/`

Figure 1: Testing the One-to-One Statistical Match



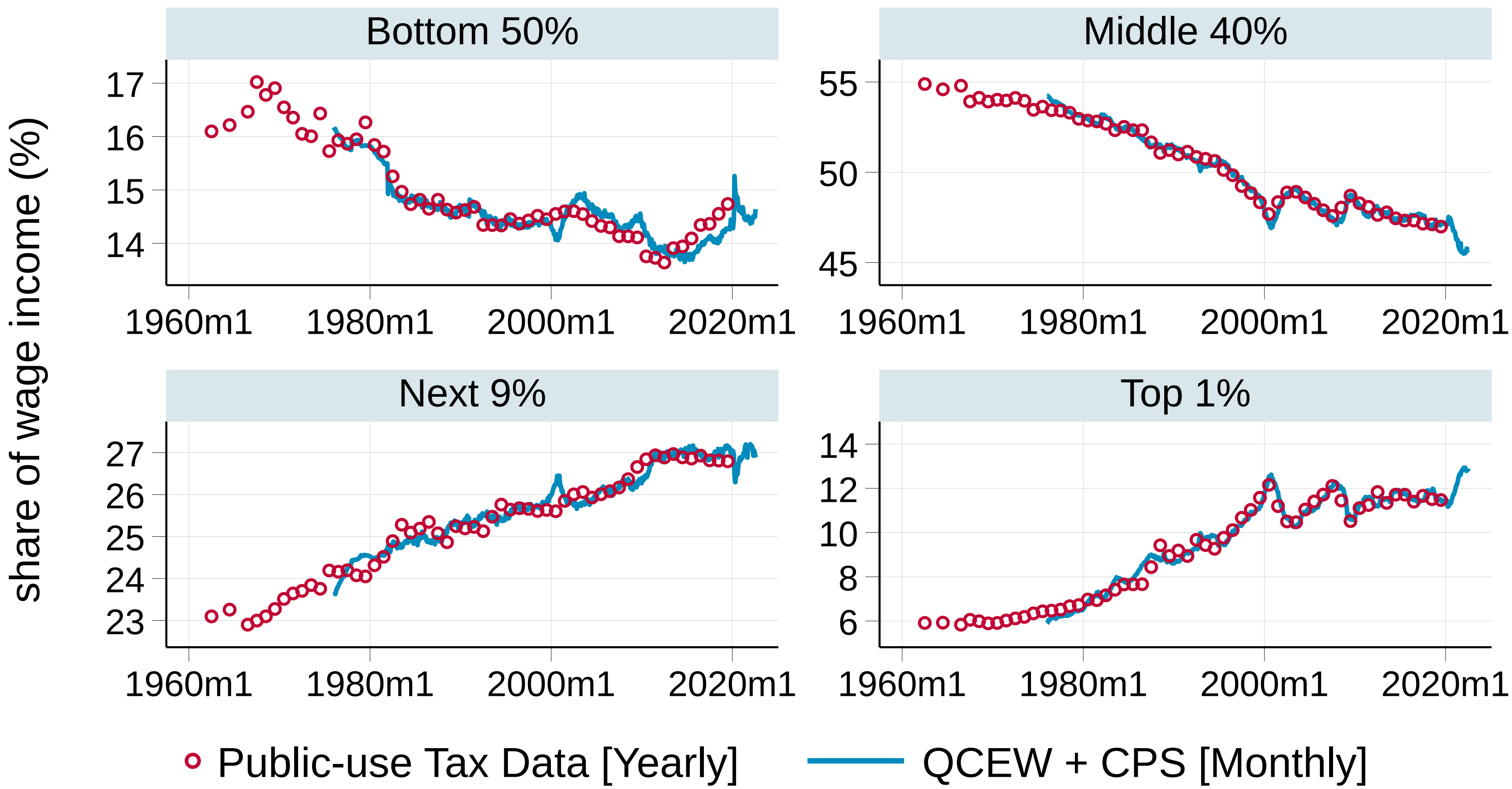
Notes: This figure compares the demographic composition along one key income or wealth variable of interest in the original survey data vs. the same key income or wealth variable from the tax data in our statistically matched dataset. Panel (a) considers gender composition by annual wage earnings brackets in the CPS vs. matched data. Panel (b) considers racial composition by annual wage earnings brackets. In both panels (a) and (b), the sample includes zeros (individuals with no wage earnings) but only includes working-age individuals (20–64). Panel (a) uses individual earnings, while panel (b) divides earnings equally among married spouses. Panel (c) considers race composition by annual total income bracket in the CPS, the SCF, and matched data. For the purpose of that exercise, total income is the sum of wages, pensions, Social Security benefits, business income, interest, dividends, and rents. Panel (d) considers the same race composition by wealth bracket in the SCF vs. matched data. In both panels (c) and (d), we consider the entire CPS (and/or SCF) samples.

Figure 2: Top 1% Wage Share: Adjusting the QCEW



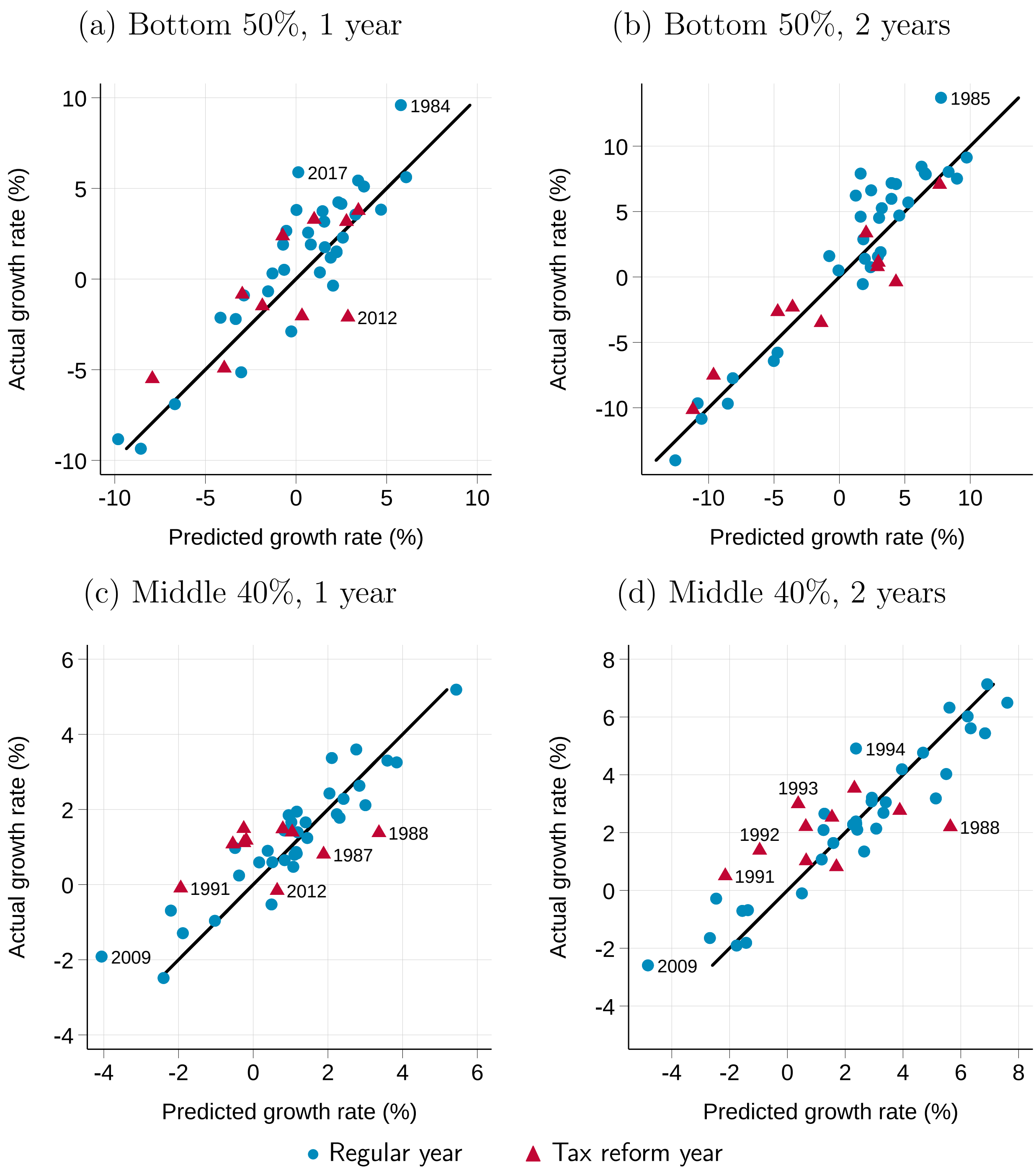
Notes: This figure depicts the top 1% wage share (among individuals with positive wages) in the annual tax data (in red circles) and in the raw but de-seasonalized QCEW data (orange series for the old QCEW based on SIC industry codes, and green series for the modern QCEW based on more granular NAICS industry codes). The top 1% wage share is substantially lower in the raw QCEW because the data is aggregated by county×industry cells, but the trends are similar. Adjusting the QCEW series with a multiplier plus level shift constant over the period (based on a time series regression approach as described in the text) generates the blue adjusted QCEW series that almost perfectly aligns with the tax data both in levels and trends over the full period.

Figure 3: Wage Distributions: Tax Data vs. QCEW & CPS Based Estimates



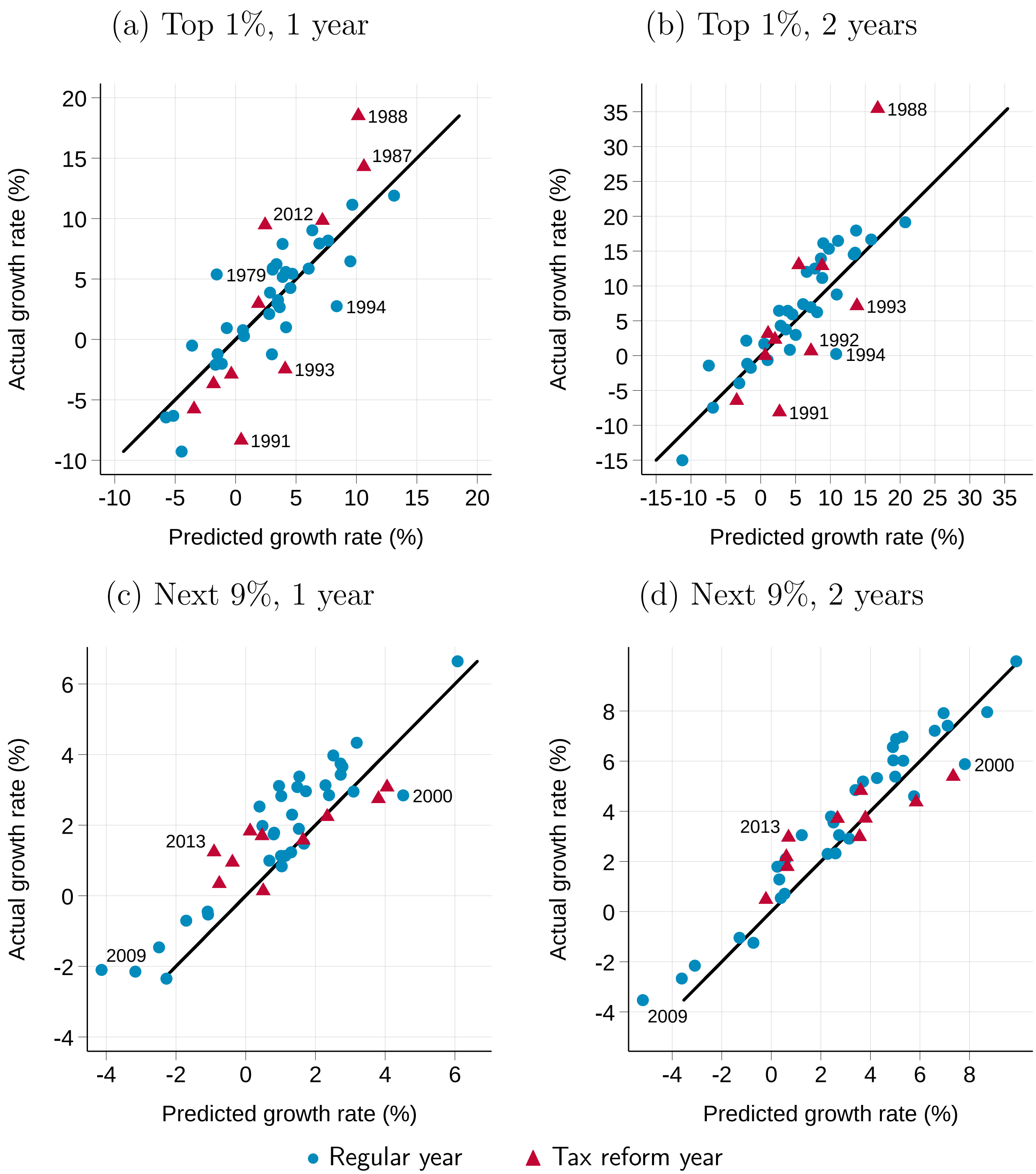
Notes: This figure compares the wage distributions in the annual distributional national accounts of Piketty, Saez and Zucman (2018, updated), which are based on public tax micro-data, and those obtained in our monthly micro-files using the QCEW and the CPS as described in Section 4.3. Each panel depicts the share of total wage income earned by a specific group (bottom 50%, middle 40%, next 9%, and top 1%) of individuals with positive wage income. Wages are individualized (they are not equally split between married spouses). The monthly series estimated from the QCEW and CPS track the annual micro-data closely for all groups, including the top 1% which is not measured well in traditional survey data.

Figure 4: Actual vs. Predicted Growth at the Bottom



Notes: This figure compares predicted to actual growth in average real factor income per adult (with income equally split among married spouses) for the bottom 50% (top panels) and the next 40% (bottom panels). Growth is computed from year t to $t+1$ (left panels) and from t to $t+2$ (right panels) for each year t from 1975 to 2018 (2017 in the right panels). Actual growth is obtained using the annual distributional national account micro-data for both years t and $t+1$ (or $t+2$). Predicted growth is obtained using the annual micro-data for year t but the projected micro-data using our methodology for $t+1$ (or $t+2$). Years of significant tax reforms (which can generate income shifting) are shown in red. Years $t+1$ (or $t+2$) with significant prediction errors are labelled. Overall, the dots align well with the 45-degree line depicted on the graphs: our methodology accurately predicts growth at the bottom and in the middle of the distribution.

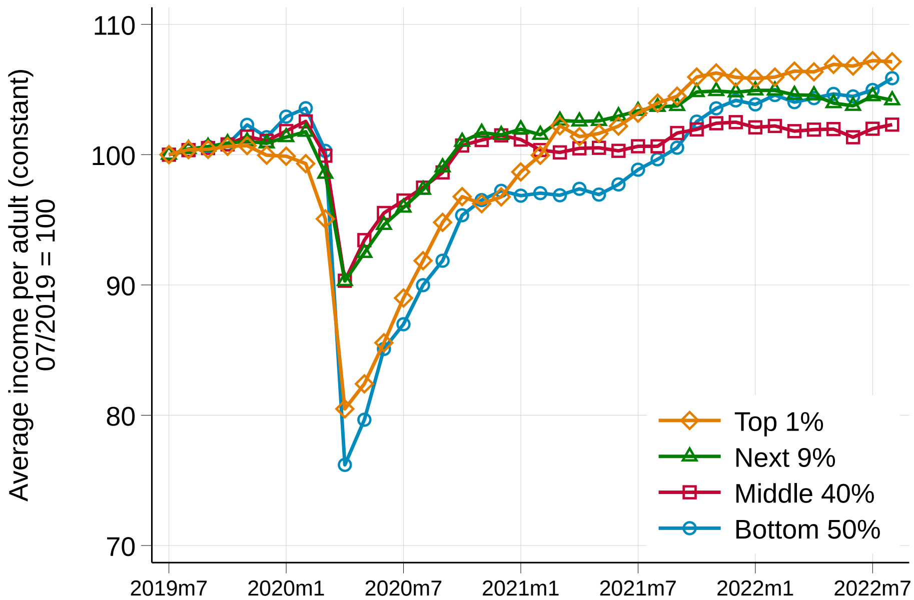
Figure 5: Actual vs. Predicted Growth at the Top



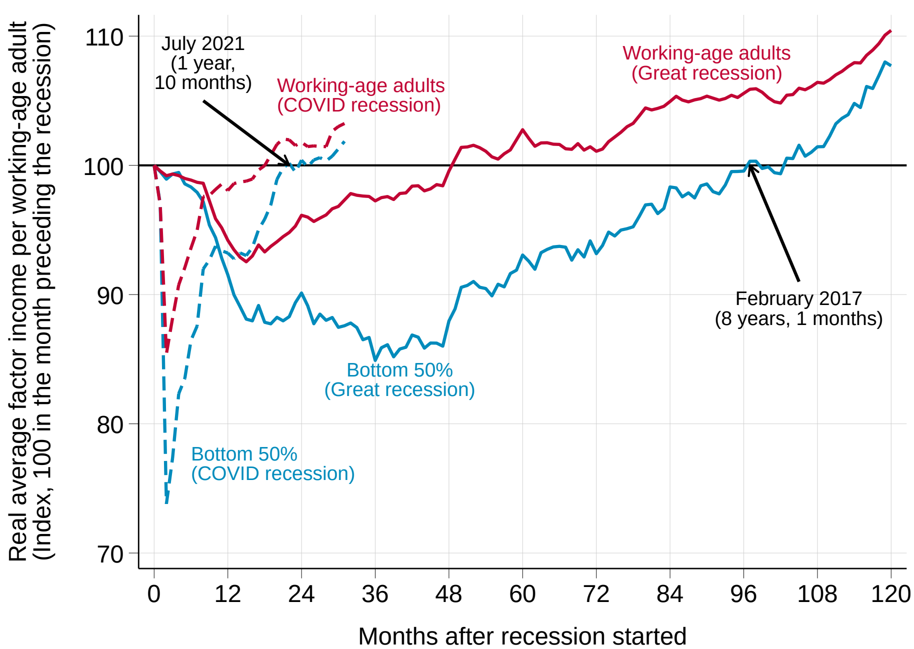
Notes: This figure compares predicted to actual growth in average real factor income per adult (with income equally split among married spouses) for the top 1% (top panels) and the next 9% (bottom panels). Growth is computed from year t to $t+1$ (left panels) and from t to $t+2$ (right panels) for each year t from 1975 to 2018 (2017 in the right panels). Actual growth is obtained using the annual distributional national account micro-data for both years t and $t+1$ (or $t+2$). Predicted growth is obtained using the annual micro-data for year t but the projected micro-data using our methodology for $t+1$ (or $t+2$). Years of significant tax reforms (which can generate income shifting) are shown in red. Years $t+1$ (or $t+2$) with significant prediction errors are labelled. Overall, the dots align well with the 45-degree line depicted on the graphs: our methodology accurately predicts the growth of top incomes.

Figure 6: Factor Income: Covid-19 vs. Great Recession

(a) Real Factor Income Around the Covid-19 Pandemic

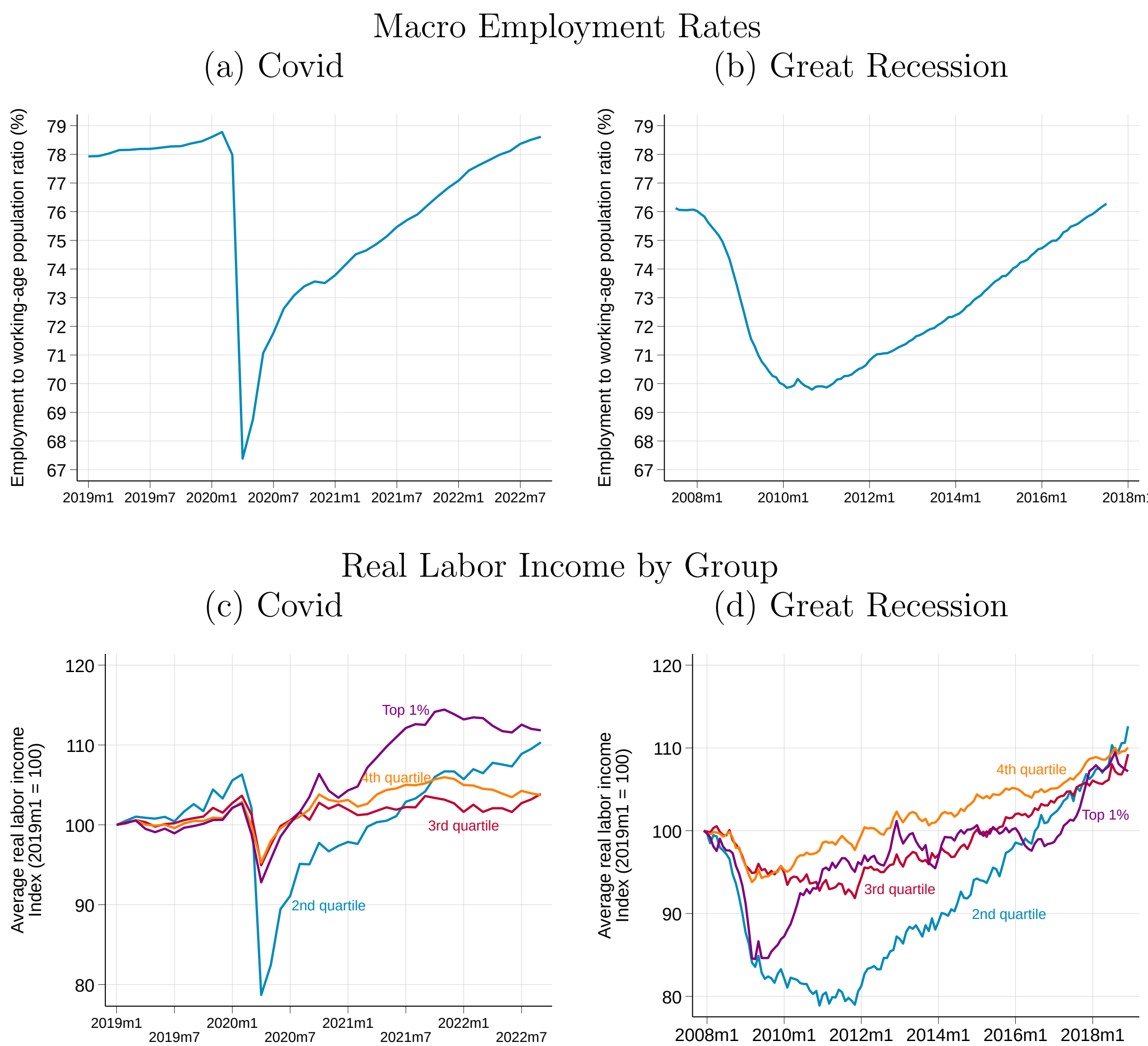


(b) Income Dynamics: Covid-19 vs. Great Recession



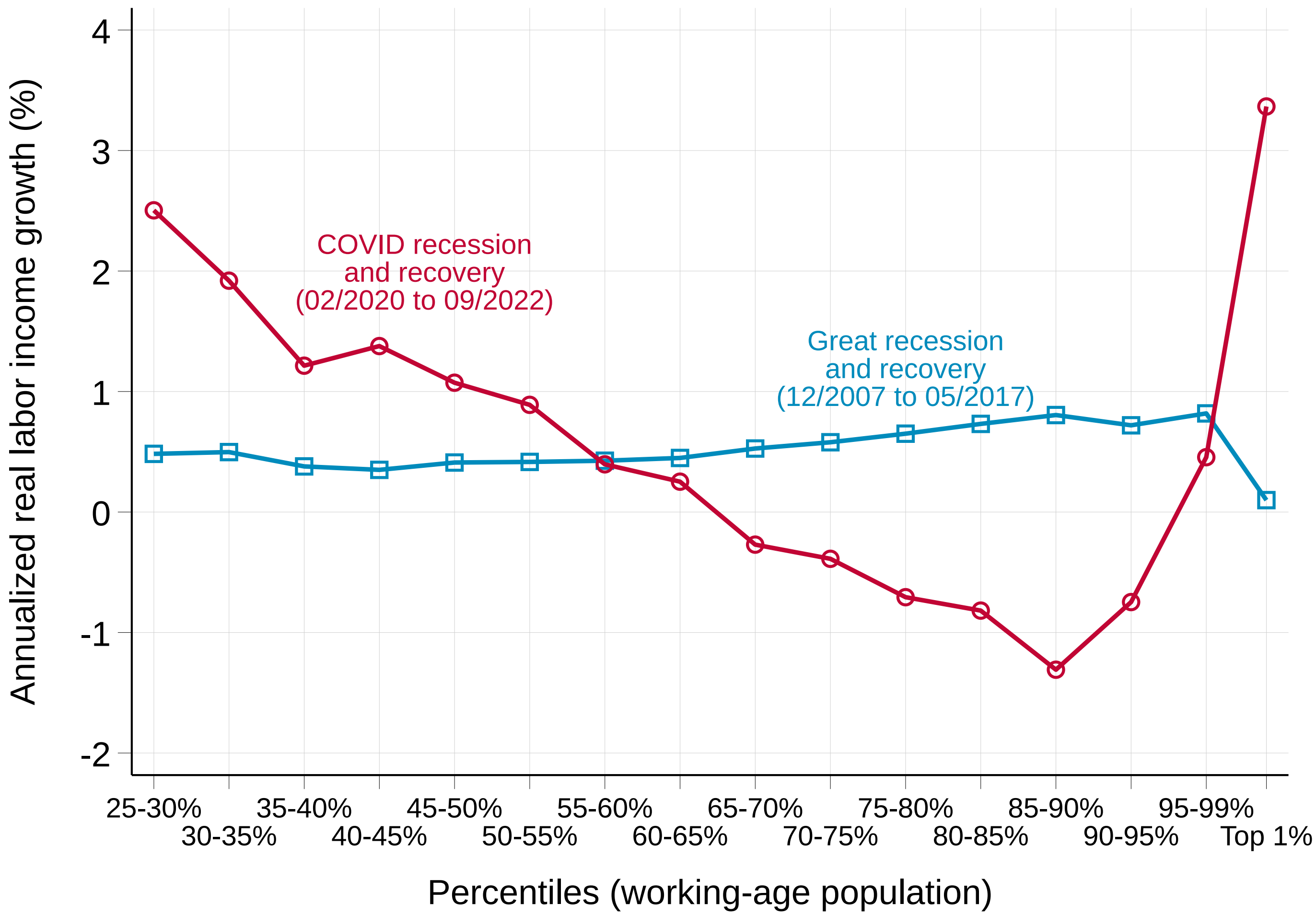
Notes: Panel (a) shows the monthly dynamic of real factor income per adult (with income equally split among married spouses) around the Covid-19 pandemic. The pandemic led to the strongest income declines for the bottom 50% and to a lesser extent for the top one percent. By October 2021 all groups had recovered their pre-crisis income level. Panel (b) compares the growth of real factor income per working-age adult (with income equally split among married spouses) on average and for the bottom 50% of the working age population during the Great Recession and the Covid-19 recession. We restrict to the working-age population (20 to 64) to control for population aging in the 2010s. Income is normalized to 100 in December 2007 for the Great Recession and February 2020 for the Covid-19 recession, corresponding to the month immediately preceding each recession. The x-axis counts the number of months since the start of each recession.

Figure 7: Employment and Earnings Growth: Covid vs. Great Recession



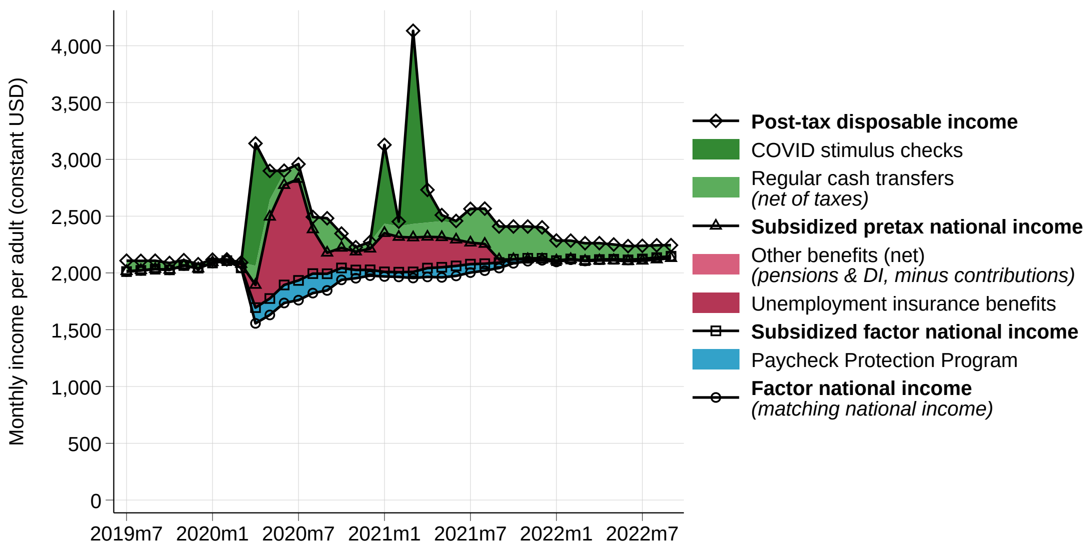
Notes: The top panels depict the evolution of the monthly employment to working-age population ratio, defined as seasonally-adjusted nonfarm employment (from the BLS Current Employment Statistics establishment survey) divided by the number of adults aged 20 to 64 during Covid in panel (a) and the Great Recession in panel (b). This measure of employment excludes proprietors, private household employees, unpaid volunteers, farm employees, and the unincorporated self-employed. The bottom panels depict the average real labor income for various fractiles of the labor income distribution among adults aged 20 to 64 (including non-workers) during Covid in panel (a) and the Great Recession in panel (b), with base 100 at the start of each period. Labor income is individualized (i.e., not equally split between married spouses) and includes all wages and salaries, supplements to wages and salaries, and 70% of self-employment income.

Figure 8: Earnings Growth Across the Distribution: Covid-19 vs. Great Recession



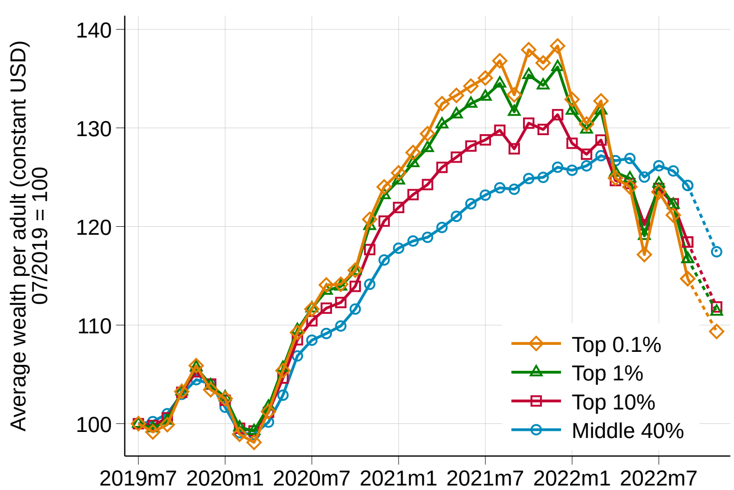
Notes: This figure shows the annualized growth rate of real labor income by vingtile of the working-age population (with a zoom on the top 1%), from the eve of the Covid recession (February 2020) to September 2022, and from the eve of the Great Recession (December 2007) to May 2017. In both cases, we capture a full employment cycle (i.e., May 2017 is the month when the employment rate had returned to its pre-Great Recession level; and by September 2022 the employment rate had returned to its pre-Covid level). Labor income is individualized (i.e., not equally split between married spouses) and includes wages, supplements to wages and salaries, and 70% of self-employment income. We include all working-age adults (aged 20 to 64), including non-workers. The graph starts at the 25^{th} percentile since the bottom quartile of working-age adults is mostly unemployed. The figure shows that the bottom (and top) of the labor income distribution experienced fast growth from the beginning of 2020 to September 2022, in contrast with the recovery from the Great Recession.

Figure 9: Income of the Bottom 50% during the Covid Crisis



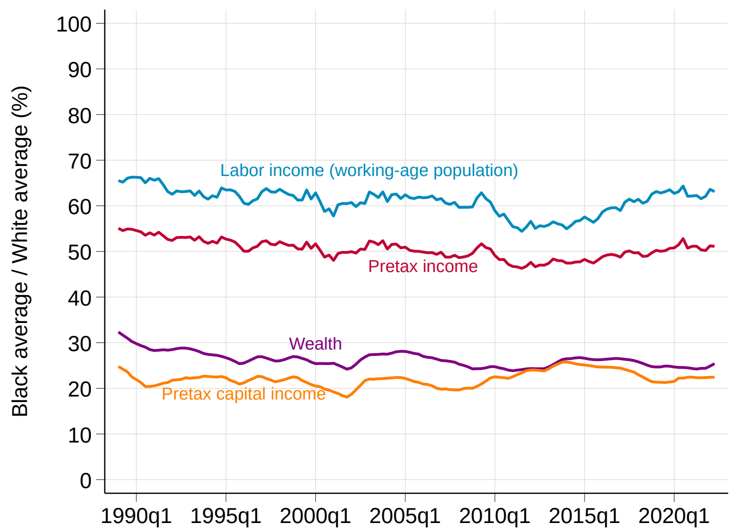
Notes: This figure decomposes the average real monthly income of the bottom 50% working age (20-64) adults from July 2019 to September 2022. Individual adults are ranked by their factor income, and income is equally split between married spouses. The figure reveals the relative importance of the different government programs enacted during the Covid-19 pandemic, most importantly the three waves of Covid-relief payments (April 2020, January 2021, and March 2021), the expansion of unemployment insurance, the expansion of refundable tax credits (EITC and child tax credit), and the Paycheck Protection Program. By the beginning of 2022 all of these programs had expired, and the only reason why average bottom 50% disposable income remained higher than pre-Covid (by about 10%) was the higher level of factor income.

Figure 10: Wealth Growth During and After Covid



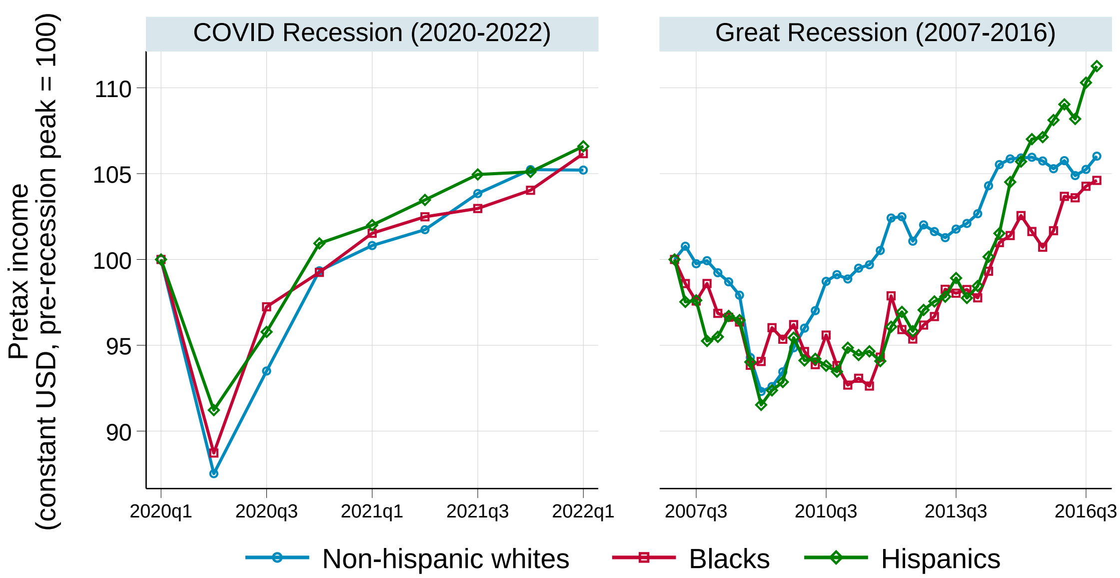
Notes: This figure shows the evolution of average real wealth per adult (with wealth equally split between married spouses) in the middle 40% of the wealth distribution (i.e., from the median to the 90^{th} percentile), the top 10%, the top 1%, and the top 0.1%, from July 2019 to September 2022 (plain line) and projected values for October 31^{st}, 2022 (dotted line) using our methodology for producing daily estimates. The average wealth of each group is normalized to 100 in July 2019. The figure shows, e.g., that the wealth of the top 0.1% increased by close to 40% (adjusted for inflation) from July 2019 to December 2021 and that more two thirds these gains were erased in the first ten months of 2022.

Figure 11: Black-White Economic Disparities



Notes: This figure shows Black-white differences in income and wealth. The unit of observation is the individual adult. The labor income line restricts to the working-age population (individuals aged 20 to 64); other series include the entire adult population. The series presented are quarterly and start in 1989, the first year of the Survey of Consumer Finances.

Figure 12: Income Dynamics by Race: Covid vs. Great Recession



Notes: This figure shows the evolution of average real pretax national income by racial group during the Covid recession and its aftermath (left panel) and the Great Recession and its aftermath (right panel). Income is normalized to 100 in the quarter preceding each recession.

Table 1: Overview of Sources

Usage	Source	Producer	Usage	Frenquency	Lag	Notes
Construction of annual microdata	Public-use tax data	Internal Revenue Service	Main microdata source.	Yearly	1–2 years	The last public-use microfile dates from 2014. For later years, we update the microdata using IRS tabulations of income.
	Social Security Wage Statistics	Social Security Administration	Complementary data on the wage income distribution.	Yearly	6 months	We use the wage income distribution from the SSA because it is better at capturing low wages.
	Current Population Survey (ASEC)	Census Bureau (via IPUMS)	Integration of socio-demographic information into the tax data (bottom 95%).	Yearly	6 months	We match the CPS to the tax data using optimal transport on detailed income variables.
	Survey of Consumer Finances	Federal Reserve	Integration of socio-demographic information into the tax data (top 5%).	Triennial	1 year	We match the SCF to the tax data using optimal transport on detailed income and wealth variables.
Intra-annual distribution adjustments	Quarterly Census of Employment and Wages (QCEW)	Bureau of Labor Statistics (BLS)	Estimation of the wage income distribution.	Quarterly	5 months	The wage data in the QCEW is quarterly but the employment data is monthly. Therefore we treat the QCEW as a monthly dataset.
	Current Employment Statistics (State and Area)	Bureau of Labor Statistics (BLS)	Estimation of the wage income distribution.	Monthly	3 weeks	This data is similar to the QCEW but coarser and released more frequently. We match it to the QCEW to extrapolate the evolution of earnings and employment in the most recent months.
	Current Employment Statistics (National)	Bureau of Labor Statistics (BLS)	Estimation of the number of wage earners.	Monthly	3 weeks	We use aggregate nonfarm employment statistics to adjust the number of wage earners monthly.
	Unemployment Insurance Weekly Claims	Department of Labor	Estimation of the number of unemployment insurance claims.	Weekly	1 week	We average the series by month, adjust for seasonal variations, and use it to adjust the number of unemployment insurance recipients by month.
	Current Population Survey (Monthly)	Census Bureau (via IPUMS)	Estimation of the wage income distribution, and the ranks in the earnings distribution, and of the number of wage earners by age, gender, education, race and marital status.	Monthly	2 weeks	We use the CPS to estimate wage income for the bottom 90% of the distribution only (because of top-coding), and average that estimate with the one from the QCEW. For the top 10% we only rely on the QCEW.
	Real-time Billionaires	Forbes Magazine	Wealth for the 400 richest Americans.	Daily	None	We adjust the overall wealth of the 400 wealthiest households to match the real-time Forbes estimate.
Estimation and rescaling to macro aggregates	National Income and Product Accounts (NIPA)	Bureau of Economic Analysis (BEA)	Estimation of monthly income aggregates.	Monthly / Quarterly	1 month	When needed, we disaggregate the quarterly or yearly data at the monthly frequency using Denton's (1971) method. With the exception of corporate profits, the most important income components are available monthly.
	Financial Accounts	Federal Reserve	Estimation of quarterly wealth aggregates.	Quarterly	2 months	When needed, we disaggregate yearly data at the quarterly frequency using Denton's (1971) method. The most important wealth components are avaialble quarterly.
	Wilshire 5000 Total Market Index	Wilshire Associates (via FRED)	Disaggregation of the value of stocks at the monthly frequency.	Daily	None	We use the index to disaggregate the quarterly estimate of stocks values at the monthly level using Denton's (1971) method.
	Case-Shiller National Home Price Index	Standard & Poor's (via FRED)	Disaggregation of the value of housing at the monthly frequency.	Monthly	2 months	We use the index to disaggregate the quarterly estimate of housing values at the monthly level using Denton's (1971) method.
	Zillow Home Value Index	Zillow	Disaggregation of the value of housing at the monthly frequency.	Monthly	2 weeks	We use the Zillow index to extrapolate the Case-Shiller in the latest months.

Table 2: Prediction Errors for Growth Rates of Income and Wealth

Concept	Bracket	All years					Excl. tax reforms				Recessions			
		Std. Dev.	Correct sign	RMSE	Bias	Std. Err.	Correct sign	RMSE	Bias	Std. Err.	Correct sign	RMSE	Bias	Std. Err.
Factor Income	Bottom 50%	4.0 pp.	82%	2.1 pp.	-0.8 pp.	2.0 pp.	85%	2.1 pp.	-0.9 pp.	1.9 pp.	92%	1.7 pp.	-0.4 pp.	1.7 pp.
	Middle 40%	1.5 pp.	82%	0.9 pp.	-0.3 pp.	0.9 pp.	91%	0.7 pp.	-0.2 pp.	0.7 pp.	83%	1.2 pp.	-1.0 pp.	0.7 pp.
	Next 9%	1.8 pp.	93%	1.1 pp.	-0.7 pp.	0.9 pp.	100%	1.1 pp.	-0.8 pp.	0.8 pp.	83%	1.1 pp.	-0.6 pp.	0.9 pp.
	Top 1%	6.0 pp.	89%	3.4 pp.	-0.2 pp.	3.4 pp.	91%	2.5 pp.	-0.2 pp.	2.5 pp.	92%	3.6 pp.	1.4 pp.	3.3 pp.
Pretax Income	Bottom 50%	3.0 pp.	75%	2.0 pp.	-1.0 pp.	1.7 pp.	76%	2.0 pp.	-1.0 pp.	1.7 pp.	83%	1.8 pp.	-1.3 pp.	1.3 pp.
	Middle 40%	1.5 pp.	80%	1.0 pp.	-0.2 pp.	0.9 pp.	88%	0.8 pp.	-0.2 pp.	0.7 pp.	83%	1.1 pp.	-0.9 pp.	0.7 pp.
	Next 9%	2.3 pp.	91%	1.2 pp.	-0.7 pp.	0.9 pp.	100%	1.1 pp.	-0.8 pp.	0.9 pp.	75%	1.0 pp.	-0.3 pp.	0.9 pp.
	Top 1%	6.2 pp.	89%	3.4 pp.	-0.1 pp.	3.4 pp.	91%	2.6 pp.	-0.2 pp.	2.6 pp.	92%	3.7 pp.	1.4 pp.	3.4 pp.
Disposable Income	Bottom 50%	2.4 pp.	70%	2.5 pp.	-1.6 pp.	1.9 pp.	68%	2.3 pp.	-1.4 pp.	1.8 pp.	75%	3.3 pp.	-2.8 pp.	1.7 pp.
	Middle 40%	1.4 pp.	86%	0.9 pp.	-0.2 pp.	0.9 pp.	91%	0.8 pp.	-0.2 pp.	0.8 pp.	83%	0.8 pp.	-0.5 pp.	0.6 pp.
	Next 9%	2.2 pp.	91%	1.4 pp.	-0.6 pp.	1.3 pp.	88%	1.4 pp.	-0.7 pp.	1.1 pp.	100%	1.1 pp.	0.2 pp.	1.1 pp.
	Top 1%	6.4 pp.	89%	4.0 pp.	0.7 pp.	3.9 pp.	91%	3.2 pp.	0.6 pp.	3.1 pp.	83%	4.5 pp.	2.1 pp.	3.9 pp.
Post-tax Income	Bottom 50%	2.4 pp.	75%	1.9 pp.	-1.2 pp.	1.5 pp.	79%	1.8 pp.	-1.0 pp.	1.5 pp.	75%	2.5 pp.	-2.0 pp.	1.4 pp.
	Middle 40%	1.7 pp.	93%	0.9 pp.	-0.2 pp.	0.8 pp.	97%	0.7 pp.	-0.2 pp.	0.7 pp.	92%	0.7 pp.	-0.5 pp.	0.5 pp.
	Next 9%	2.6 pp.	89%	1.3 pp.	-0.6 pp.	1.1 pp.	97%	1.2 pp.	-0.7 pp.	1.0 pp.	75%	0.9 pp.	0.0 pp.	0.9 pp.
	Top 1%	6.7 pp.	84%	3.9 pp.	0.2 pp.	3.9 pp.	88%	2.9 pp.	0.2 pp.	2.9 pp.	67%	4.1 pp.	1.4 pp.	3.8 pp.
Wealth	Middle 40%	4.9 pp.	82%	1.8 pp.	0.0 pp.	1.8 pp.	88%	1.7 pp.	-0.2 pp.	1.7 pp.	75%	1.4 pp.	0.1 pp.	1.4 pp.
	Next 9%	3.9 pp.	95%	1.2 pp.	-0.2 pp.	1.2 pp.	97%	1.1 pp.	-0.1 pp.	1.0 pp.	100%	1.4 pp.	-0.3 pp.	1.4 pp.
	Top 1%	5.8 pp.	82%	2.8 pp.	-1.5 pp.	2.4 pp.	82%	2.5 pp.	-1.4 pp.	2.1 pp.	75%	2.9 pp.	-1.4 pp.	2.5 pp.

Notes: This table reports statistics for goodness of fit and noise of our 1-year ahead real income and real wealth growth predictions. “Std. dev.” is the standard deviation of observed 1-year growth. “Correct sign” is the fraction of years in which we correctly predict whether income (or wealth) is growing or falling. “Bias” is the mean difference between predicted and observed 1-year growth, “Std. Err” is the standard error of this mean, and ”RMSE” is the root mean square error capturing total error ($RMSE^2 = bias^2 + std.\ err.^2$). “All years” includes 44 observations (growth relative to the preceding year in 1976, 1977, ..., 2019). “All years excluding tax reforms” includes 34 observations (it excludes 1987, 1988, 1991, 1992, 1993, 2001, 2003, 2012, 2013). “Recessions” includes 12 observations, corresponding to recession years (1980, 1981, 1982, 1990, 1991, 2001, 2008, 2009) and their immediate aftermath (1983, 1992, 2002, 2010).

Appendix (for Online Publication)

A Link Between NIPA National Income Components and DINA Concepts

Our monthly micro-files distribute BEA's high-frequency national income accounts, starting from the annual distributional national accounts micro-files of Piketty, Saez and Zucman (2018). These files are based on internationally harmonized guidelines (Blanchet et al., 2021), which themselves are based on the UN System of National Accounts and definitions of income components that maximize consistency with components of household wealth. The concepts used by the Bureau of Economic Analysis for the US national accounts are largely consistent with the System of National Accounts, but sometimes slightly differ. To clarify how the main variable in our micro-files relate to the headline aggregates of the official US national accounts, this Section provides a mapping of the main components of national income as published by BEA (henceforth NIPA) into the main components of factor national income in our distributional national accounts micro-files (henceforth DINA).

Recall that national income as published by BEA (NIPA Table 1.12) is decomposed into compensation of employees, proprietors' income, rental income of persons, corporate profits, net interest and misc. payments, taxes on production and imports less subsidies, net business transfer payments, and current surplus of government enterprises. The main differences with DINA are the following:

- In DINA, business transfer payments are allocated to corporate profits (for corporate transfers), to proprietors' income (for non-corporate businesses' transfers) and to rental income (for housing transfers).
- in DINA, the small current surplus of government enterprises is treated as a tax on production.
- In DINA, property taxes (business and real estate) are not treated as taxes on production but as direct taxes (like wealth taxes would be), hence allocated to corporate profits (for property taxes paid by corporations), proprietors' income (for property taxes paid by non-corporate businesses), and rental income (for residential property taxes).
- In the NIPAs, there are various imputations of interest income (e.g., dividends received by life-insurance companies; notional interest on underfunded pension plans) that are re-classified in DINA for consistency with household wealth.

As a result of these and other minor other reclassifications to improve consistency with household wealth aggregates, NIPA national income concepts map onto DINA factor income concepts as follows (NIPA Table numbers and DINA variable names in parenthesis):[36]

[36] The mapping for compensation of employees, corporate profits, net interest, and taxes on production and imports less subsidies is exact. The mapping for proprietors' income and rental income is almost exact (with a discrepancy of less than 0.1% of national income).

NIPA compensation of employees (Table 1.12 line 2) = DINA compensation of employees (flemp)

NIPA proprietors' income (Table 1.12 line 9) = DINA business asset income (fkbus)
+ DINA labor component of mixed income (flmil)
– DINA business property taxes allocated to non-corporate businesses (non-corporate business share of propbustax)
+ NIPA rental income included in proprietors' income (Table 7.4.5 line 20)
– NIPA net non-corporate business transfers paid (Table 1.12 line 21 – Table 1.14 line 10 – Table 7.4.5 line 19)
– NIPA royalties (Table 7.9 line 7)

NIPA rental income of persons (Table 1.12 line 12) = DINA housing asset income (fkhou)
- NIPA residential property taxes (Table 7.4.5 line 15 = proprestax)
- NIPA mortgage interest payments (Table 7.4.5 line 18)
- NIPA housing net current transfer payments (Table 7.4.5 line 19)
- NIPA rental income included in proprietors' income (Table 7.4.5 line 20)
+ NIPA tenant-occupied rental income of nonprofits (Table 7.9 line 14)
+ NIPA royalties (Table 7.9 line 7)

NIPA corporate profits (Table 1.12 line 13) = DINA equity asset income (fkequ)
+ DINA equity income earned through pension plans (equity share of fkpen)
– DINA business property taxes allocated to corporations (corporate share of propbustax)
+ NIPA dividends received by government (Table 3.1 line 14)
+ NIPA dividends received by nonprofits (Table 2.9 line 51)
– NIPA net corporate business transfers paid (Table 1.14 line 10)
– NIPA imputed interest paid by corporations on underfunded pension plans (Table 7.12 line 192)
– NIPA dividend receipts of life-insurance companies included under "imputed interest received from life-insurance carriers" (part of Table 7.11 line 68, not separately reported).

NIPA net interest and misc. payments (Table 1.12 line 18) = DINA currency, deposits, and bond income (fkfix)
+ DINA interest income earned through pension plans (interest share of fkpen)
– DINA non-mortgage interest paid (fknmo)
+ NIPA misc. corporate payments (Table 1.14 line 9 – Table 7.11 line 101)
+ NIPA imputed interest paid by corporations on underfunded pension plans (Table 7.12 line 192)
+ NIPA dividend receipts of life-insurance companies included under "imputed interest received from life-insurance carriers" (part of Table 7.11 line 68, not separately reported)
+ NIPA interest received by nonprofits (Table 2.9 line 50)
– NIPA net interest paid by government, other than imputed for unfunded pension plans (Table 7.11 line 107 - line 50)

NIPA taxes on production and imports less subsidies (Table 1.12 line 19 – line 20) = DINA

sales and excise taxes (fkprk + flprl)
+ DINA residential property taxes (proprestax)
+ DINA business property taxes (propbustax)
– DINA subsidies on production and imports (fksubk + flsubl)
– NIPA current surplus of government enterprises (Table 1.12 line 25)

B Real-Time Wealth Distributions

To construct quarterly wealth distributions, we start from the Piketty, Saez and Zucman (2018) micro-files, last updated in February 2022,[37] and re-scale the main components of household wealth to their end-of-quarter value, using the latest quarterly release of the Financial Accounts.

The wealth components we consider are, on the asset side, tenant-occupied housing, owner-occupied housing, S-corporation equity, C-corporation equity, equity in non-corporate businesses, fixed-income assets, pension assets; and on the liability side, tenant-occupied mortgages, owner-occupied mortgages, and non-mortgage debt. The aggregate value of all of these components are published quarterly by the Federal Reserve in the Financial Accounts of the United States, around 70 days after the end of each quarter. Following Saez and Zucman (2016), our estimates exclude unfunded pensions (such as promises of future Social Security benefits), consumer durables (which are not assets in the System of National Accounts and thus excluded from wealth in other countries; see United Nations, 2009), and the assets and liabilities of non-profit institutions such as private foundations.

We then assume that distributions are stable within each of these components from one quarter to the other and compute the implied distribution of wealth. We use real-time *Forbes* data to adjust the wealth of the top 400 tax units so that it matches the *Forbes* estimate at the end of each quarter. Thus our estimates of wealth inequality by construction match *Forbes* at the very top, like the annual estimates constructed in Saez and Zucman (2020b) and the SCF-year estimates of Batty et al. (2019). The Forbes estimates contain valuable information on high-frequency wealth dynamics at the top-end because they combine public information on ownership of stock in listed companies (from mandatory Securities and Exchange Commission filings) with daily stock price changes for these companies. Close to half of the wealth of the *Forbes* 400 is in public equity in recent years. The limitations of annual *Forbes* estimates (lack of public information on diversified portfolios and on debts, imperfect information on the value of private businesses) carry over to the real-time *Forbes* estimates.

Additionally, we construct monthly and daily wealth distributions as follows. Starting from the quarterly wealth totals by components described above, we use housing and equity price indices to update total housing wealth and total equity wealth at the monthly frequency. We do so using the Denton (1971) method for available quarters, and extrapolating using the indexes' growth rate in the most recent months, before the last quarter becomes available. The stock market index that we use is the Wilshire 5000, a comprehensive index of the stock performance of publicly traded U.S. firms. For housing prices, we use the Case-Shiller index, extrapolated

[37] The micro-files are updated annually. A description of each update is available at `http://gabriel-zucman.eu/usdina`, as are current micro-files, computer code, and tabulations of key findings. All vintage releases and corresponding code are also published at this address.

using the Zillow house price index in the most recent months. For daily estimates, we simply update total equity (including S-corporations equity) using the daily Wilshire 5000. For both monthly and daily estimates, we keep distributions constant within components and rescale the wealth of the top 400 tax units to match real-time *Forbes* numbers.

C Adjustment of Employment Status and Wages

C.1 Adjustment of Employment and UI Recipients

In our monthly microfiles, we adjust at the margin whether someone (i) is employed and (ii) receives UI benefits based on the information described in Section 4.2, namely (i) Bureau of Labor Statistics (BLS) monthly releases of non-farm employment at the national level, (ii) the Department of Labor's weekly unemployment claims statistics, and (iii) labor force status by race $\times$ education $\times$ gender $\times$ 5-year age group $\times$ marital status cell in the CPS. Here we provide details on our adjustment procedure.

Consider an annual microfile for date t, which was constructed as a weighted average of year y_1 (with wealth $1-\lambda$) and y_2 (with weight λ, with $y_2 = 2019$ and $\lambda = 1$ for the most recent years) to represent the average distribution over the 12-month period $\{t_{\min}, \dots, t_{\max}\}$. Let us use the letter i to denote a given cell of race $\times$ education $\times$ gender $\times$ 5-year age group $\times$ marital status. Let $x_{it} \in [0,1]$ be the employment rate of cell i, and let $n_{it} \in [0,1]$ be the relative size of cell i, and let $x_t \equiv \sum_i n_{it} x_{it}$ be the aggregate employment rate in the microfile. In the CPS-based employment series constructed above, let $y_{it} \in [0,1]$ be the employment rate of cell i at time t. From the BLS-based aggregate employment series, let z_t be the aggregate employment rate at time t.

Let $y_i \equiv \frac{1}{12} \sum_{t=t_{\min}}^{t_{\max}} y_{it}$ be the average employment rate from the CPS series over the period notionally covered by the synthetic microfile. Let $\Delta y_{it} \equiv \frac{y_{it} - y_i}{y_i(1-y_i)}$ be the relative difference between in the series' values at date t and their average over $\{t_{\min}, \dots, t_{\max}\}$.[38] We calculate the new employment rate for cell i in the microfile as $x_{it} + x_{it}(1-x_{it})(\Delta y_{it} + z_t - x_t - \gamma)$ where $\gamma = \frac{\sum_i n_{it} x_{it}(1-x_{it}) \Delta y_{it}}{\sum_i n_{it} x_{it}(1-x_{it})}$ is a renormalization term that ensures that we match the aggregate employment rate. This method has the desirable property of reproducing the relative changes in employment between cells while matching aggregate employment overall and without modifying employment within a given cell unless the CPS-based warrants such a change.[39]

Using this newly calculated employment rate, we adjust the number of employed individuals within each cell as follows. If the employment rate has increased, we randomly select non-employed units and mark them as employed to match the new rate. Conversely, if the employment rate has decreased, we randomly select employed units and mark them as non-employed. If employment has not changed, we do nothing. We give a wage to all the observations we marked as employed, using the procedure described in Section 4.3.

We follow the exact same procedure as above, using the unemployment rates to adjust the

[38] Note that we normalize by a factor $y_i(1-y_i)$ and not y_i to get relative changes since we are working with ratios between 0 and 1.

[39] In rare cases where the adjusted employment rate is below 0 or above 1, we truncate it and repeat the procedure until the constraint is satisfied.

number of UI recipients. When we need to increase the number of UI recipients within a cell, we select in priority individuals who are not employed.

C.2 Adjustment of Wage Income

To attribute a wage to every individual marked as employed in our monthly microfiles, we combine (i) the marginal wage income distribution estimated by averaging QCEW and CPS predictions (as described in Section 4.3) with (ii) the rank of each individual in the wage distribution, which we adjust as explained below.

Estimation of the average wage rank by individual characteristic. We estimate the average wage rank by race × education × gender × 5-year age group × marital status in the CPS. Since the monthly wage data is top-coded in the CPS, we use an interval-censored regression model. In each monthly CPS file, we estimate the wage rank of each observation (or, for top-coded observations, the lower bound of the rank). We transform these ranks using the logistic function and then run an interval-censored regression of the rank against race, education, 5-year age groups, and marital status interacted with gender, and assuming normal residuals. We use the prediction from that regression to construct monthly series of the average wage rank for each cell. We correct these series for seasonal variations using the X11 procedure.

Adjustment of wage income in the microfiles. In the monthly microfiles, (i) we adjust each observation's wage rank at the margin using the relative variations in wage rank from the CPS, and then (ii) we attribute the wage income corresponding to their rank.

Consider an annual microfile for date t, which was constructed as a weighted average of year y_1 (with wealth $1 - \lambda$) and y_2 (with weight λ, with $y_2 = 2019$ and $\lambda = 1$ for the most recent years) to represent the average distribution over the 12-month period $\{t_{\min}, \dots, t_{\max}\}$. For individuals with a positive wage, define x_{ikt} the wage rank of individual k in cell i at date t. In the CPS-based wage rank series constructed above, let $y_{it} \in [0, 1]$ be the average wage rank of cell i at time t.

Let $y_i \equiv \frac{1}{12} \sum_{t=t_{\min}}^{t_{\max}} y_{it}$ be the average wage rank from the CPS series over the period notionally covered by the synthetic microfile. Define $\Delta y_{it} \equiv \frac{y_{it} - y_i}{y_i(1-y_i)}$ be the relative difference between in the series' values at date t and their average over $\{t_{\min}, \dots, t_{\max}\}$.[40]

We calculate the a new rank for observation i, k as $x_{ikt} + x_{ikt}(1 - x_{ikt})\Delta y_{it}$. If an observation does not have an initial rank because it was previously non-employed, we directly give it the average wage rank of its cell.[41] Then we interpolate the monthly wage distribution estimated by percentile using the generalized Pareto interpolation (Blanchet et al., 2022) and attribute to each observation the wage income that corresponds to its rank. We adjust ranks similarly in the UI benefits distribution and distribute UI benefits by keeping their marginal distribution identical.

[40]Note that we normalize by a factor $y_i(1 - y_i)$ and not y_i to get relative changes since we are working with ranks between 0 and 1.

[41]We sort observations according to the new rank, and re-compute the rank from this order, to ensure that the distribution of ranks is uniform.

D Structure of Programs and Files

The section describes the overall structure of the program files used in this paper. The programs (as well as a more detailed and up-to-date version of the following description and instructions) are available online at `https://github.com/thomasblanchet/real-time-inequality`.

D.1 Description of programs/code

- The folder `raw-data` contains the raw input data, primarily in cases where direct download/scraping is not possible or not justified, or in cases where data files are heavy (like the QCEW) and therefore downloading them over the internet every time is not desirable.
- The folder `work-data` contains intermediary data files that are produced by the code. It is divided into subfolders corresponding to each code file, and no intermediary data file is may be changed by two distinct code files.
- The folder `graphs` contains the all the figures (and a few tables) generated by the code. It is divided into subfolders corresponding to each code file.
- The folder `programs` contains the codes (except those performing the optimal transport).
  - The codes named `programs/01-*` handle the retrieval of the raw data, either directly from the internet or from the folder `raw-data`.
  - The codes named `programs/02-*` handle preliminary treatments of the data.
  - The codes named `programs/03-*` produce the synthetic microfiles and related outputs.
  - The codes named `programs/04-*` produce the figures and tables used for the analysis.
- The folder `transport` contains the code and data specifically related to the optimal transport: it is meant to run separately from the main code on a computing cluster.

D.2 License for Code

The code is licensed under the Modified BSD License.

D.3 Instructions for Replication

- Edit the `$root` global in `programs/00-setup.do` to correspond to the project's directory.
- Run the file `programs/00-run.do`.
- To also run the transport, run programs until `programs/02-export-transport-dina.do` and then execute the Python code under `transport/transport.py` preferably using Slurm and the Shell script `transport/transport.sh`. Then resume the execution of `programs/00-run.do`.

Programs are organized as follows:

- `programs/01-*`
  - The codes retrieve the data from the internet directly to the extent that it is possible.
  - Unless there has been changes in the structure of the data, they should run without any change for each update.
  - In some cases, the data needs to be manually updated in the `raw-data` folder at each update.
  - Instruction for each file in included in `00-run.do`.
- `programs/02-*`
  - The codes primarily generate data in the `work-data` folder that is used to generate the synthetic microfiles.
- `transport`
  - This folder includes the data and code necessary for the optimal transport.
  - These codes are meant to run on the computing cluster.
  - They do not need to be updated every time (only when new tax microdata is available).
  - The CSV data files included in this folder are produced by the codes before.
- `programs/03-*`
  - The codes in that folder produce the synthetic microfiles, including backtesting versions of the microfiles that use older tax data, and rescaling versions that only use information on macro aggregates.
  - The globals `$date_begin` and `$date_end` at the beginning of these files can be used to generate only the files for specific months. This can be useful since not all the files need to be constructed for every update.
  - Codes in that section also produce the aggregated version of the database by group that is used for the website `http://realtimeinequality.org/`. These files are stored in the folder `website`.
- `programs/04-*`
  - Use the microfiles and related outputs to create the tables and figures included in the paper.

D.4 Description of Microfiles and Codebook

One output of this paper is a set of harmonized monthly micro-files in which an observation is a synthetic adult (obtained by statistically matching public micro-data) and variables include income, wealth, and their components. These variables add up to their respective national accounts totals and their distributions are consistent with those observed in the raw input data. With these micro-files one can reproduce all the results of the paper and the statistics shown at `https://realtimeinequality.org` exactly. The files are available here.[42]

There is one file per month starting in January 1976. The files are at the adult individual (aged 20 and above) level, so the sum of weights (variable `weight`) adds up to the total adult population, 248.9 million in May 2022. The variable `id` identifies a household (as defined by the tax code, i.e., either a single person aged 20 or above or a married couple, in both cases with children dependents if any). Income is individualized. To compute our benchmark equal-split adult statistics, compute average income or wealth by `id`. To compute statistics at the household level, take the sum of income or wealth by `id` and the average weight by `id`.

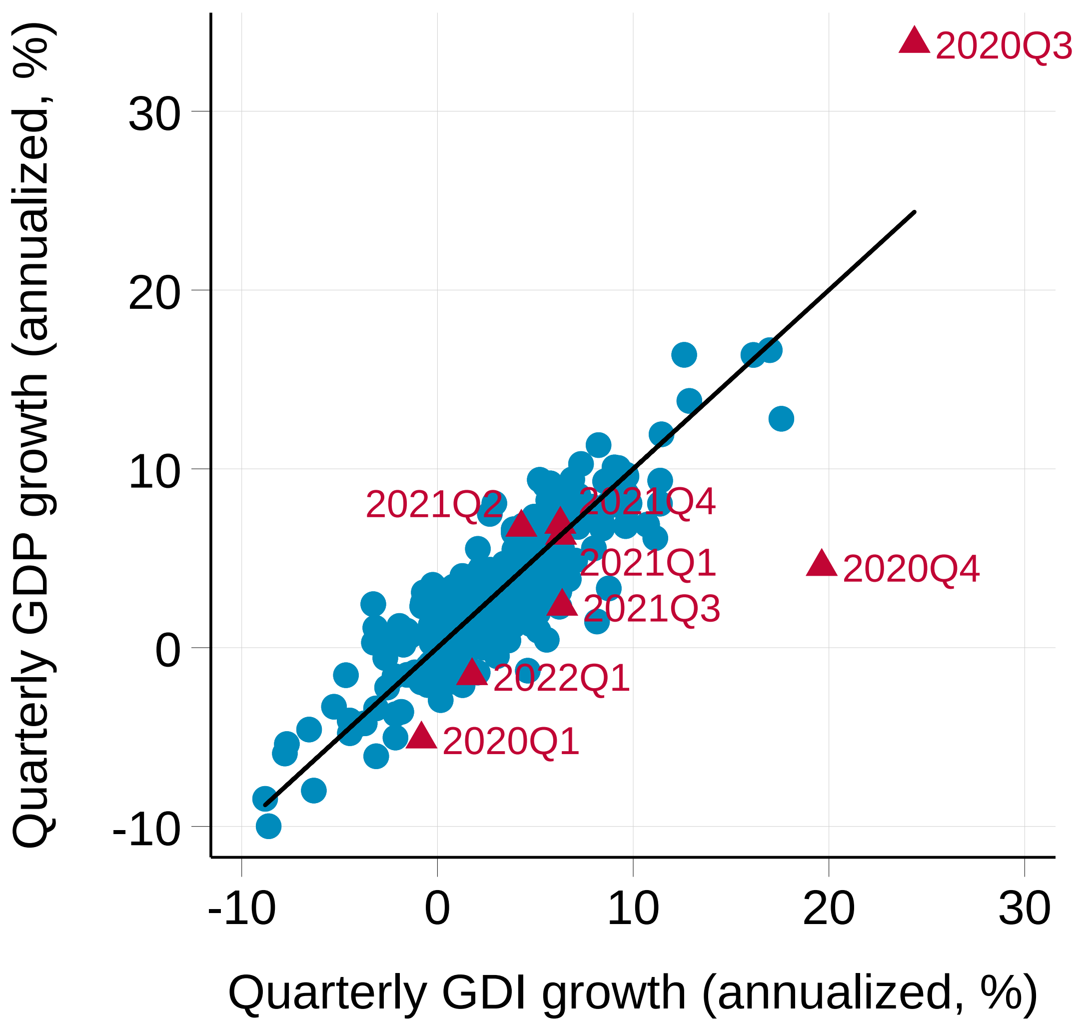
The files include socio-demographic information: `age`, `sex`, marital status (`married`), `race`, educational attainment (`educ`), and the following income and wealth variables:

- `princ`: factor national income
- `peinc`: pretax national income
- `poinc`: posttax national income
- `dispo`: disposable income
- `flemp`: compensation of employees
- `proprietors`: proprietors' income
- `rental`: rental income
- `profits`: after-tax corporate profits
- `corptax`: corporate income tax
- `fkfix`: interest income
- `govin`: government interest income
- `fknmo`: non-mortgage interest payments
- `prodtax`: production taxes
- `prodsub`: production subsidies
- `contrib`: contributions to pensions, disability insurance, & unemployment
- `penben`: pension and disability insurance benefits
- `uiben`: unemployment insurance benefits
- `taxes`: current taxes on income and wealth
- `estatetax`: estate tax
- `othercontrib`: contributions for government social insurance other than pension, unemployment, and disability
- `govcontrib`: total contributions for government social insurance
- `vet`: veterans' benefits
- `othcash`: other cash benefits
- `medicare`: Medicare spending

[42]Full link in case hyperlinks break: `https://www.dropbox.com/home/SaezZucman2014/RealTime/repository/real-time-inequality/work-data/03-build-monthly-microfiles/microfiles`.

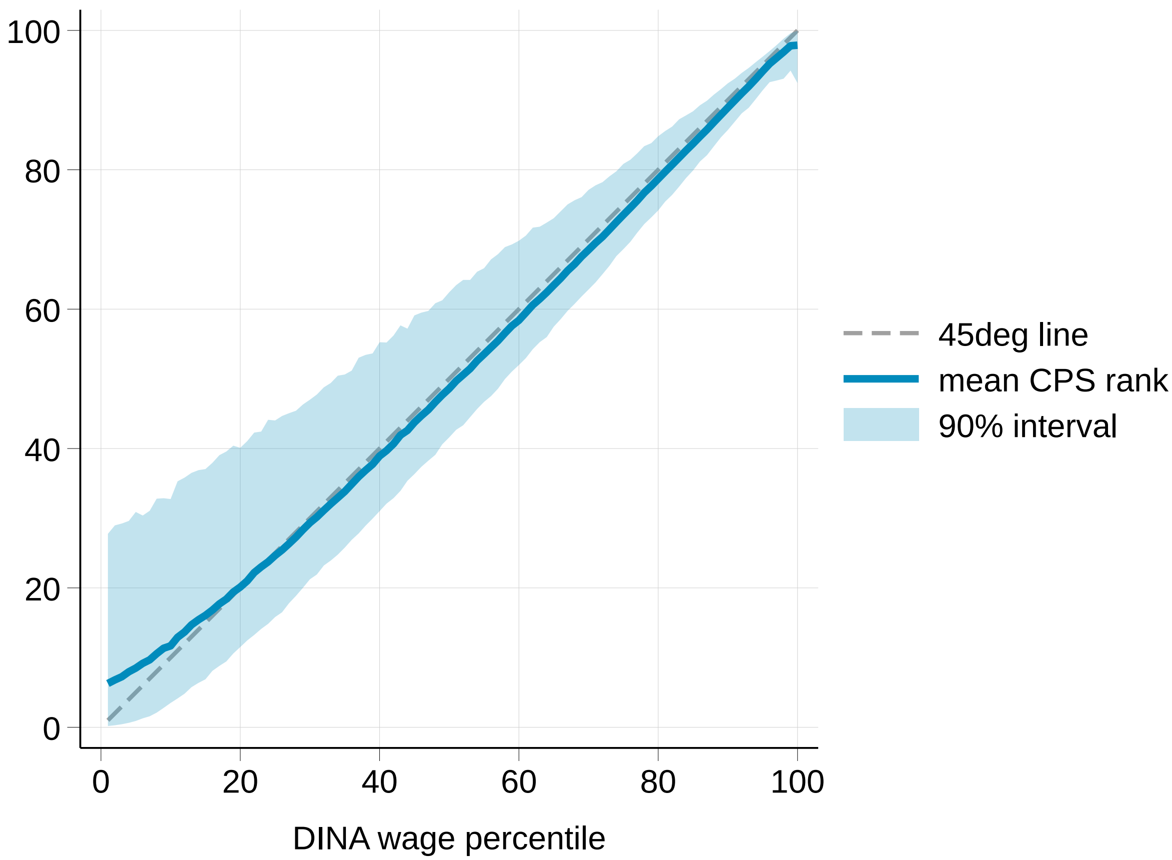
- `medicaid`: Medicaid spending
- `otherkin`: other in-kind government spending
- `colexp`: collective consumption expenditure
- `covidrelief`: covid-19 economic impact payments
- `covidsub`: paycheck protection program
- `surplus`: surplus/deficit of government and private social insurance
- `surplus_ss`: surplus/deficit of government social insurance
- `prisupenprivate`: surplus/deficit of private pension system
- `prisupgov`: primary surplus/deficit of government
- `hweal`: net wealth
- `housing_tenant`: tenant-occupied housing wealth
- `housing_owner`: owner-occupied housing wealth
- `equ_scorp`: equity in S-corporations
- `equ_nscorp`: equity in corporations other than S-corporations
- `business`: equity in non-corporate businesses
- `pensions`: pension wealth
- `fixed`: currency, deposits, and interest-bearing assets
- `mortgage_owner`: mortgages on owner-occupied housing
- `mortgage_tenant`: mortgages on tenant-occupied housing
- `nonmortage`: non-mortgage debt
- `top400`: dummy if observation is in the top 400 of households by wealth
- `acs`: dummy if observation lives in group quarters (nursing homes, dormitories, etc.)

Figure A1: GDP vs. GDI Growth



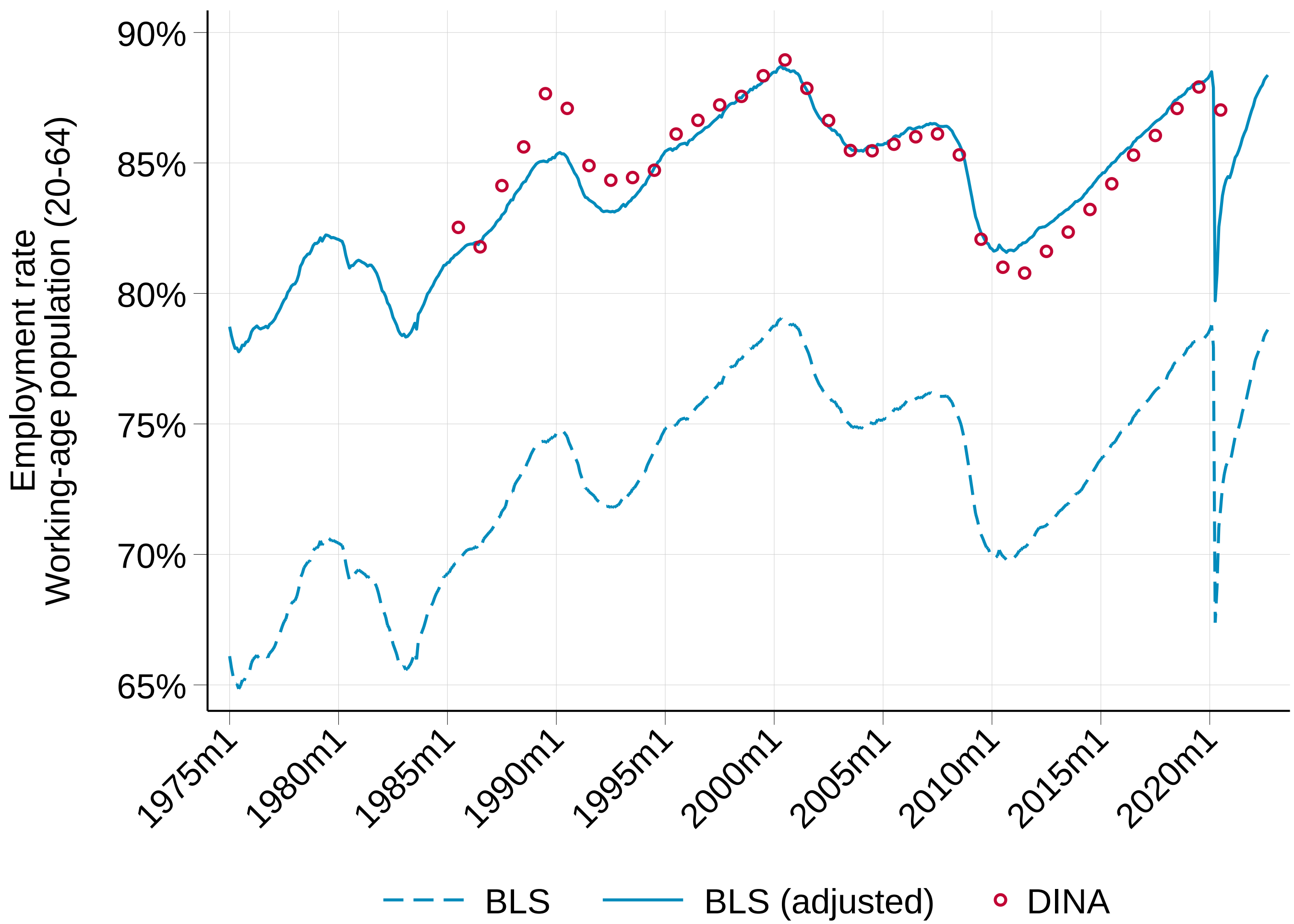
Notes: This figure compares the quarterly growth of real annualized GDP to the quarterly growth of real annualized GDI from 1947Q2 to 2022Q1. Observations corresponding to the Covid-19 recession and recovery are shown in red triangles. The point corresponding to 2020Q2 (an outlier with around -35% annualized GDP growth and -35% annualized GDI growth) is omitted. During the recovery from Covid-19, GDI has recovered faster than GDP.

Figure A2: Comparing Wage Earnings Ranks in the One-to-One Statistical Match



Notes: This figure compares the ranks in wage earnings between the wage earnings in the DINA data and the wage earnings in the CPS data in the one-to-one statistically matched dataset. The x-axis is the DINA wage earnings percentile and y-axis is the average percentile in the CPS data. The figure show averages over 1976–2019. The data refers to household earnings and the sample is limited to households that earn at least a full-time Federal minimum wage. The figure shows that ranks in wage earnings are very close implying that the one-to-one statistical match respects ranks well along the crucial wage earnings variable.

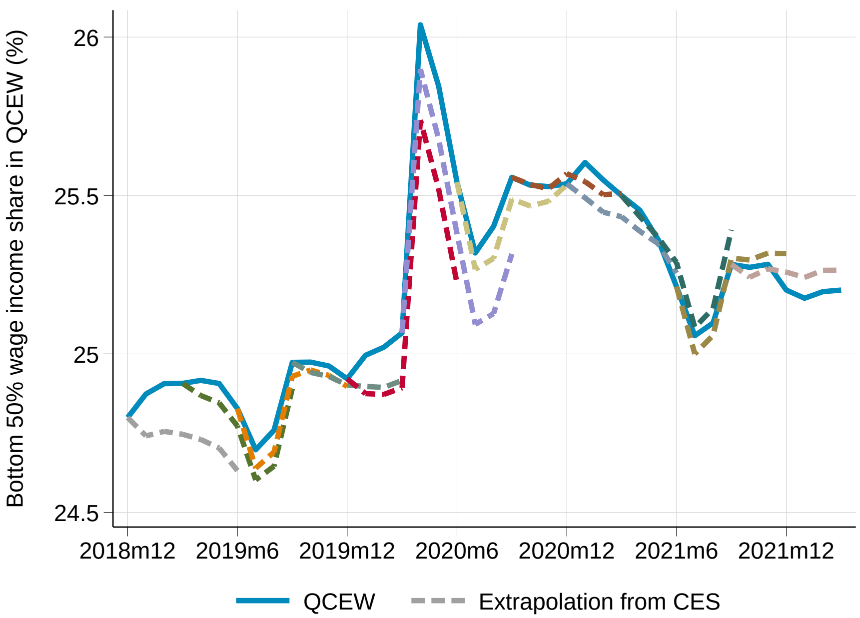
Figure A3: Monthly vs. Annual Employment Rates



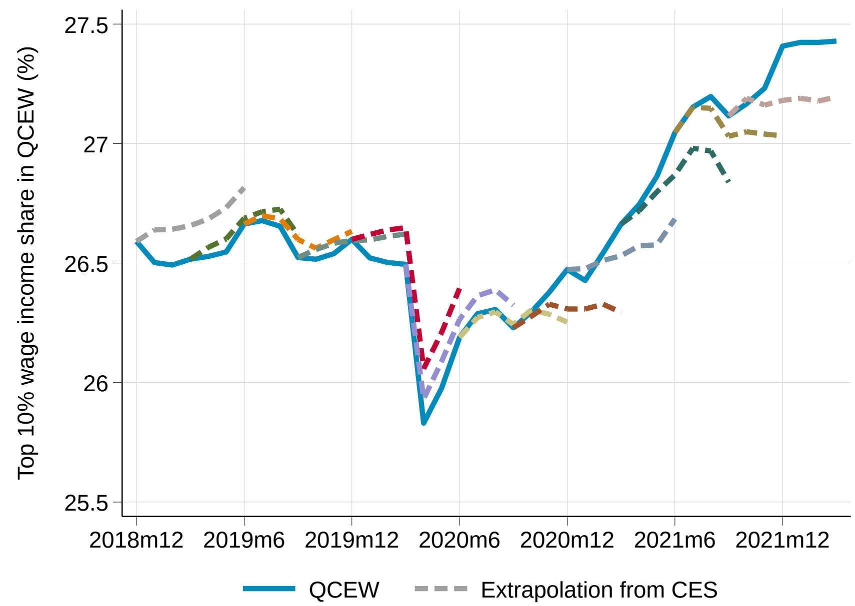
Notes: This figure depicts, for the working age population 20-64, the raw monthly employment rates from the BLS based on the monthly CPS (in dashed blue line), the annual employment rates from the Social Security tax data (in red circles), and the monthly employment rates adjusted to match annual rates (in solid blue line). Employment rates in our monthly data is given by the blue solid series. Quantitatively, in recent decades (when the catching-up female labor force participation trend stabilizes), raw monthly employment rates are around 75% while adjusted monthly employment rates (that track annual levels) are 10 points higher around 85% because of part-year workers.

Figure A4: Capturing Most Recent Months Using CES instead of QCEW

(a) Bottom 50% Wage Share



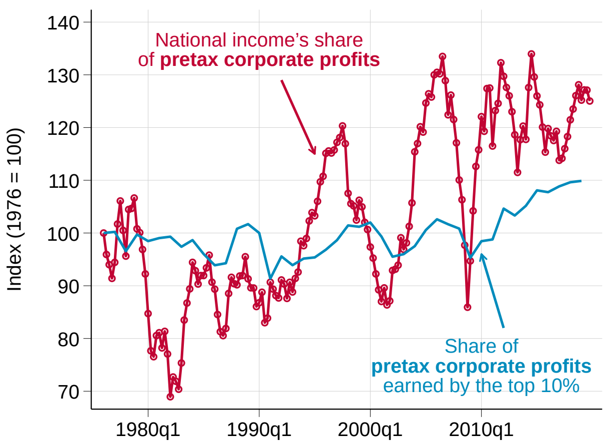
(b) Top 10% Wage Share



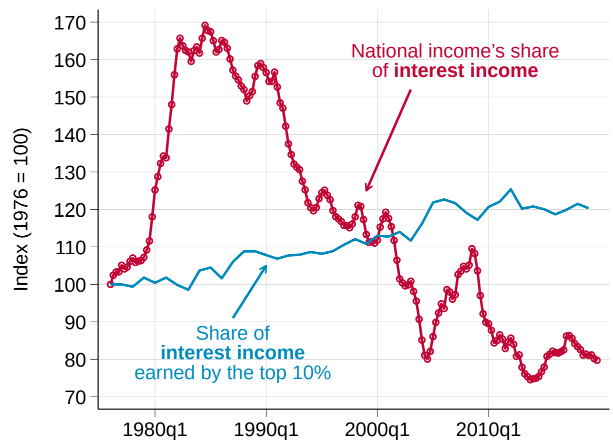
Notes: This figure depicts the quality of the projection using the monthly Current Employment Statistics (CES) data to extrapolate the QCEW data for the last 2 quarters (when the QCEW is not yet available but the CES is). The blue series depicts the bottom 50% wage share in panel (a) and the top 10% wage share in panel (b) using the raw (de-seasonalized) QCEW data. For each quarter, the figure also depicts in dashed line the projected shares using the CES data to extrapolate the QCEW data up to 6 months out. By definition, the projection matches the QCEW level in the last month of the quarter, and then projects out over the next two quarters. To carry out the projection, we match each QCEW cell to three CES series. The first matches the location of the QCEW cell as precisely as possible, the second matches it at the state level, and the third at the national level. We average the trend from these three series in each cell and use this average trend to extend the QCEW data in the most recent quarters. Overall, the CES projection comes pretty close to the QCEW. The fit is better for the bottom 50% share than for the top 10% wage share.

Figure A5: Capital Income Volatility

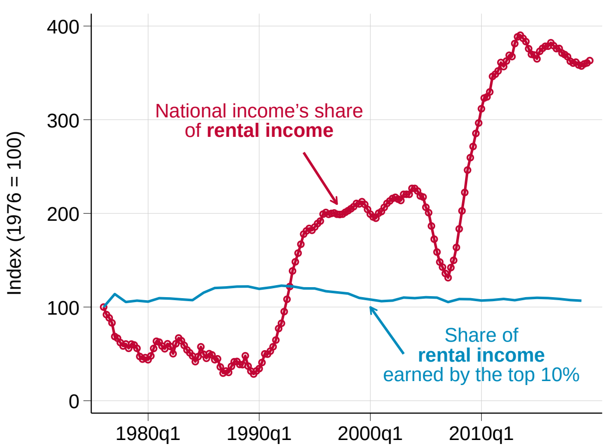
(a) Corporate Profits



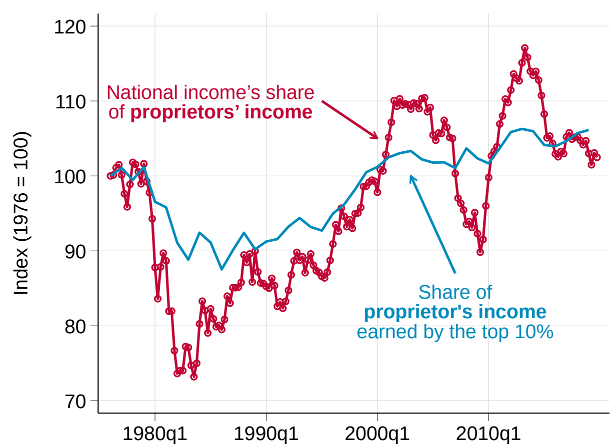
(b) Interest



(c) Rental Income



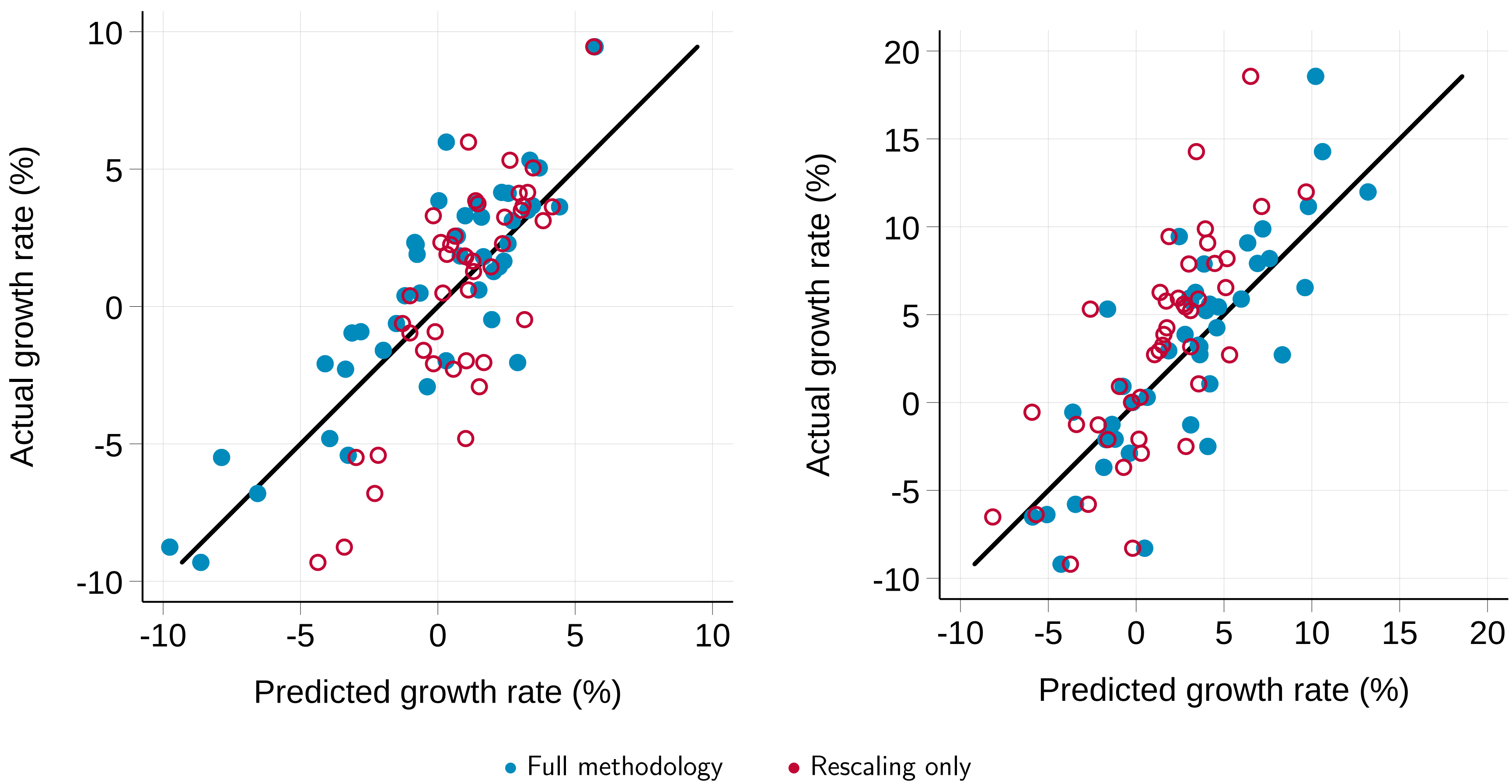
(d) Proprietor's Income



Notes: This figure compares the volatility of the size of capital income components (corporate profits, interest, rental income, proprietor's income), as measured by their share of national income, and the volatility of their concentration, as measured by the share of each component going to the top 10% of the pretax income distribution. For example the top left panel shows the share of pretax corporate profits in national income and the share of pretax corporate profits accruing to the 10% of adults at the top of the pretax income distribution, from 1976 to 2019. All series are depicted relative to a base 100 in 1976 to compare volatility. The figure shows that the size of capital income components can be highly volatile at high-frequency while their concentration is slow-moving. This lends support to our methodology which captures high-frequency changes in aggregate capital income and assumes that their concentration is unchanged in the short run.

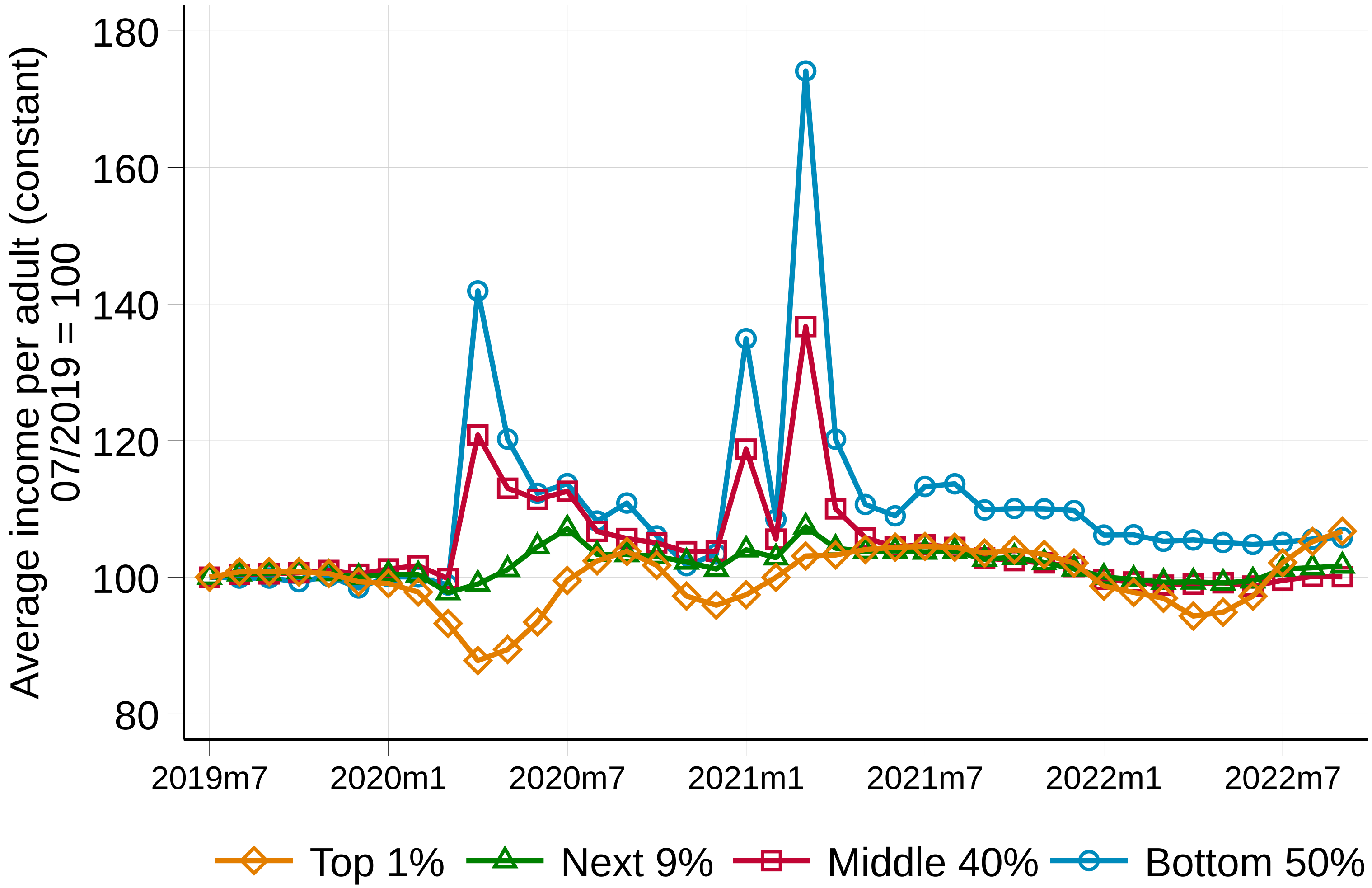
Figure A6: Predicted Growth: Our Method vs. Simplified Macro Method

(a) Bottom 50% (b) Top 1%



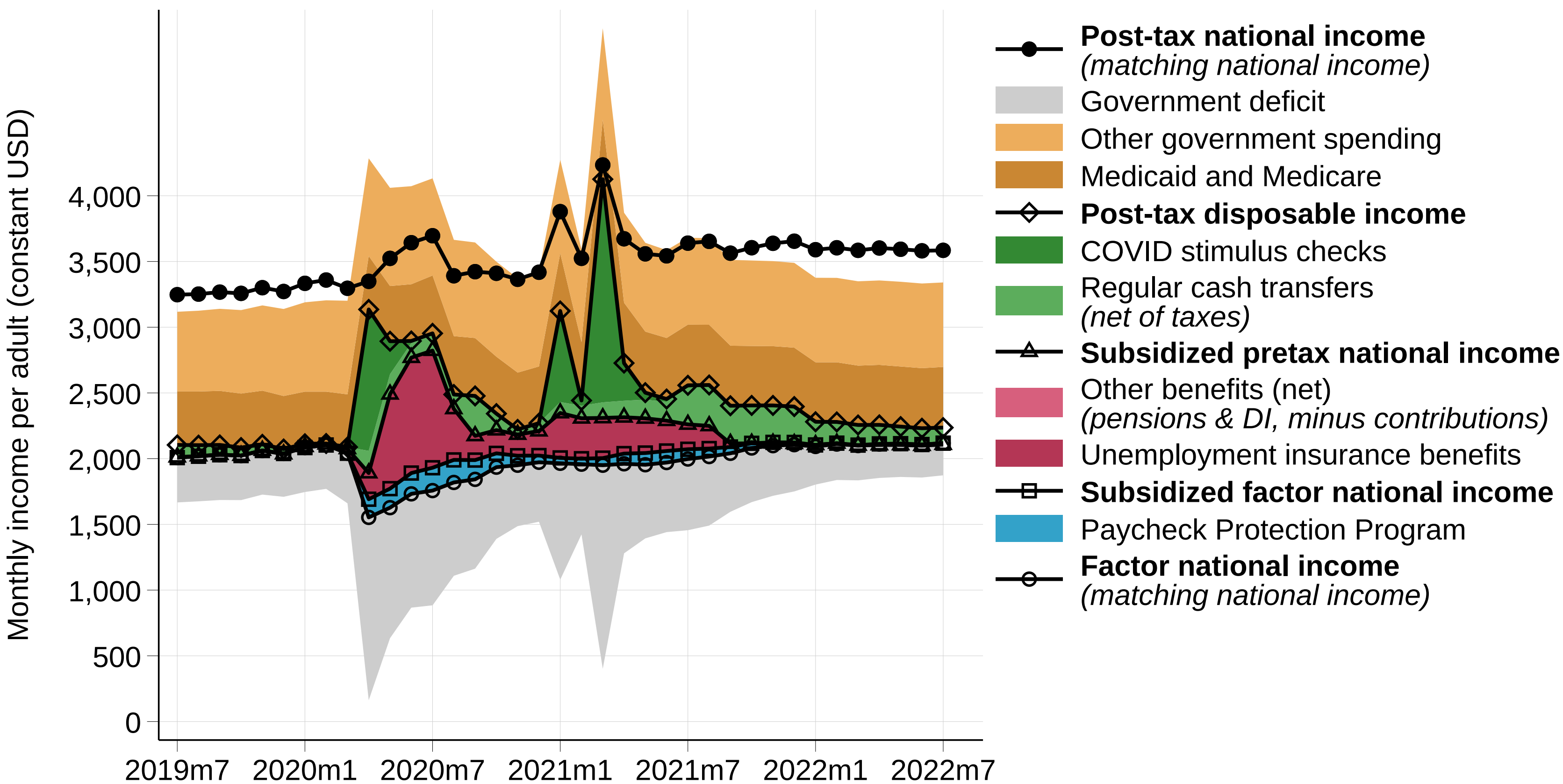
Notes: This figure compares the quality of our baseline estimates with the quality of simpler estimates that only rescales macroeconomic aggregates without adjusting within-component distributions. This figure depicts predicted to actual growth in average real factor income per adult (with income equally split among married spouses) from year t to $t+1$ for the bottom 50% (left panel) and the top 1% (right panel) for each year from 1976 to 2019. Actual growth is obtained using the annual distributional national account micro-data for both years t and $t+1$. Predicted growth is obtained using the annual micro-data for year t but the projected micro-data using our full methodology for $t+1$ (blue full dots) or a simplified methodology that only rescales macroeconomic aggregates without adjusting within-component distributions (red empty dots). The simplified methodology performs worse, especially in recessions years for the bottom 50%, showing that adjusting distributions is critical to accurately project real-time inequality during recessions.

Figure A7: Real Disposable Income Around the Covid-19 Pandemic



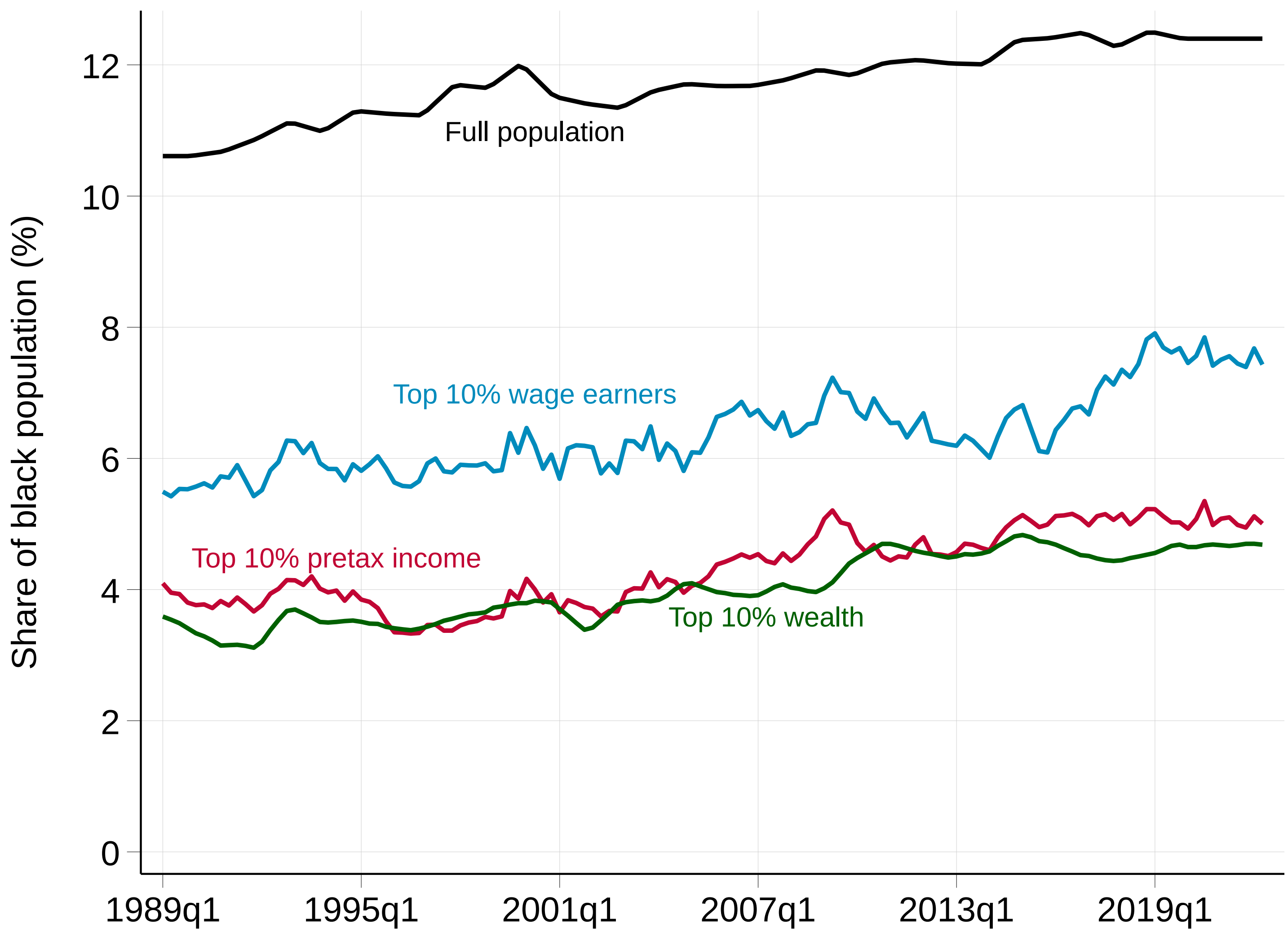
Notes: This figure shows the monthly evolution of real disposable income per adult from July 2019 to March 2022 in the full adult population (not restricting to working-age adults). Individual adults are ranked by their factor income, and income is equally split between married spouses. The figure shows that for the bottom 50% of the factor income distribution, monthly disposable income was nearly twice as large in March 2021 as in July 2019 (as a result of the third wave of Economic Impact Payments). By the spring of 2022, disposable income had returned to its pre-Covid level for all groups except for the bottom 50% for which it was about 15% higher.

Figure A8: Income of the Bottom 50% During the Covid Crisis



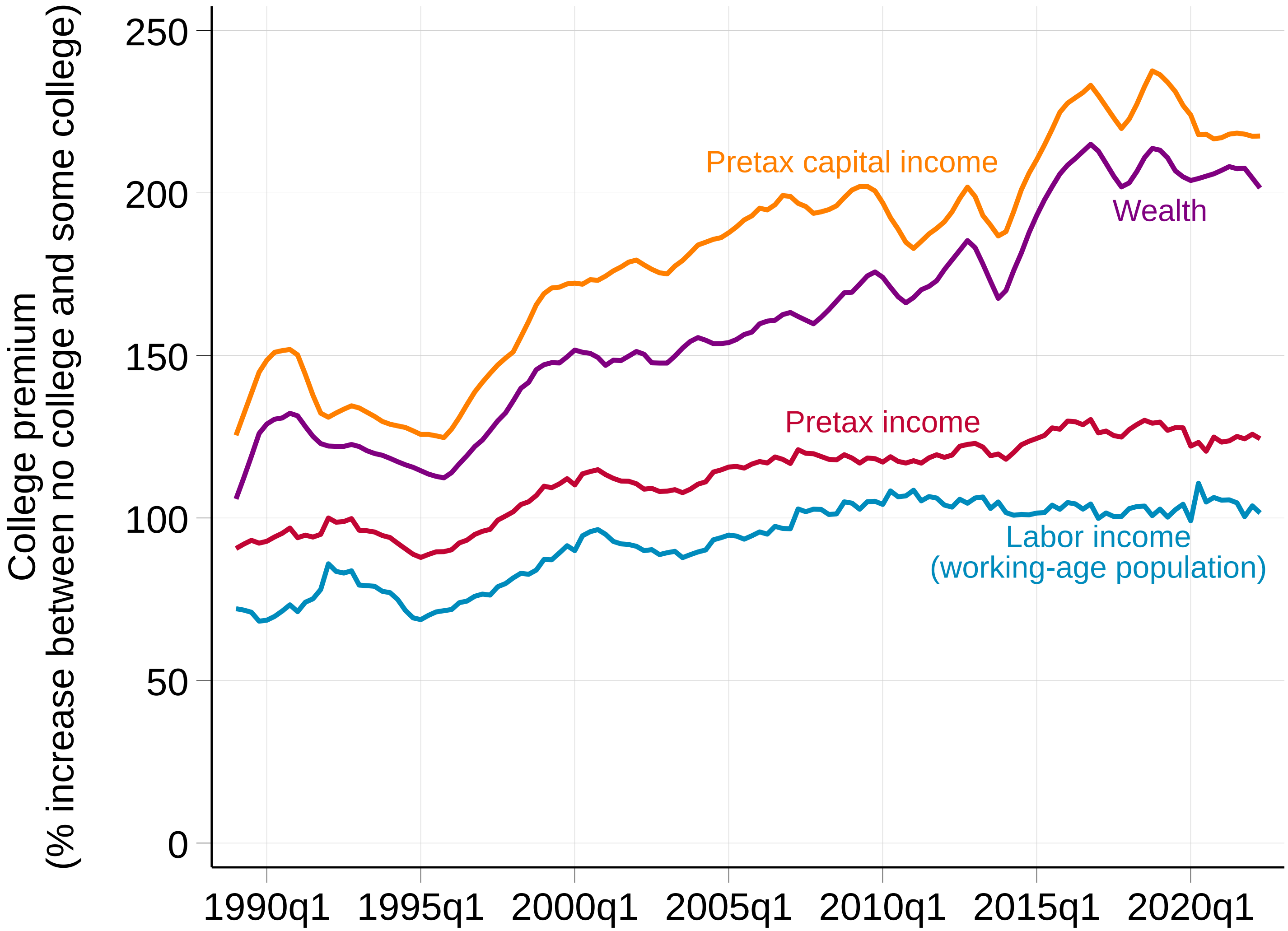
Notes: This figure decomposes the average real monthly posttax national income of the bottom 50% of the working-age population ranked by factor income from July 2019 to May 2022. This is the same Figure as Figure 9 but adding Medicaid and Medicare, other government spending, and the government deficit, so as to go all the way to posttax national income. See notes to Figure 9.

Figure A9: Share of Black adults in the Full Population vs. Top 10%



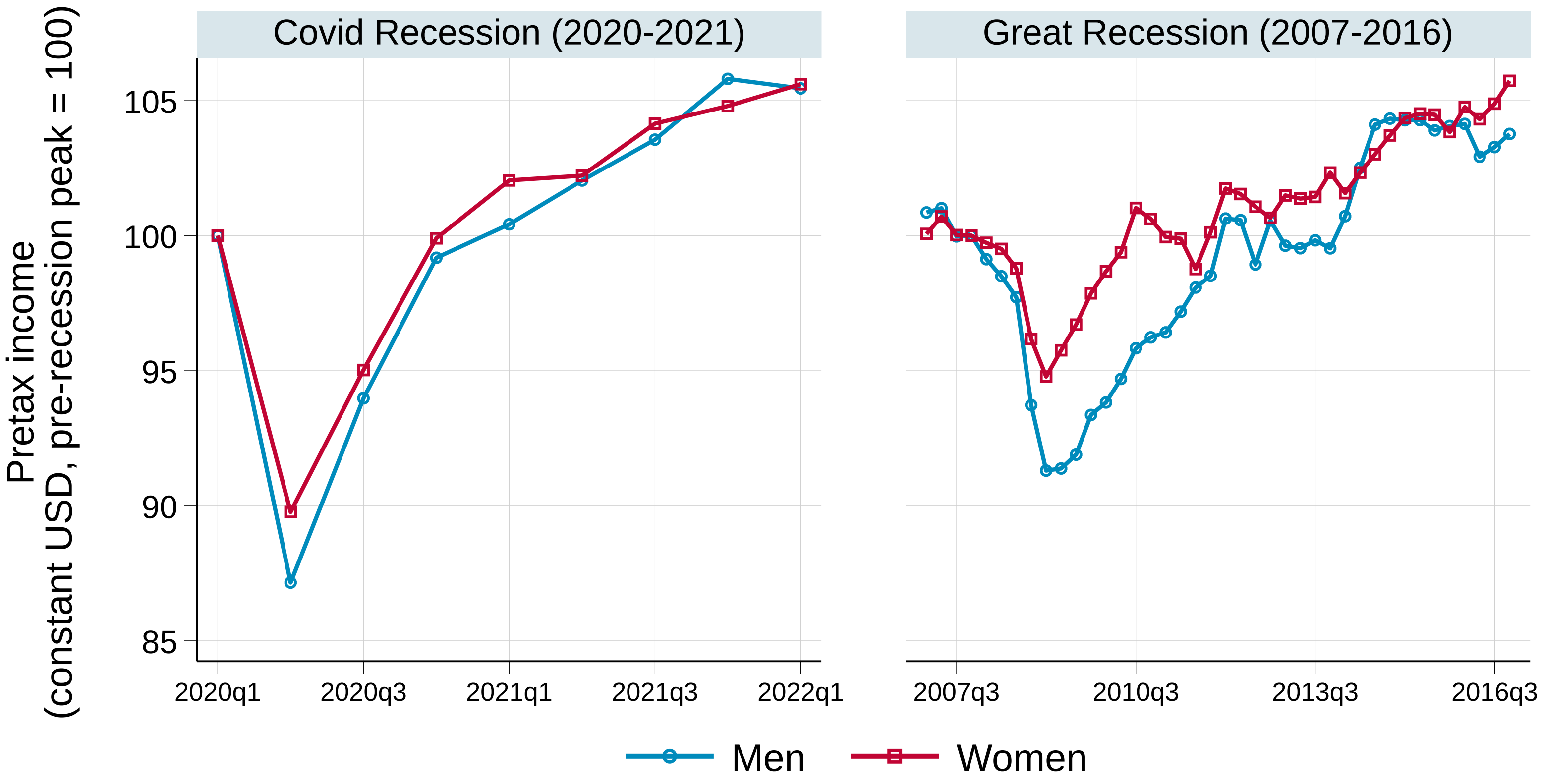
Notes: This figure shows the fraction of Black adults in the full adult population, in the top 10% of the wage distribution, 10% of the pretax income distribution, and top 10% of the wealth distribution. The series start in 1989, the first year of the Survey of Consumer Finances.

Figure A10: College Premium



Notes: This figure shows differences in income and wealth between people with no college education and people with at least some college education. The unit of observation is the individual adult. The labor income line restricts to the working-age population (individuals aged 20 to 64); other series include the entire adult population. The series start in 1989, the first year of the Survey of Consumer Finances.

Figure A11: Income Dynamics by Gender: Covid Recession vs. Great Recession



Notes: This figure shows the evolution of average pretax national income by gender during the Covid recession and its aftermath (left panel) and the Great Recession and its aftermath (right panel). Income is normalized to 100 in the quarter preceding each recession.

Table A1: Prediction Errors for Growth Rates of Income & Wealth (2 Years)

Concept	Bracket	All years					Excl. tax reforms				Recessions			
		Std. Dev.	Correct sign	RMSE	Bias	Std. Err.	Correct sign	RMSE	Bias	Std. Err.	Correct sign	RMSE	Bias	Std. Err.
Factor Income	Bottom 50%	6.6 pp.	91%	2.4 pp.	-0.7 pp.	2.3 pp.	91%	2.4 pp.	-1.0 pp.	2.2 pp.	92%	1.4 pp.	-0.4 pp.	1.4 pp.
	Middle 40%	2.3 pp.	93%	1.3 pp.	-0.2 pp.	1.3 pp.	97%	1.0 pp.	-0.1 pp.	1.0 pp.	83%	1.6 pp.	-1.2 pp.	1.0 pp.
	Next 9%	3.0 pp.	98%	1.2 pp.	-0.6 pp.	1.0 pp.	100%	1.1 pp.	-0.6 pp.	0.9 pp.	92%	0.9 pp.	-0.5 pp.	0.8 pp.
	Top 1%	9.1 pp.	93%	5.1 pp.	-1.0 pp.	5.0 pp.	94%	3.8 pp.	-1.1 pp.	3.6 pp.	83%	4.8 pp.	1.5 pp.	4.5 pp.
Pretax Income	Bottom 50%	4.9 pp.	88%	2.5 pp.	-1.2 pp.	2.2 pp.	88%	2.5 pp.	-1.1 pp.	2.3 pp.	83%	2.2 pp.	-1.8 pp.	1.4 pp.
	Middle 40%	2.4 pp.	95%	1.3 pp.	-0.1 pp.	1.3 pp.	100%	1.1 pp.	-0.0 pp.	1.1 pp.	83%	1.4 pp.	-1.1 pp.	0.9 pp.
	Next 9%	3.7 pp.	98%	1.2 pp.	-0.4 pp.	1.1 pp.	97%	1.1 pp.	-0.5 pp.	1.0 pp.	92%	0.7 pp.	0.1 pp.	0.7 pp.
	Top 1%	9.4 pp.	88%	5.2 pp.	-1.0 pp.	5.1 pp.	94%	3.8 pp.	-1.1 pp.	3.7 pp.	75%	4.9 pp.	1.5 pp.	4.6 pp.
Disposable Income	Bottom 50%	3.8 pp.	88%	3.5 pp.	-2.2 pp.	2.7 pp.	91%	3.1 pp.	-1.9 pp.	2.5 pp.	75%	4.8 pp.	-4.2 pp.	2.4 pp.
	Middle 40%	2.2 pp.	95%	1.3 pp.	-0.2 pp.	1.2 pp.	94%	1.1 pp.	-0.2 pp.	1.1 pp.	92%	1.1 pp.	-0.5 pp.	1.0 pp.
	Next 9%	3.3 pp.	93%	1.6 pp.	-0.0 pp.	1.6 pp.	91%	1.4 pp.	-0.2 pp.	1.4 pp.	83%	1.5 pp.	1.2 pp.	0.9 pp.
	Top 1%	9.3 pp.	93%	6.4 pp.	0.3 pp.	6.3 pp.	97%	4.4 pp.	-0.0 pp.	4.4 pp.	83%	6.7 pp.	1.9 pp.	6.5 pp.
Post-tax Income	Bottom 50%	3.8 pp.	88%	2.7 pp.	-1.6 pp.	2.2 pp.	97%	2.3 pp.	-1.2 pp.	2.0 pp.	92%	3.5 pp.	-2.9 pp.	1.9 pp.
	Middle 40%	2.9 pp.	93%	1.2 pp.	-0.2 pp.	1.2 pp.	97%	1.0 pp.	-0.2 pp.	1.0 pp.	83%	0.9 pp.	-0.5 pp.	0.8 pp.
	Next 9%	4.1 pp.	95%	1.4 pp.	-0.2 pp.	1.4 pp.	94%	1.2 pp.	-0.4 pp.	1.2 pp.	83%	1.0 pp.	0.7 pp.	0.7 pp.
	Top 1%	10.0 pp.	91%	6.2 pp.	-0.4 pp.	6.2 pp.	94%	4.3 pp.	-0.7 pp.	4.3 pp.	83%	6.0 pp.	1.2 pp.	5.9 pp.
Wealth	Middle 40%	8.2 pp.	88%	2.1 pp.	0.6 pp.	2.0 pp.	97%	1.8 pp.	0.1 pp.	1.8 pp.	92%	1.8 pp.	0.5 pp.	1.8 pp.
	Next 9%	6.6 pp.	93%	1.9 pp.	0.2 pp.	1.9 pp.	97%	1.7 pp.	0.2 pp.	1.7 pp.	83%	2.3 pp.	0.1 pp.	2.3 pp.
	Top 1%	9.9 pp.	93%	4.7 pp.	-3.1 pp.	3.5 pp.	94%	4.1 pp.	-2.7 pp.	3.1 pp.	83%	4.5 pp.	-2.8 pp.	3.5 pp.

Notes: This table report statistics for goodness of fit and noise of our 2-year ahead real income and real wealth growth predictions. See notes to Table 2.