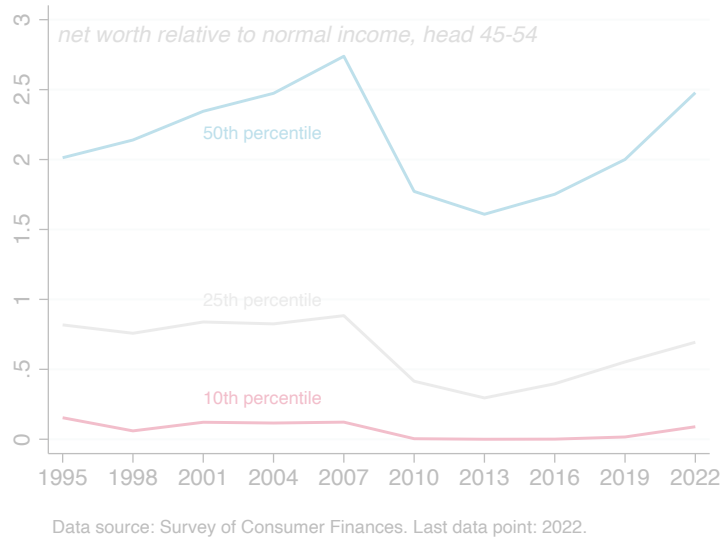
Opportunities to Improve Research and Policy with a System of Household Income, Consumption, and Wealth Data



Karen Dynan
Harvard University

ASSA Session on "**Creating an Integrated System of US Household Income, Consumption, and Wealth Data and Statistics to Improve Research and Policy**"
January 6, 2024

Outline

Fundamental research questions that could be informed by an integrated ICW system

Policy applications

[Additional] reasons why we need to go beyond currently available data sources

Implications for what features an integrated ICW data set should have

Fundamental research questions that could be informed by an integrated ICW system

Fundamental research questions that could be informed by an integrated ICW system [an incomplete list]

Trends and levels of **inequality and economic mobility**

Assessing **well-being**, **hardship**, and **exposure to risk**—at the business cycle frequency and over the longer-run

Better understanding consumer behavior—**propensities to consume, borrow, save, take on risk**

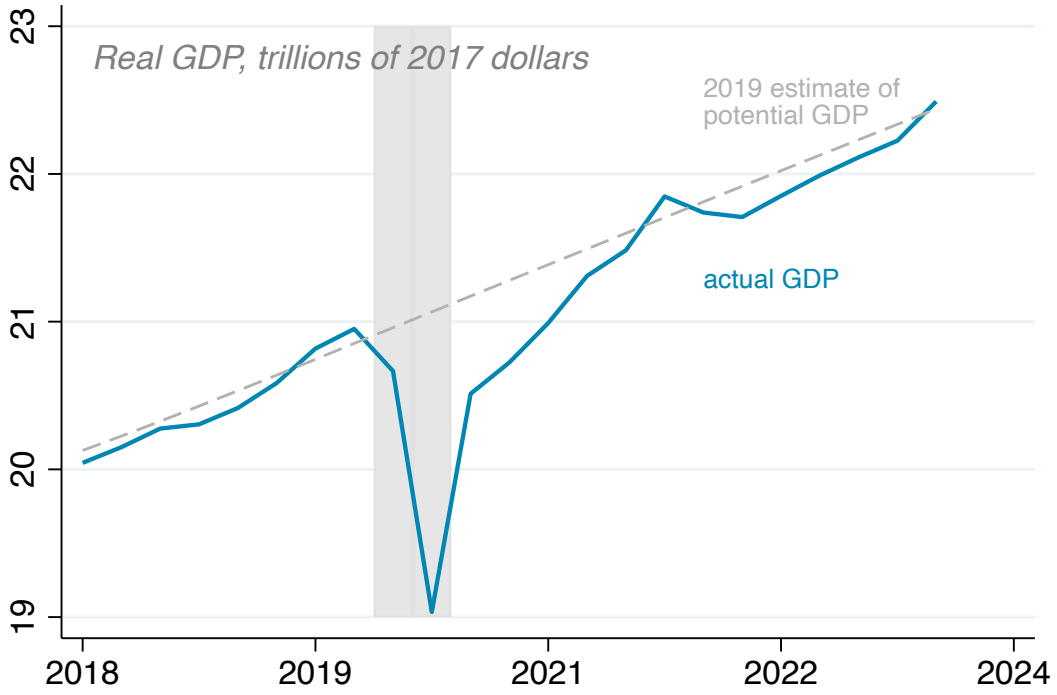
- heterogeneity by income/wealth/access to financial opportunities/other characteristics
- heterogeneity by type of resource shocks (permanent/transitory; income/wealth)

Implications of shifts in demographics, household structure, and the economic environment for the dynamics of the economy

Policy applications

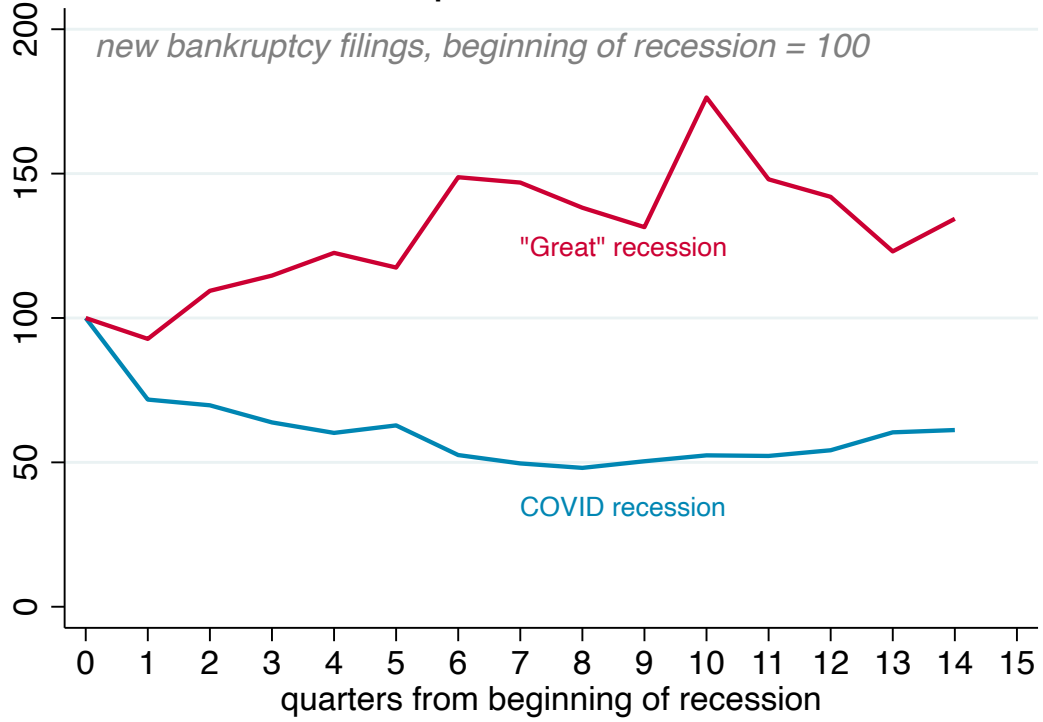
Policy application 1: Taking lessons from the vigorous US policy response to the economic fallout from the pandemic

Macro outcomes compared favorably to the last recession and to other countries



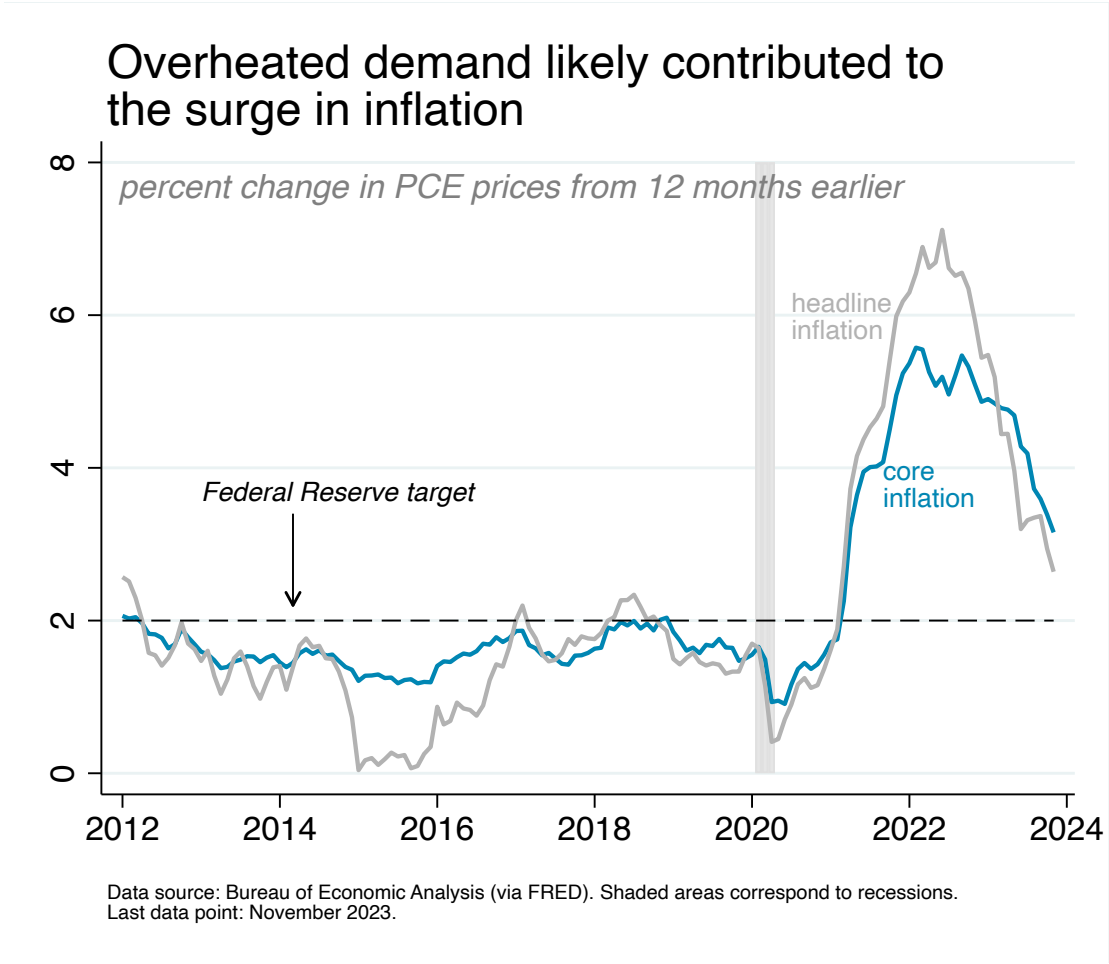
Data sources: Congressional Budget Office and Bureau of Economic Analysis.
Potential GDP series level-adjusted to reflect revisions to GDP since 2019.
Shaded area corresponds to recession.

Outcomes at the household level were much better than post-financial-crisis

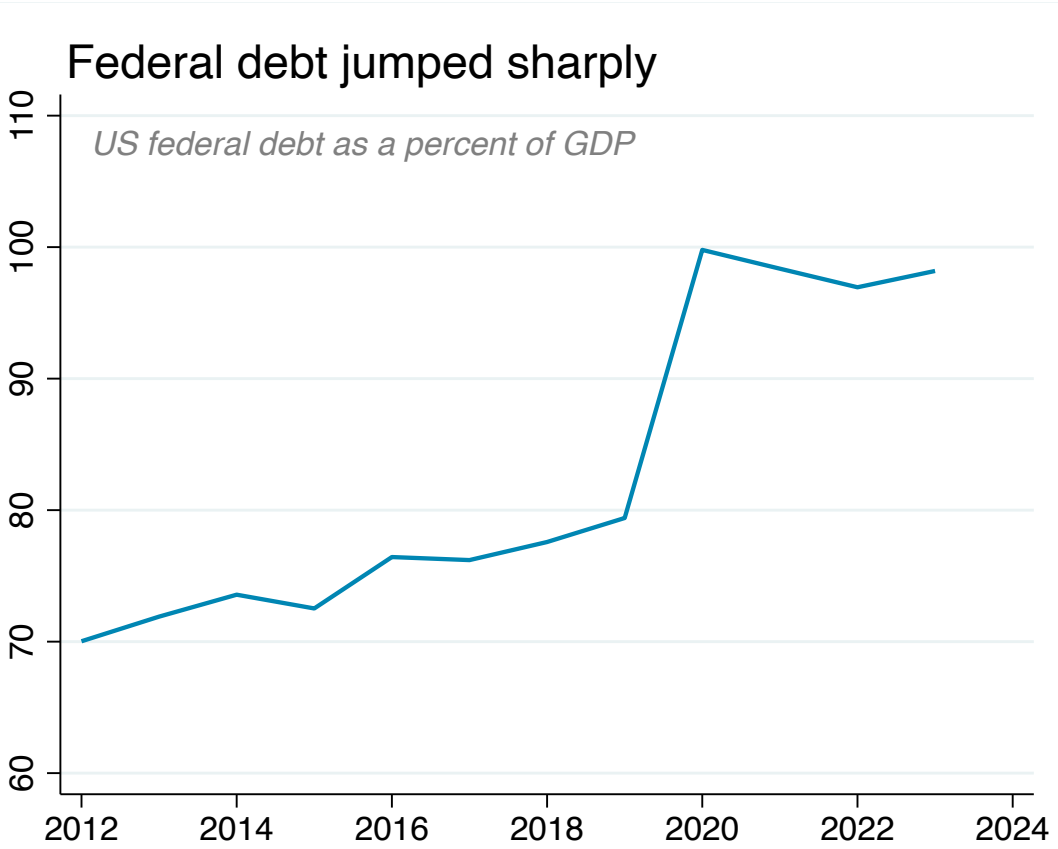


Data source: Federal Bank of New York. Shaded areas correspond to recessions.
Last data point: 2023q3.

But there were also costs to the pandemic policy response



Data source: Bureau of Economic Analysis (via FRED). Shaded areas correspond to recessions.
Last data point: November 2023.



Data source: Congressional Budget Office 2023 Long-term Budget Outlook.
Last data point: 2023.

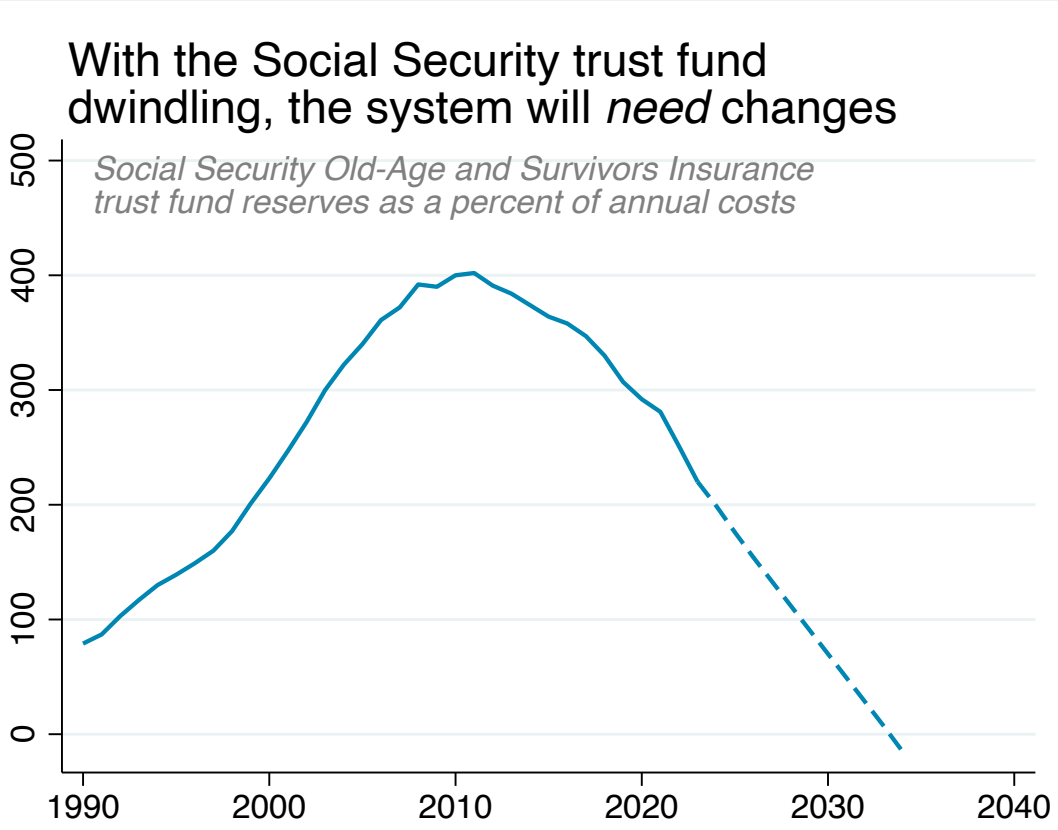
Research based on better ICW data could help mitigate macro fluctuations and reduce their costs

Better ICW data could **help reveal "cracks"** before they turn into big problems—e.g., how many borrowers taking on unsustainable levels of debt in the early 2000s?

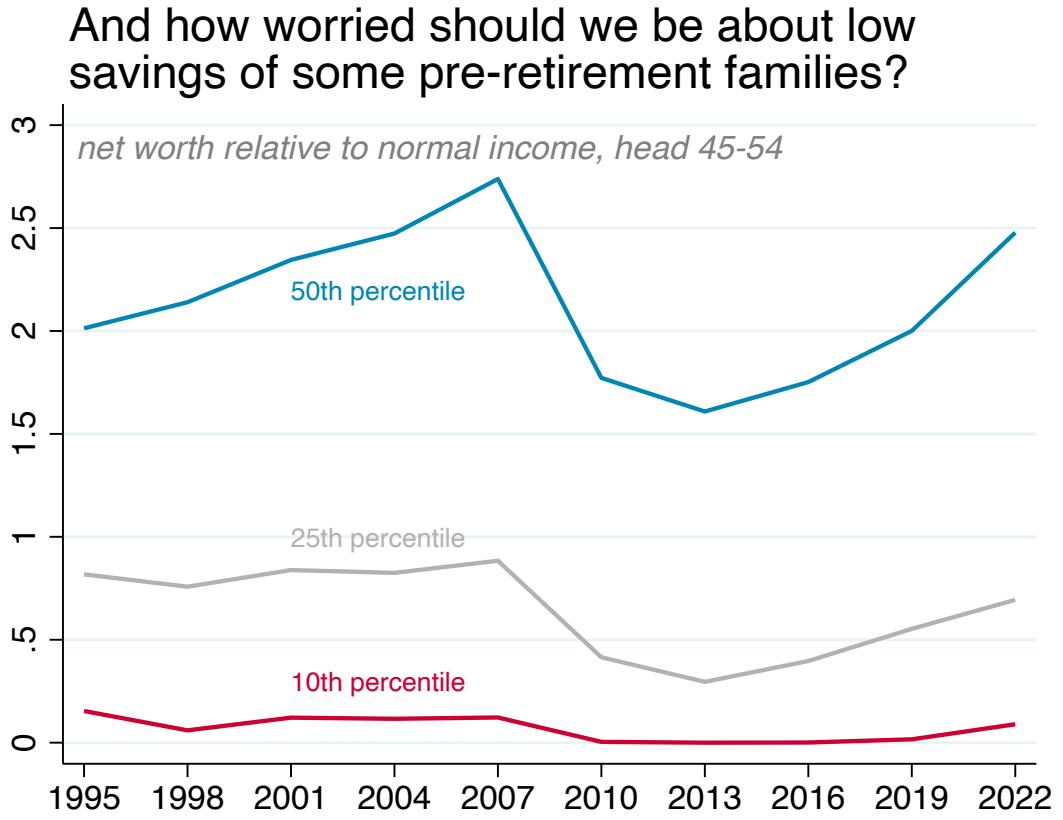
Better ICW data could **help with targeting**—e.g., who really needed another "COVID check" in March 2021?

Better ICW data could shed more light on **the likely response to countercyclical interventions**—e.g., a better grasp of likely MPCs could help us understand whether demand stimulus is not enough (as was probably the case in the early 2010s) or too much (as was probably the case in the early 2020s)?

Policy application 2: Informing changes to Social Security and broader retirement savings policies



Data source: 2023 Social Security Trust Fund Report.
Dashed line shows projection. Last data point: 2053.



Data source: Survey of Consumer Finances. Last data point: 2022.

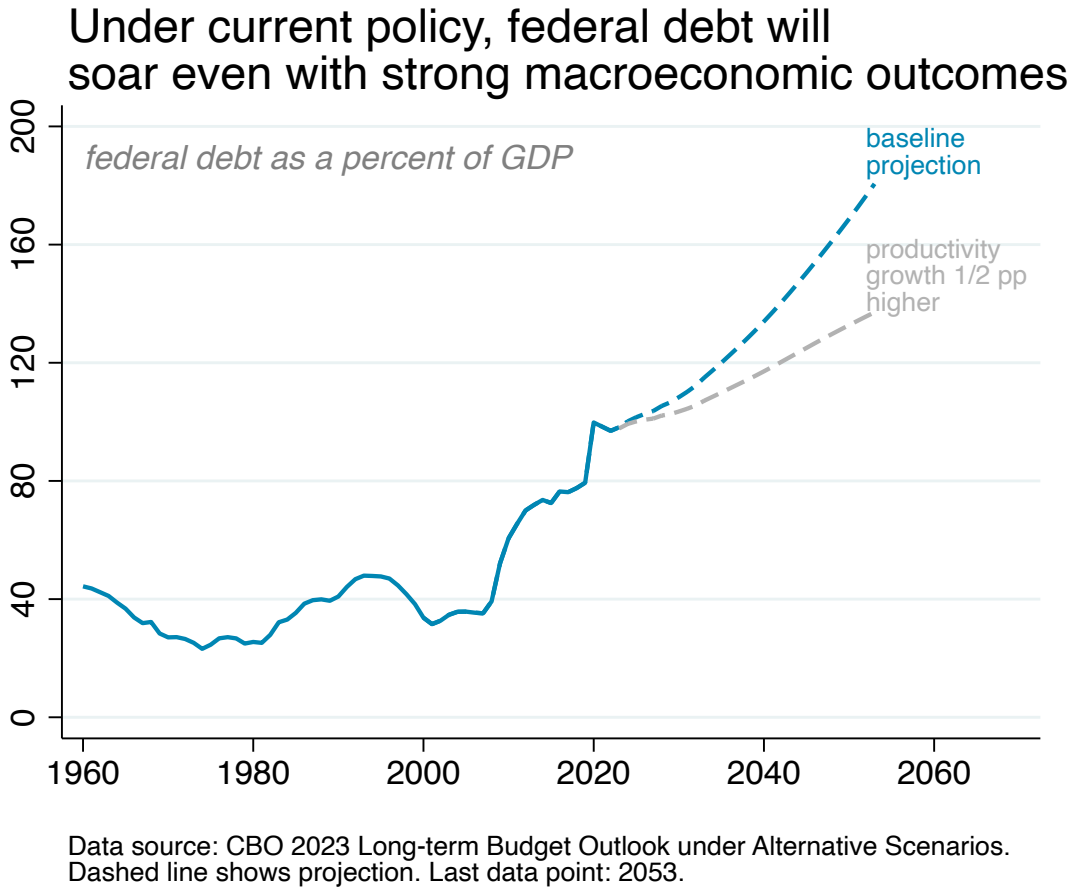
Research based on better ICW data could yield essential background for the retirement security policy discussion

Better ICW data could provide a **truly comprehensive picture of household resources** (including pension wealth, family structure, and financial connections across generations)—which is needed to understand who is at risk of hardship at older ages

Better ICW data could resolve debates over trends in inequality, which **bear on what people view as "fair"** when it comes to cutting benefits or raising taxes

Better ICW data could help **the many outstanding questions about how retirement policy affects incentives** for people to accumulate their own retirement savings

Policy application 3: Informing other fiscal policy choices that will be *needed* to put federal debt on a sustainable path

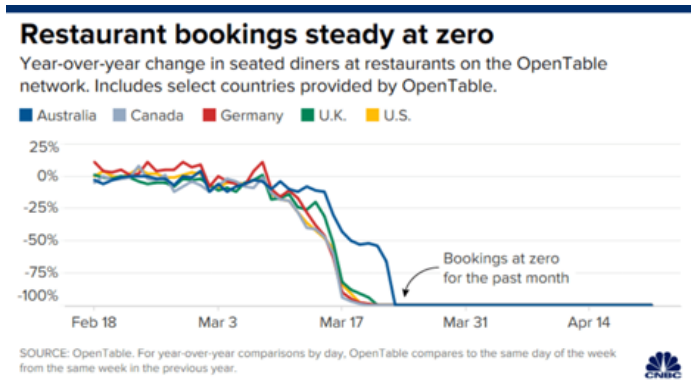


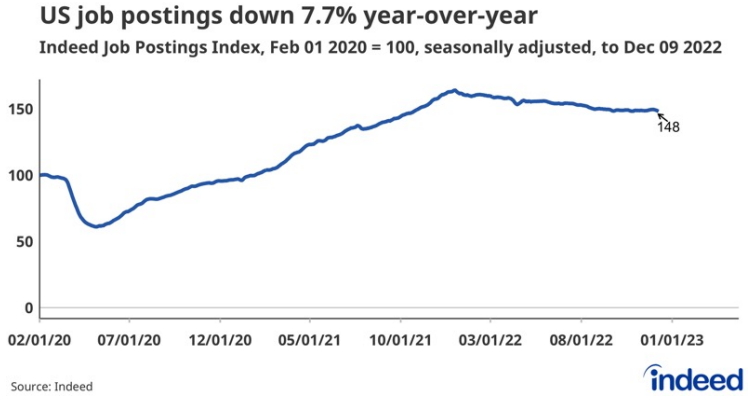
In addition to changes in Social Security and other spending on older Americans, **changes to safety net programs and the tax system** will be on the table when it comes to addressing the unsustainable path of federal debt

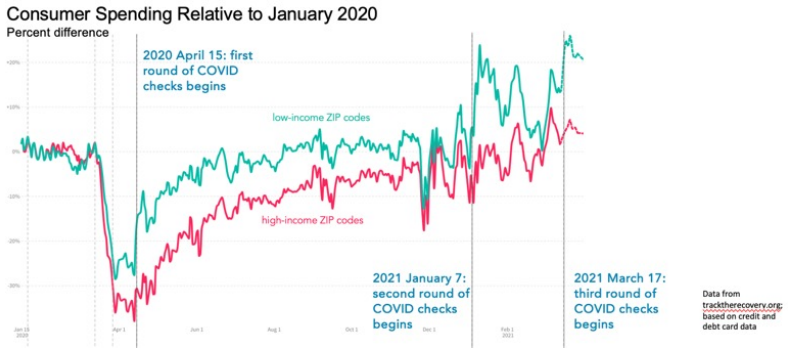
Research based on ICW data could produce better outcomes by shedding more light on **the trade-offs associated with these decisions**

[Additional] reasons why we need to go beyond currently available data sources

Recent years have seen a burst of “big data” innovations that have supported policy-related measurement and research







Essential pandemic research from [a special Brookings volume] produced just months after the pandemic set in

Initial impacts of the pandemic on consumer behavior

Natalie Cox, Peter Ganong, Pascal Noel, Joseph Vavra, Arlene Wong, Diana Farrell, Fiona Greig, and Erica Deadman ·
Thursday, June 25, 2020

Safety net programs and poverty during the COVID-19 crisis

Jeehoon Han, Bruce D. Meyer, James X. Sullivan, Marianne Bitler, Hilary W. Hoynes, and Diane Whitmore Schanzenbach ·
Thursday, June 25, 2020

COVID-19 and labor markets

Alexander W. Bartik, Marianne Bertrand, Feng Lin, Jesse Rothstein, Matthew Unrath, Tomaz Cajner, Leland D. Crane, Ryan A. Decker, John Grigsby, Adrian Hamins-Puertolas, Erik Hurst, Christopher J. Kurz, and Ahu Yildirmaz ·
Thursday, June 25, 2020

But private “big data” sources are usually not designed for such purposes and thus can yield confusing and conflicting results

Private big data sources are often:

- **Not representative**
- **Not aligned with economic concepts** [e.g., credit card spending is not “consumption”]
- Based on methodologies that are **not transparent** [nor always developed by trained experts]
- Proprietary and thus **only available to a limited set of researchers** [typically well-resourced]

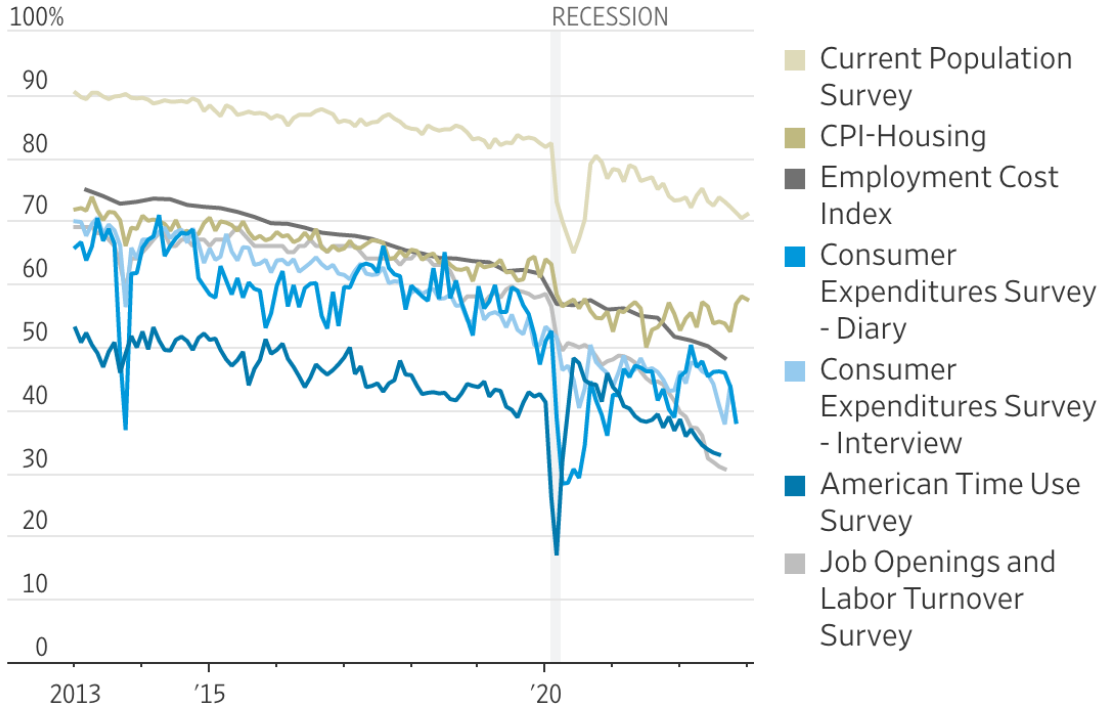
The risk with “letting a thousand flowers bloom” with big data is that inconsistent results [even if reconcilable] will add **noise to the policy conversation** or let people choose the results that fit their priors or interests

Cherry picking, **suppressing evidence**, or the **fallacy of incomplete evidence** is the act of pointing to individual cases or data that seem to confirm a particular position while ignoring a significant portion of related and similar cases or data that may contradict that position. Cherry picking may be committed intentionally or unintentionally.[2]

WIKIPEDIA
The Free Encyclopedia

While new opportunities are also arising with traditional government data sources, there are also challenges

Falling response rates may undermine the quality of traditional government surveys (particularly for more granular information)



Screenshot from *Wall Street Journal*

Changes made to protect privacy may limit access and the information available to study

Tweetorial: formal privacy for social scientists. If you collect, publish or analyze data, understand the revolution happening in safe data publication. Stat agencies, @Google, @Apple, @Microsoft, @Facebook, @LinkedIn are all struggling with the same problem.
#dataprivacy
2. What is formal privacy? Mathematical definitions and theorems that translate concepts from cryptography into algorithms that provably bound the worst-case information leakage due to the publication of a collection of statistics using confidential data.
#differentialprivacy
3. What is information leakage? Think of the confidential data as an encryption. Published statistics are clues to the encryption (deliberately, they describe properties of the data). The more statistics published, the closer one gets to full knowledge of the confidential data.
4. This is called #databasereconstruction. Original paper: Dinur and Nissim 2003 http://www.cse.psu.edu/~ads22/privacy598/papers/dn03.pdf.
5. Easier read: @xchatty Garfinkel et al. 2018. (https://queue.acm.org/detail.cfm?id=3295691).
6. There is an unavoidable tension between publishing statistics and protecting confidentiality. Crypto lesson 1: publishing too many statistics, too accurately, leaks all the confidential data with near certainty. (https://arxiv.org/abs/1701.00752)
7. What's the harm? Data are collected to be analyzed. #databasereconstruction rebuilds a record-level image of the confidential data outside the data curator's firewall. Can individual records be re-identified from this image? Does the re-identification harm those individuals?

Start of a 46-part "tweetorial" from John Abowd, former Chief Scientist at the Census Bureau

Implications for what features an integrated ICW data set should have

What ICW features are needed to answer these questions?

1. **Comprehensive information** about resources [including pension wealth, insurance, bequests, credit access, informal risk-sharing channels]
2. Attention to **measuring the low-end** [given the greater likelihood of hardship, under-reporting of benefits, greater complexity of resources/family structure]
3. Capacity to produce **high-quality granular results** [necessary information and sample sizes]
4. **Broad access** (with appropriate privacy protection) [not just for researchers with resources]
5. **Panel** dimension [or ability to construct useful pseudo-panels]

Harder but would be nice to have …

6. **I, C, W independently measured** so that errors are uncorrelated [if not, transparency about potential for correlations, so researchers can work around them]
7. **Timeliness**, **high frequency** [second-best: usefulness as a check on private "big data" indicators]