Motivation: Lack of timely data on inequality

A major gap in government statistics is the lack of timely data on inequality:

- Macro national accounts published monthly and quarterly.
- But distributions from tax statistics available with lag of about 2 years.
- Current Population Survey available monthly, but covers only small fraction of GDP → impossible to know who benefits from economic growth in real time

This gap limits the ability of policymakers to design effective policies:

- **During Covid:** Did government redistribution undershoot? Overshoot?
- **Today:** Are real wages rising for low-income workers?

Our goal: Mobilize all public data to build realtime inequality stats
Contribution: Monthly microdata matching macro totals

This project: **prototype real-time monthly distributional national accounts:**

- **Output:** Monthly synthetic microdata, which distributes all of national income and wealth to individuals, matching macro totals.
- Following a recession, this can be used to compute “distributional output gaps:” which groups of the population are below their pre-crisis income level or trend.
- Incorporate all taxes and government transfers → reveal how national income is distributed and redistributed month-to-month.

**Estimates available on realtimeinequality.org** within a few hours of the publication of the national accounts:

- Based solely on **public data**.
- Hope that this will eventually be incorporated into official government statistics.
**Real income growth per adult in 2021**

Growth rates, gains, and income levels are annualized.

<table>
<thead>
<tr>
<th>Period</th>
<th>Last Calendar Year</th>
<th>Last Quarter</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Group</strong></td>
<td><strong>Growth (%)</strong></td>
<td><strong>Gain ($)</strong></td>
</tr>
<tr>
<td>Total</td>
<td>7.9%</td>
<td>$6.2k</td>
</tr>
<tr>
<td>Bottom 50%</td>
<td>13.4%</td>
<td>$2.2k</td>
</tr>
<tr>
<td>Middle 40%</td>
<td>5%</td>
<td>$4.2k</td>
</tr>
<tr>
<td>Top 10%</td>
<td>9.5%</td>
<td>$34k</td>
</tr>
<tr>
<td>Top 1%</td>
<td>13.4%</td>
<td>$190k</td>
</tr>
<tr>
<td>Top 0.1%</td>
<td>14.6%</td>
<td>$930k</td>
</tr>
<tr>
<td>Top 0.01%</td>
<td>14.7%</td>
<td>$4.2M</td>
</tr>
</tbody>
</table>
Two main methodological contributions

We combine all the publicly available data:

- Monthly/quarterly national accounts, employment, public-use tax microdata, Current Population Survey, etc.
- **Highlight:** Quarterly Census of Employment and Wages (QCEW) to estimate monthly wage inequality, e.g., top 1% wage income share.

We implement a first-of-its-kind **statistical match between survey data (CPS, SCF) and public-use tax data** using optimal transport:

- First comprehensive measures of income decomposable by race, gender, age, education, etc.
Methodology and Validation
Methodological Overview

Starting point: annual tax-based Distributional National Accounts microdata of income and wealth by Piketty, Saez, Zucman (2018) continuously updated/improved:

- Monthly file using moving average of current and adjacent yearly microdata.

How we move to high frequency:

- **Capital income/wealth:** rescaling to macro aggregates. Works because aggregate volatility much bigger than concentration volatility.
- **Labor income:** distribution estimated monthly using the Quarterly Census of Employment and Wages and from other sources.
Corporate profits: aggregate vs. concentration volatility

National income’s share of pretax corporate profits

Share of pretax corporate profits earned by the top 10%
Quarterly Census of Employment and Wages to estimate wage inequality

- Exhaustive administrative dataset (nearly all wage earners covered).
- Published quarterly with monthly data.
- Data by (6-digit NAICS industry code) × (county) × (ownership sector).
- ≈ 1,000,000 observations each month.
- Can be used to infer complete wage distribution (Lee, 2020).

→ QCEW predicts the wage distribution remarkably well, incl. the top 1%
Monthly CPS + QCEW Data predict wage income inequality

Bottom 50%

Middle 40%

Next 9%

Top 1%

○ Public-use Tax Data [Yearly]

---

QCEW + CPS [Monthly]
Statistical matching between the public use tax micro data, the CPS, the American Community Survey and the Survey of Consumer Finances.

- We match observations “one-to-one” using **optimal transport** using common variables.
- Resulting micro dataset respects joint distribution of all variables from each dataset.
- Much better than one sided matching or matching without replacement.
- Much more computationally demanding but doable today.
- Obviously not as good as comprehensive admin socio-economic (confidential) data

- Update employment, unemployment, and earning rank by 5-year age group × gender × race × education × marital status in the microdata to match aggregate employment and UI claims
Our monthly/quarterly data are directly comparable to annual data in both levels and distribution.

National Accounts also present monthly and quarterly data on an annualized basis.

True monthly/quarterly data have more inequality that gets smoothed out at annual frequency.

Our method bypasses this issue which is good for 2 reasons:
- Limited longitudinal high frequency data (Fixler, Gindelsky, Kornfeld 2021)
- Our Monthly, quarterly, and annual data are all directly comparable.

Our monthly statistics tell us how annual statistics would look like if monthly distributions stayed the same for 1 year.
Predicting Bottom 50% Growth Before Tax Data Becomes Available

![Graphs showing the relationship between actual and predicted growth rates for different years. The graphs illustrate how well the predicted growth rates align with the actual growth rates for the years 1984, 2012, 2017, and 2018.](image-url)
Predicting Top 1% Growth Before Tax Data Becomes Available

Top 1%: shares
The Distribution and Redistribution of National Income During Covid
Income of the bottom 50% since Covid (working age population, 20–64)
Income of the bottom 50% since Covid (working age population, 20–64)

Monthly income per adult (constant USD)

- Paycheck Protection Program
- Factor national income (matching national income)
Income of the bottom 50% since Covid (working age population, 20–64)

Monthly income per adult (constant USD)

Subsidized factor national income
Paycheck Protection Program
Factor national income (matching national income)
Income of the bottom 50% since Covid (working age population, 20–64)
Income of the bottom 50% since Covid (working age population, 20–64)

Monthly income per adult (constant USD)

- Regular cash transfers (net of taxes)
- Subsidized pretax national income
- Other benefits (net) (pensions & DI, minus contributions)
- Unemployment insurance benefits
- Subsidized factor national income
- Paycheck Protection Program
- Factor national income (matching national income)
Income of the bottom 50% since Covid (working age population, 20–64)
Income of the bottom 50% since Covid (working age population, 20–64)

- Post-tax disposable income
- COVID stimulus checks
- Regular cash transfers (net of taxes)
- Subsidized pretax national income
- Other benefits (net) (pensions & DI, minus contributions)
- Unemployment insurance benefits
- Subsidized factor national income
- Paycheck Protection Program
- Factor national income (matching national income)
Factor income: Before transfers, all groups lost, especially the bottom 50%
But disposable income gives a different picture
The Post-Covid Recovery: Are Real Wages Growing? For Whom?
Employment rate changes sharply over business cycle especially at bottom:
▶ Comparing wages of workers only is misleading
▶ Longitudinal comparisons of workers also affected by who stays employed

Solution: Consider labor income quantiles for the full working-age (20-64) population:
▶ Captures both job effects and pay effects
▶ Captures both employees and self-employed
▶ Can compare pay growth at points with equal macro-employment rates

Findings:
▶ Much faster real pay growth at bottom in 2018-2022 than 2007-2017
▶ Inflation eats pay increases only in early 2022
Working-age (20–64) Employment Rate

Employment to working-age population ratio (%) over time.
Real Labor Income Growth during Covid

Average real labor income index (2019m1 = 100)

- Top 1%
- 4th quartile
- 3rd quartile
- 2nd quartile

Year:
- 2019m1
- 2020m1
- 2021m1
- 2022m1
Labor Income, Working-age Population (20–64)

Annualized real labor income growth (%)

Percentiles (working-age population)

COVID recession and recovery (02/2020 to 05/2022)

Great recession and recovery (12/2007 to 05/2017)
Crisis and Recovery by Race and Gender
A Comprehensive Estimate of Black/White Economic Disparities

Labor income (working-age population)
A Comprehensive Estimate of Black/White Economic Disparities

Labor income (working-age population)

Wealth
A Comprehensive Estimate of Black/White Economic Disparities

Labor income (working-age population)

Wealth

Pretax capital income
A Comprehensive Estimate of Black/White Economic Disparities

- Labor income (working-age population)
- Pretax income
- Pretax capital income
- Wealth
A Covid “Shecession”? Recessions and Recovery by Gender

Covid Recession (2020-2021)

Great Recession (2007-2016)

Pretax income (constant USD, pre-recession peak = 100)

Men

Women

2020q1 2020q3 2021q1 2021q3 2022q1
2007q3 2010q3 2013q3 2016q3
Conclusion

- It’s possible to track inequality in near real-time.
- Estimates based solely on public data.
- Prototype to be improved and run by government agencies down the road.
- Stark contrast between the recoveries from the last two recessions.

realtimeinequality.org
Supplementary Slides
First dataset of size $n$ with weights $u = (u_1, u_2, \ldots, u_n)$.

Second dataset of size $m$ with weights $v = (v_1, v_2, \ldots, v_m)$.

(Without restriction, assume weights sum to one.)

The optimal transport map $\Gamma \in \mathbb{R}^{n \times m}$ matches observations “one-to-one” while minimizing the sum of distances between matched observations.

Formally, it is the solution of:

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

where $D_{ij}$ is the distance between observation $i$ in the first dataset and $j$ in the second (we use the $L^1$ norm).

At the optimum, the map $\Gamma$ is sparse and contains at most $m + n - 1$ entries.
How we Simulate COVID-specific Programs

- Stimulus checks
- Federal Pandemic Unemployment Compensation
- Paycheck Protection Program:
  - Incidence: 70% to capital, 30% to labor (Autor et al., 2022)
How we Simulate COVID-specific Programs

- Stimulus checks

- Federal Pandemic Unemployment Compensation

- Paycheck Protection Program:
  - Incidence: 70% to capital, 30% to labor (Autor et al., 2022)
  - **Novel estimate of the program’s distributional effects on labor:**
    - Data for each individual individual loan publicly available.
    - Match to QCEW cells by (date) × (county) × (6-digit NAICS industry code).
    - About 5,700,000 individual loans matched to 5,500,000 QCEW cells.
    - **Extensive margin:** workforce covered by wage percentile.
    - **Intensive margin:** share of wage bill covered by wage percentile.
Rule-based simulation of stimulus checks.

- 04/2020: $1,200, phasing out from $75,000 to $99,000 (≈ $300bn, 1.5% of NI).
- 01/2021: $600, phasing out from $75,000 to $87,000 (≈ $170bn, 0.9% of NI).
- 03/2021: $1,400, phasing out from $75,000 to $80,000 (≈ $400bn, 2% of NI).
Paycheck Protection Program: Distribution of Proceeds

Percentile of wage distribution

Proceeds spent on:
- EIDL refinancing
- Debt interest
- Mortgage interest
- Utilities
- Rent
- Health care
- Payroll

Share of PPP loan amounts (%)
Paycheck Protection Program: Share Forgiven

Percentile of wage distribution vs. Share forgiven (%)

- X-axis: Percentile of wage distribution
- Y-axis: Share forgiven (%)

The graph shows the distribution of share forgiven across different percentiles of wage distribution.
Paycheck Protection Program: Share of Workforce Covered

Share of workforce (%) vs. Percentile of wage distribution.
Prediction: Bottom 50% (Share)

1 year ahead

2 years ahead

Actual change in share (pp.)

Predicted change in share (pp.)
Prediction: Top 1% (Share)

1 year ahead

2 years ahead

Actual change in share (pp.) vs. Predicted change in share (pp.) graphs for 1 year and 2 years ahead.
Income of the bottom 50% since COVID (working age population, 20–64)

- Post-tax national income (matching national income)
- Government deficit
- Other government spending
- Medicaid and Medicare
- Post-tax disposable income
- COVID stimulus checks
- Regular cash transfers (net of taxes)
- Subsidized pretax national income
- Other benefits (net) (pensions & DI, minus contributions)
- Unemployment insurance benefits
- Subsidized factor national income
- Paycheck Protection Program
- Factor national income (matching national income)