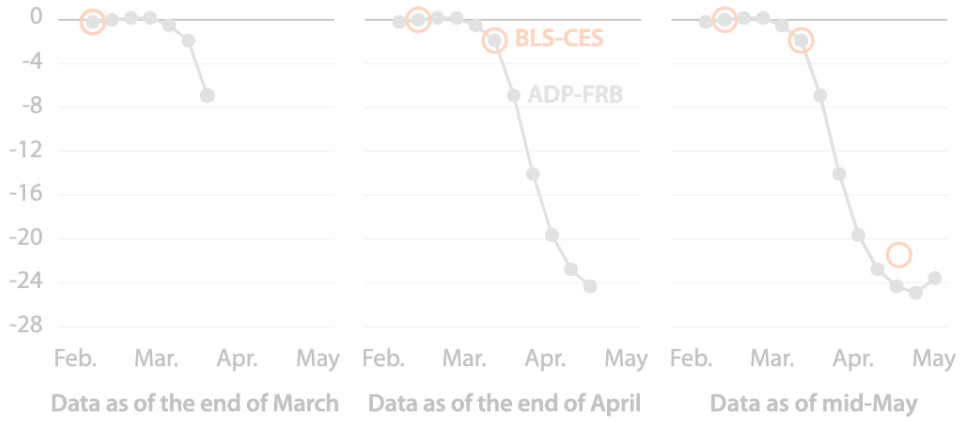
Using Data to Make Better Economic Policy in the 21st Century



Karen Dynan
Harvard University

MDI Distinguished Lecture
Georgetown Massive Data Institute
March 23, 2023

For decades, as the primary collectors, processors, and curators of the raw information underlying economic statistics, government statistical offices were near monopoly providers of [the source information for economic measurement].

Today, in contrast, staggering volumes of digital information relevant to measuring and understanding the economy are generated each second by an increasing array of devices that monitor transactions and business processes as well as track the activities of workers and consumers.

Abraham, Jarmin, Moyer, and Shapiro (2022)

Policymakers use economic data in many ways

To set key provisions of government programs

To evaluate whether new interventions are effective

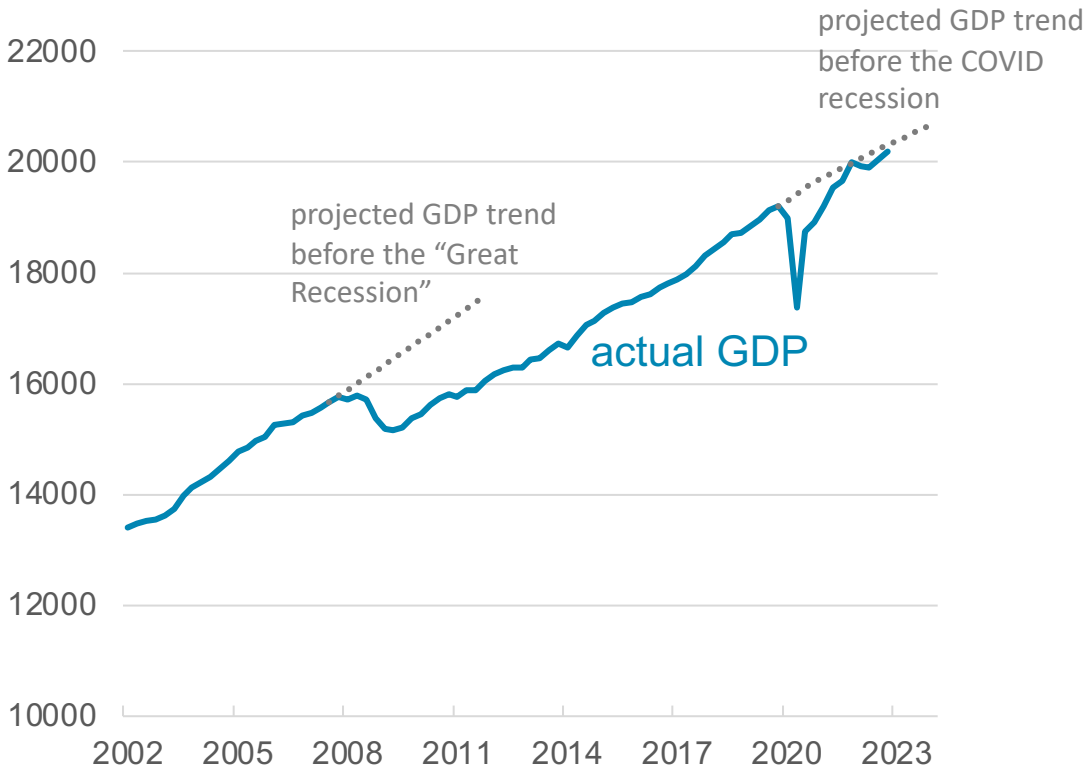
To do cost-benefit analysis for regulations

To track economic developments and support policies that stabilize the economy and help individuals in the face of economic disruption

focus of today's talk

The stakes are high with macroeconomic stabilization policy

At the national level



Real GDP in 2012 dollars from the Bureau of Economic Analysis via FRED, last data point 2022:Q4; projected trends from the Congressional Budget Office

And at the individual level

THE WALL STREET JOURNAL.

By Sara Murray
Updated May 9, 2009 11:59 p.m. ET

The Curse of the Class of 2009

For College Grads Lucky Enough to Get Work This Year, Low Wages are Likely to Haunt Them for a Decade or More

The traditional data toolkit for economic policymakers

Important new additions to the economic data toolkit

Challenges associated with the new tools

Constructive next steps

The traditional data toolkit for economic policymakers

Important new additions to the economic data toolkit

Challenges associated with the new tools

Constructive next steps

Data has long been the cornerstone of good economic policymaking

Key decisions about monetary and fiscal policy have been informed by a set of **high quality traditional indicators**, mostly produced **by experienced statistical experts at government agencies**

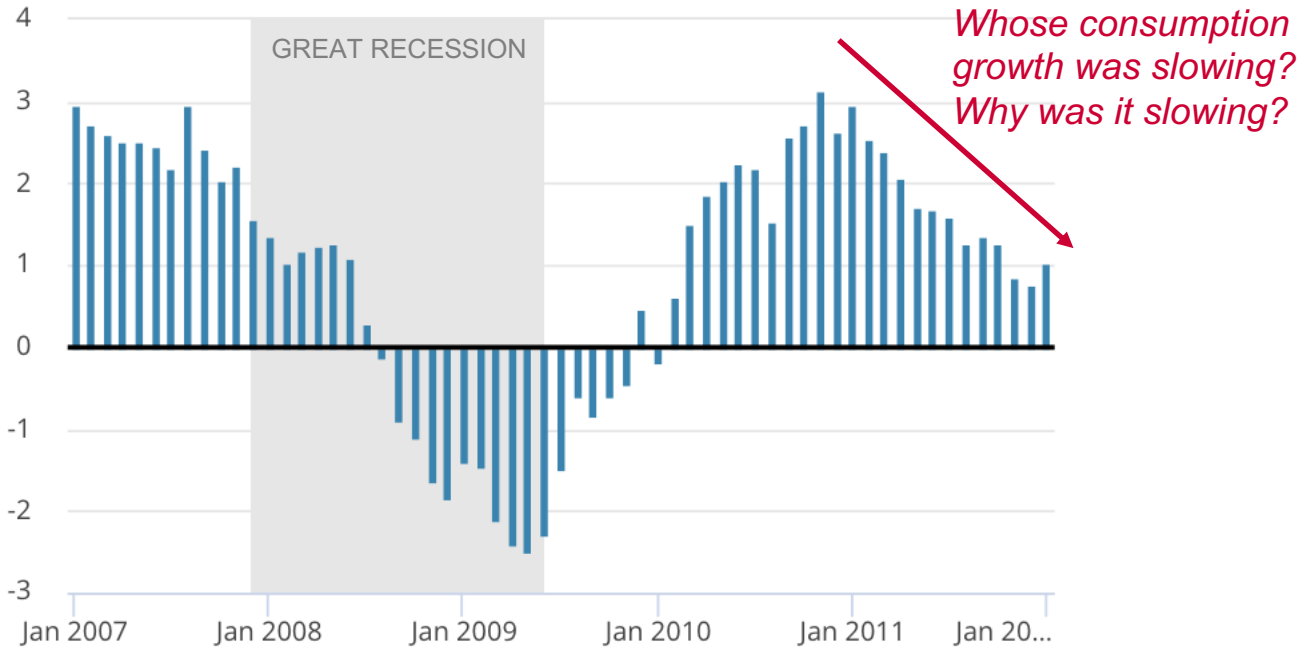
- GDP and its components, along with source data
- Labor statistics
- Financial flows
- [With a lag] some individual-, household-, and firm-level data sets

But this traditional data toolkit **also had its limitations**

It was often hard to know what was going on “under the hood” with just the traditional data toolkit

Aggregate data only speaks to what is happening **on average in the economy**

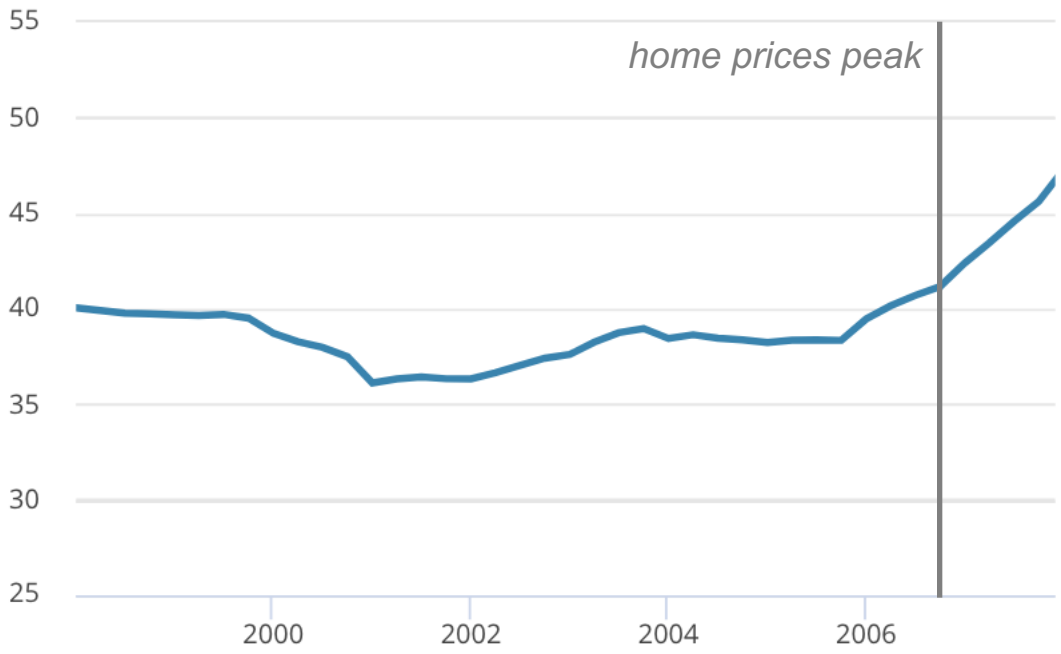
Consumer Spending Growth around the Time of the Great Recession



12-month change in real PCE from the Bureau of Labor Statistics via FRED

The traditional data toolkit missed important developments in the tails of the distribution

Aggregate Mortgage Loan-to-Value Ratio before the Financial Crisis



100 minus owners' equity as a percentage of household real estate from US Financial Accounts via FRED

Aggregate mortgage leverage rose during the housing bubble, but at <50%, **it looked like there was plenty of room for home prices to fall** without putting homeowners underwater

Analysis done since then using more granular data shows that the median subprime borrower in 2006 had an LTV of 100%

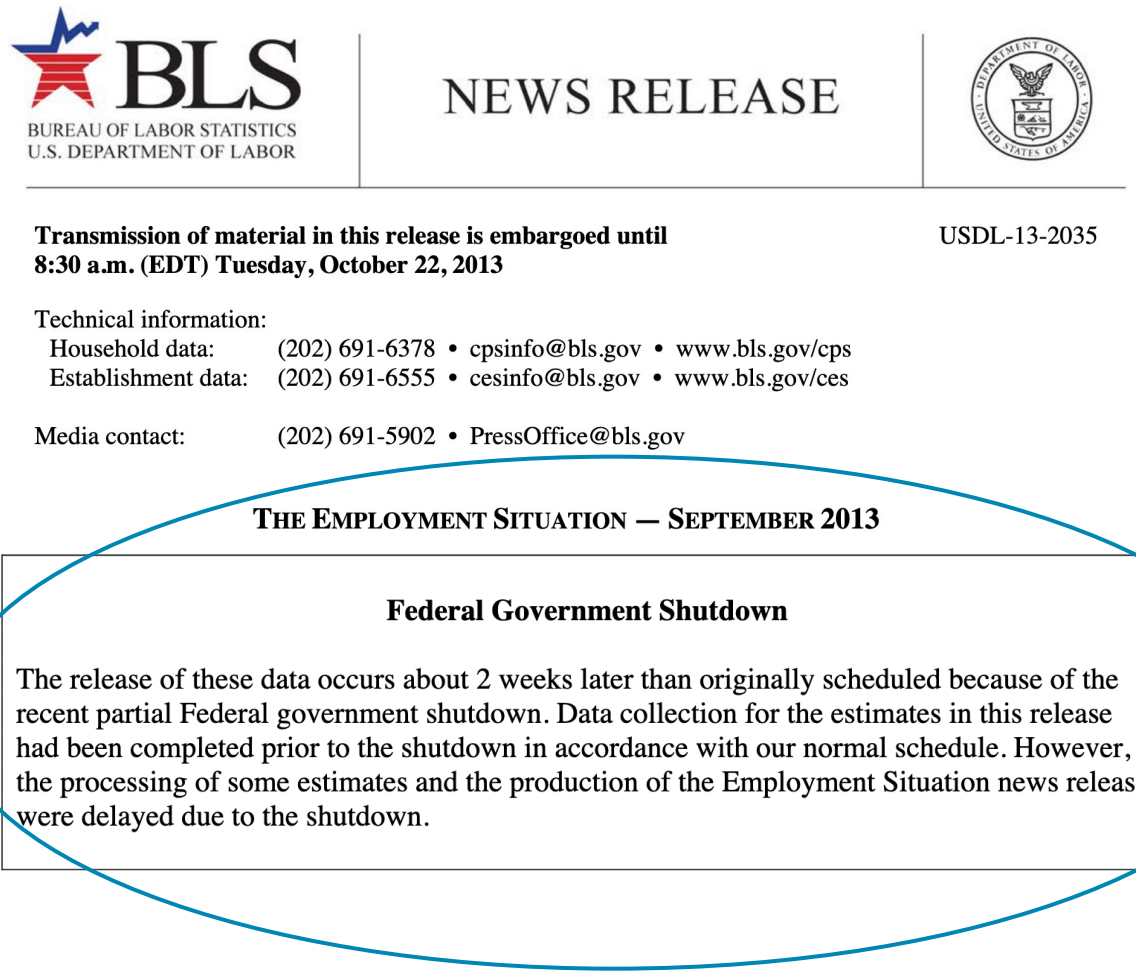
Traditional data can be noisy and subject to revision

Cajner, Feiveson, Kurz, and Tevlin (2022) discuss **the significant downward revisions to data for the period preceding the Great Recession**:

> *Had the revised data, or an expansive set of nontraditional data, been in policymakers' hands at the time of the August [2007] meeting, a better picture of a less robust state of the economy might have assisted policymakers.*

Accurately capturing business cycle turning points is essential to make good macroeconomic policy, but they can be obscured by provisional estimates of traditional indicators if prior trends are used to fill in still-missing source data

Traditional data may not be available if the government closes



BLS
BUREAU OF LABOR STATISTICS
U.S. DEPARTMENT OF LABOR

NEWS RELEASE

Transmission of material in this release is embargoed until 8:30 a.m. (EDT) Tuesday, October 22, 2013

USDL-13-2035

Technical information:
Household data: (202) 691-6378 • cpsinfo@bls.gov • www.bls.gov/cps
Establishment data: (202) 691-6555 • cesinfo@bls.gov • www.bls.gov/ces

Media contact: (202) 691-5902 • PressOffice@bls.gov

THE EMPLOYMENT SITUATION — SEPTEMBER 2013

Federal Government Shutdown

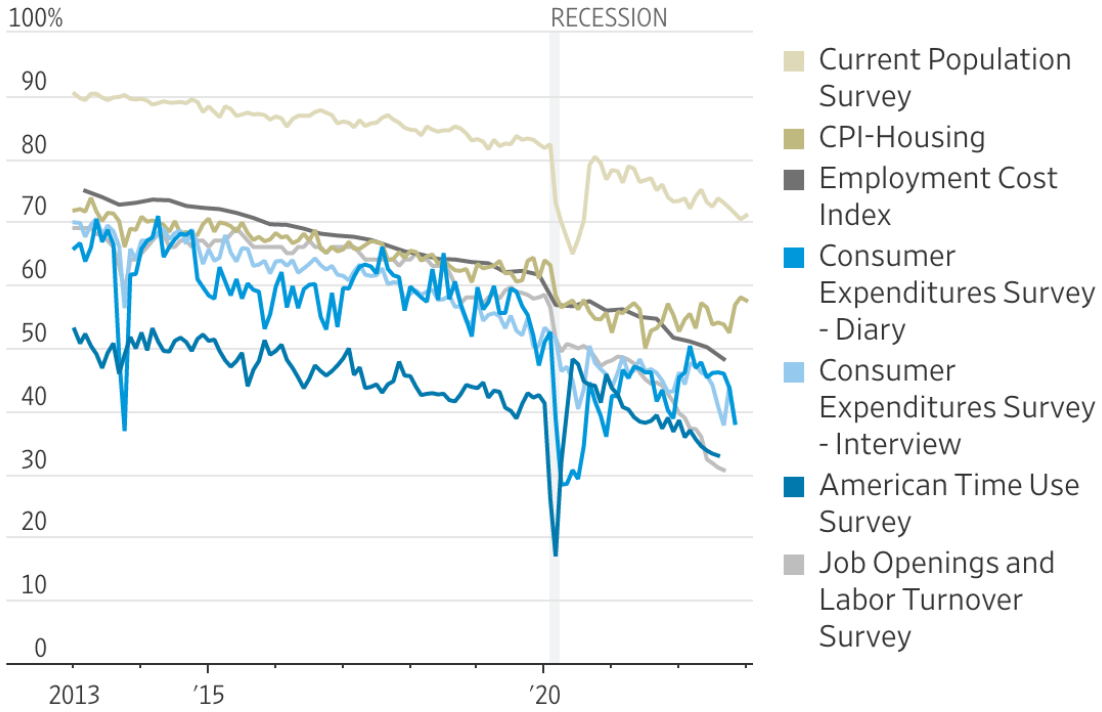
The release of these data occurs about 2 weeks later than originally scheduled because of the recent partial Federal government shutdown. Data collection for the estimates in this release had been completed prior to the shutdown in accordance with our normal schedule. However, the processing of some estimates and the production of the Employment Situation news release were delayed due to the shutdown.

This is a real issue!

The 16-day federal government shutdown that began on October 1, 2013 delayed the release of the September employment report and affected collection for October

The producers of traditional data have hurdles in front of them

Falling survey response rates



Screenshot from *Wall Street Journal*

The need to protect the privacy of government survey respondents

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, Chief Scientist at the Census Bureau

The traditional data toolkit for economic policymakers

Important new additions to the economic data toolkit

Challenges associated with the new tools

Constructive next steps

Recent years have seen important new additions to the economic policy data tool kit

Policymaking has been enhanced by a wave of innovation in several categories:

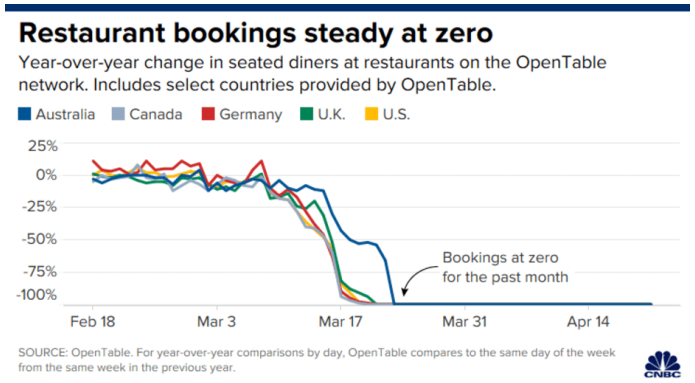
New "trackers" based on administrative records (and, in some cases, the underlying micro data)

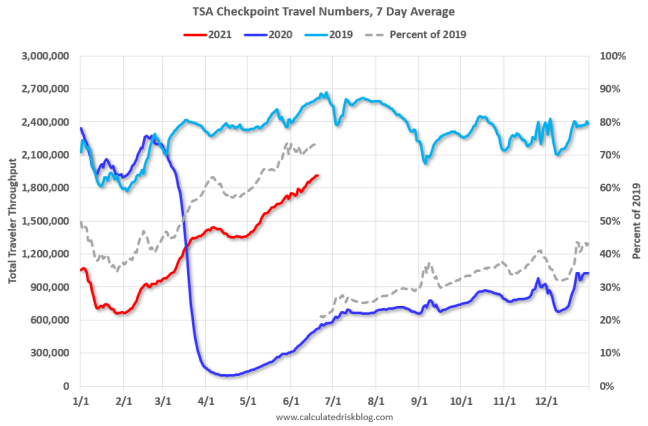
"Mash-ups" of existing (and sometimes newly available) data

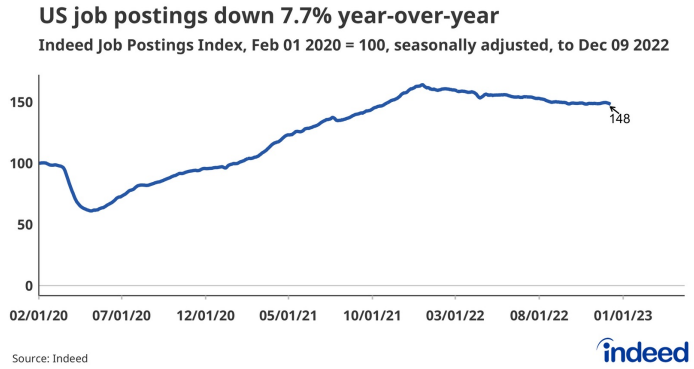
New data collections

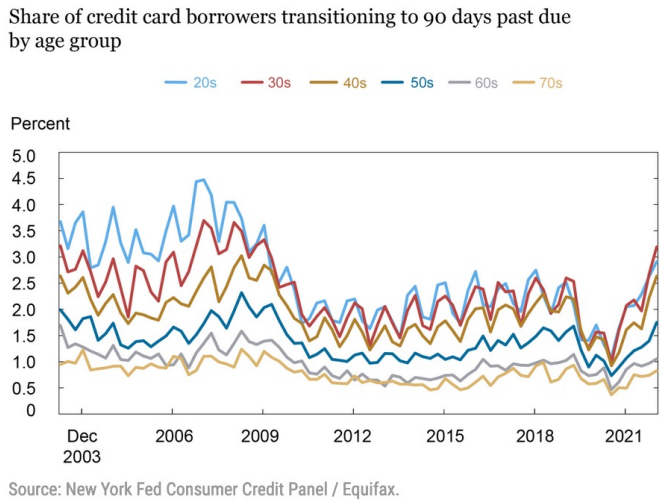
Facilitated by advances in private-sector practices, technology, and (relatedly) the capacity and willingness to disseminate data

A wide range of new economic activity trackers based on administrative records have been developed





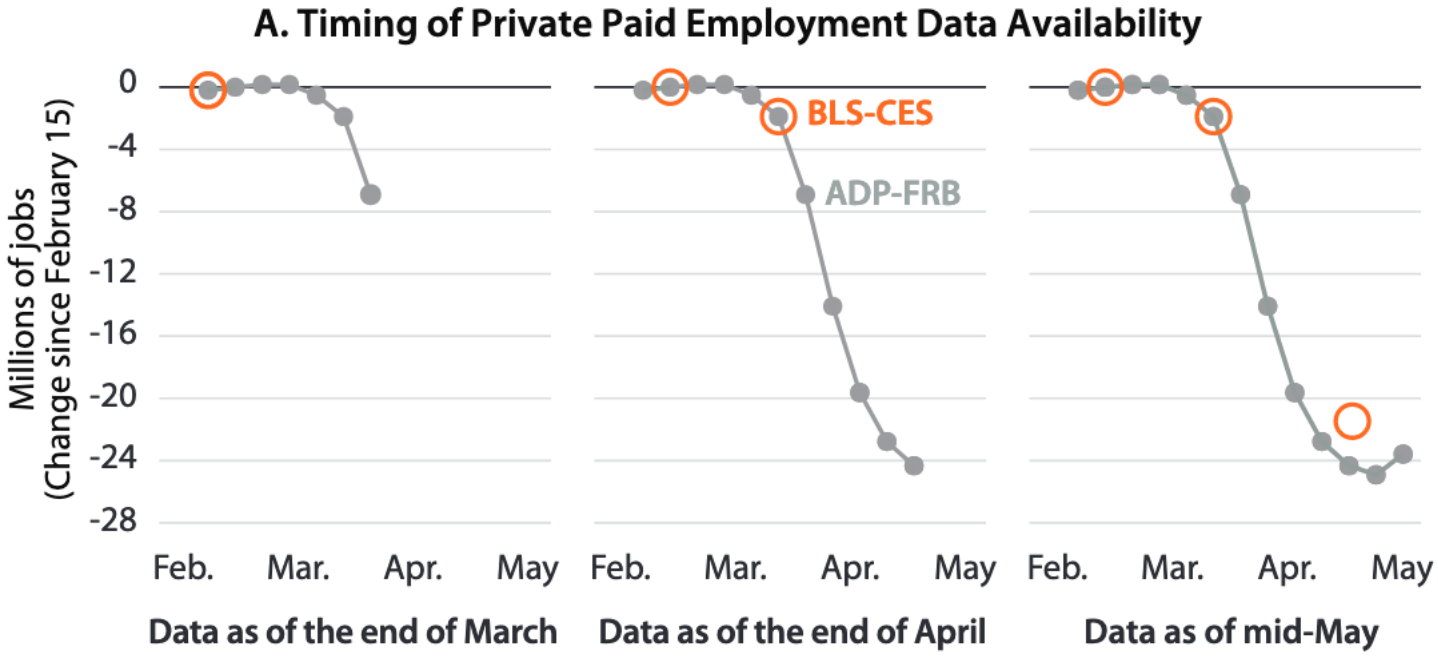




New trackers shed light on the 2020 economic collapse well ahead of traditional indicators

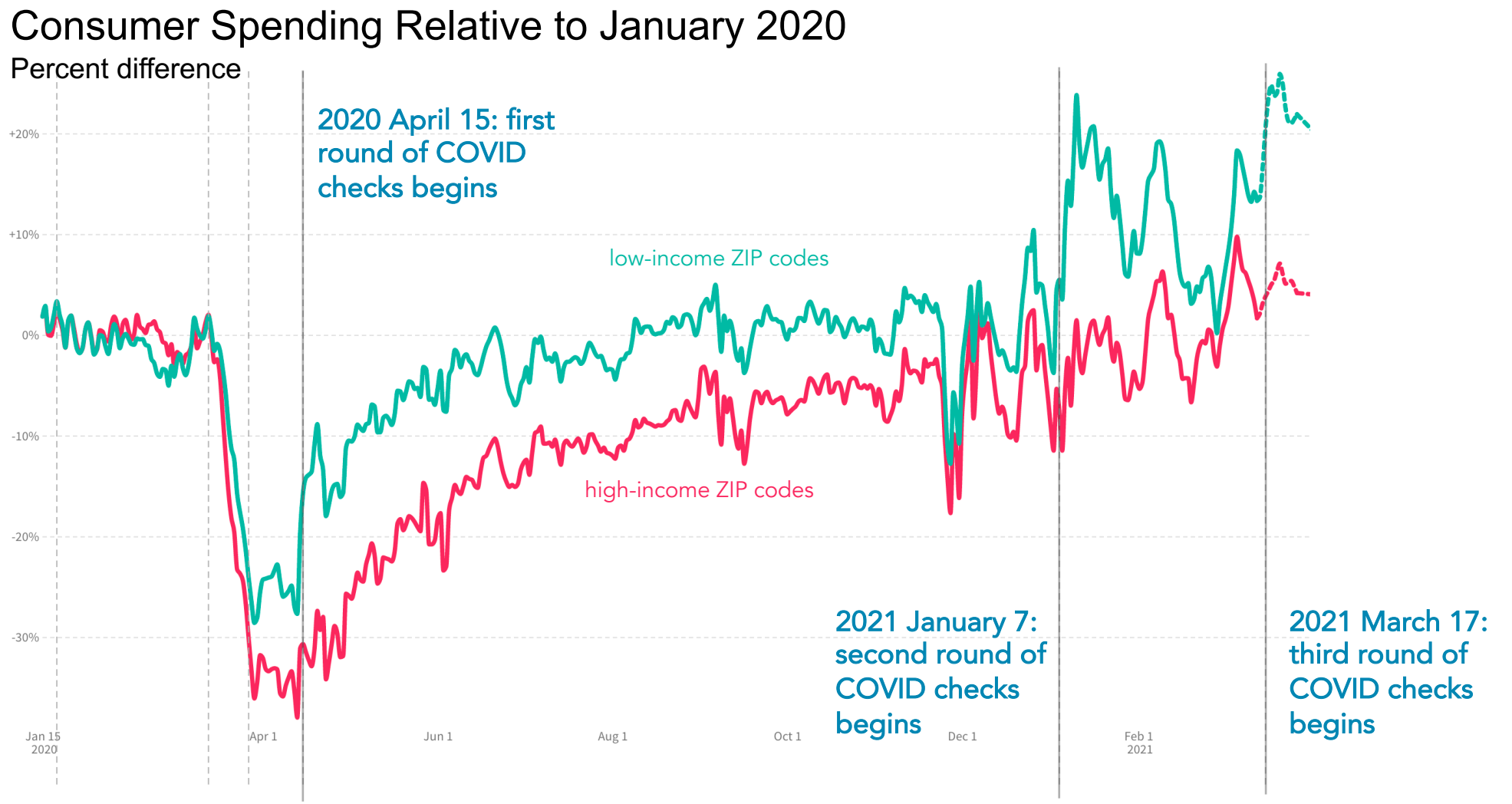
Many trackers are released weeks before traditional data and are available at a higher frequency

Using traditional data, policymakers would not have seen the degree of job loss in 2020 until mid May (the plunge in GDP would not be seen in government data until July)



Screenshot from Cajner et al (2022)

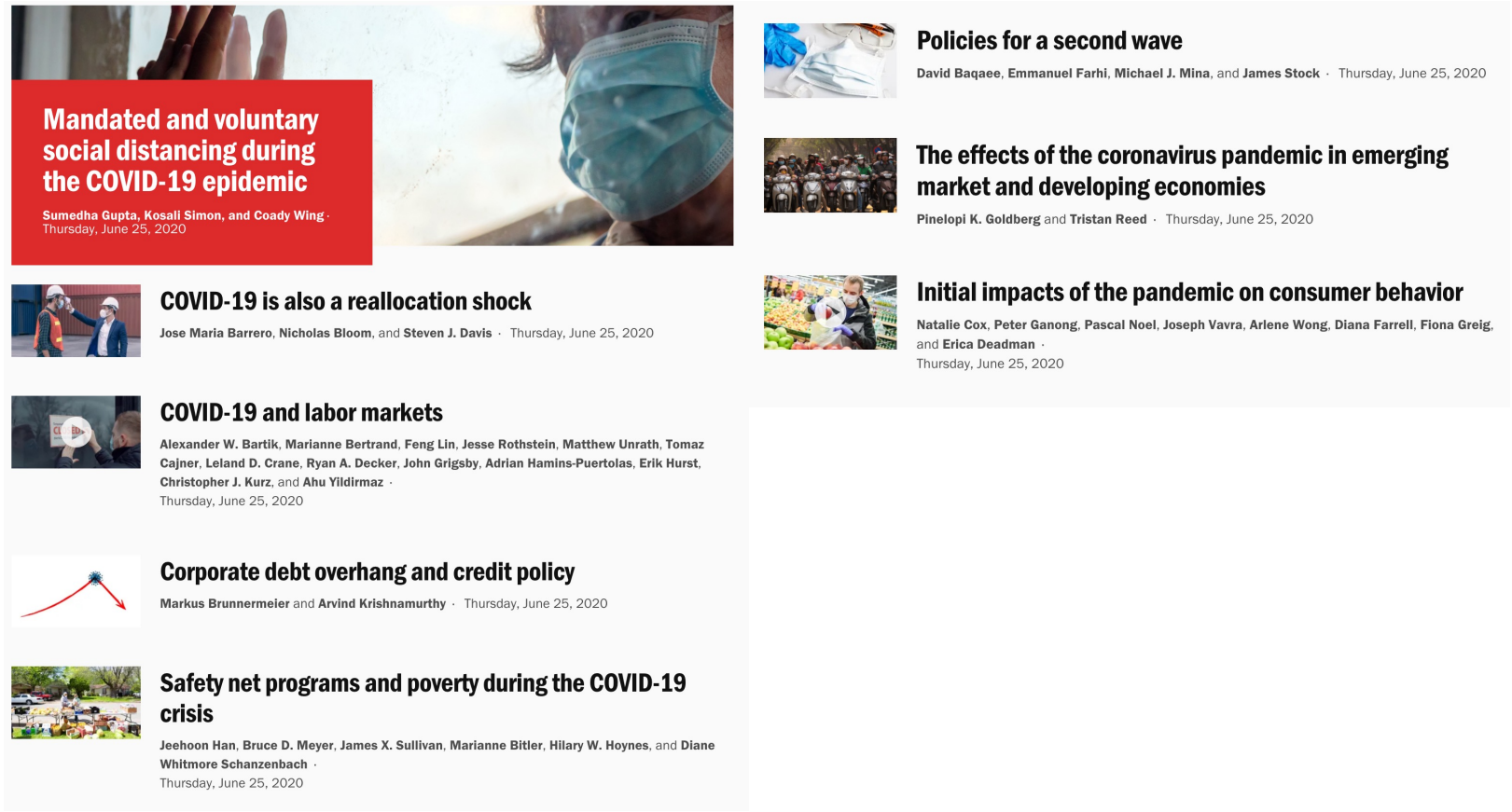
Many new trackers are more granular => visibility into the different experiences of different groups



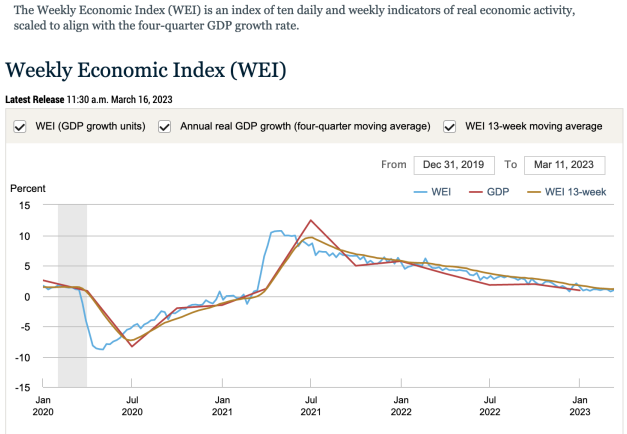
Data from tracktherecovery.org; based on credit and debt card data

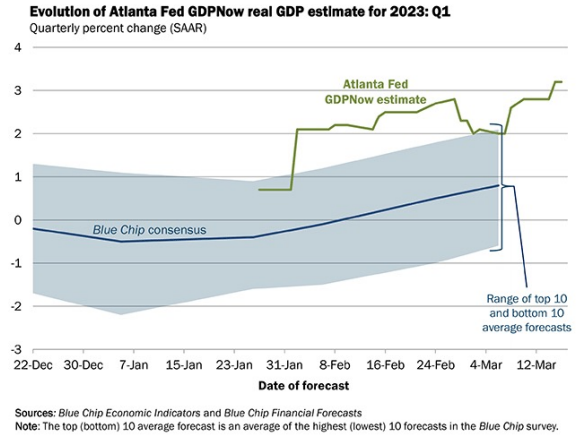
The underlying administrative records allowed for quick analysis of the policies deployed

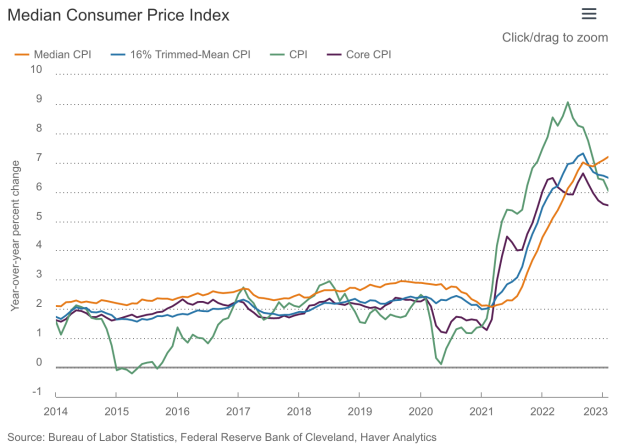
On June 25, 2020, the Brookings Papers on Economic Activity held a special conference featuring 8 new academic papers evaluating the COVID shock and policies to mitigate the economic damage



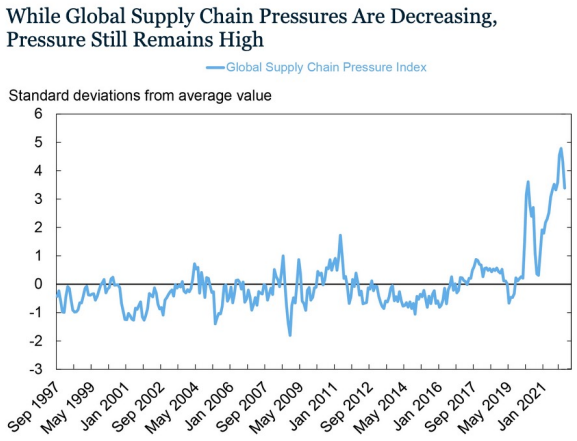
Statistical advances have led to “mash-ups” that have yielded additional insight into developments



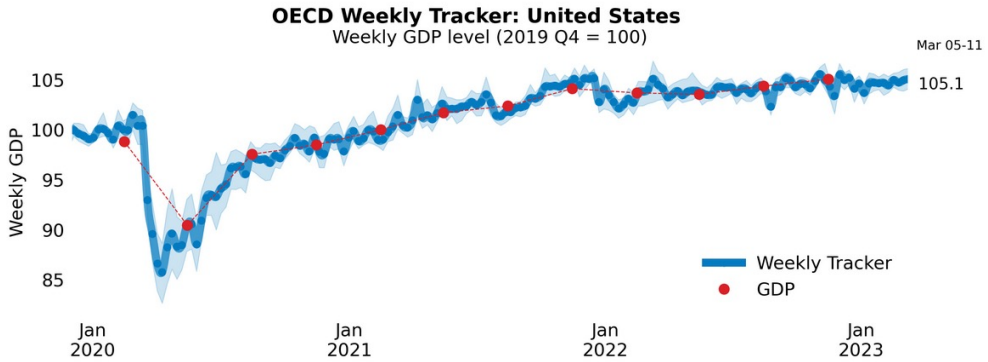




Source: Bureau of Labor Statistics, Federal Reserve Bank of Cleveland, Haver Analytics

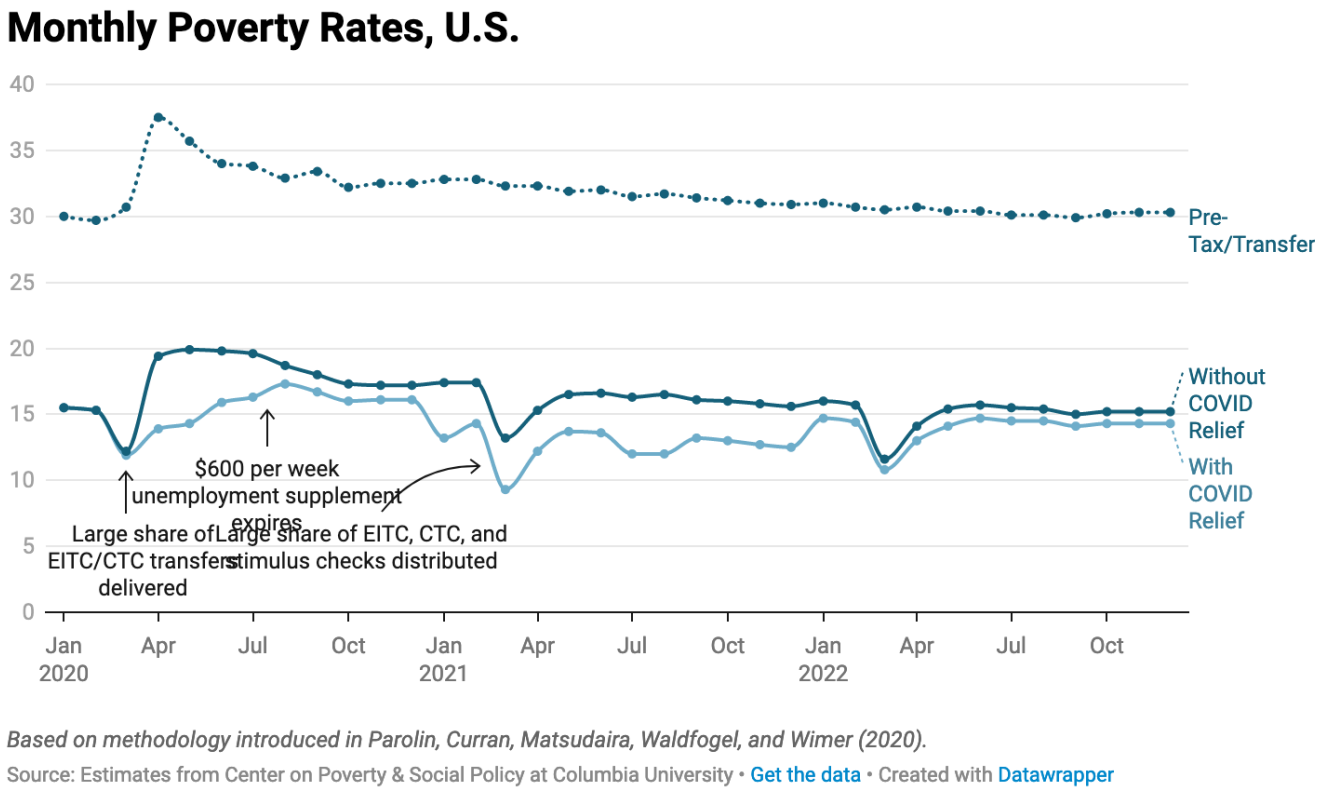


Sources: Bureau of Labor Statistics; Harper Petersen Holding GmbH; Baltic Exchange; IHS Markit; Institute for Supply Management; Haver Analytics; Bloomberg L.P.; authors' calculations.
Note: Each index is scaled by its standard deviation.



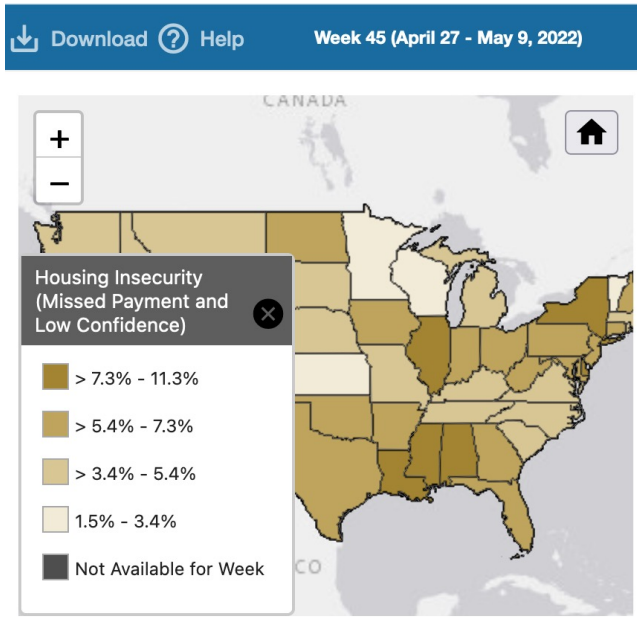
Note: The Weekly Tracker provides an estimate of weekly GDP based on Google Trends search data and machine learning.
Source: OECD Weekly Tracker (Woloszko, 2020), https://www.oecd.org/economy/weekly-tracker-of-gdp-growth; OECD Quarterly National Accounts.

For example, we now have a timely understanding of how macro conditions and policy are affecting hardship because of real-time estimates of poverty

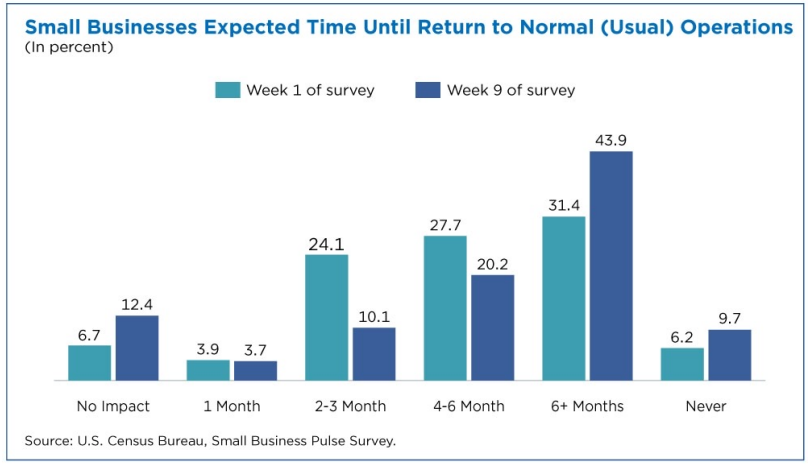


Screenshot from Columbia University Center on Poverty

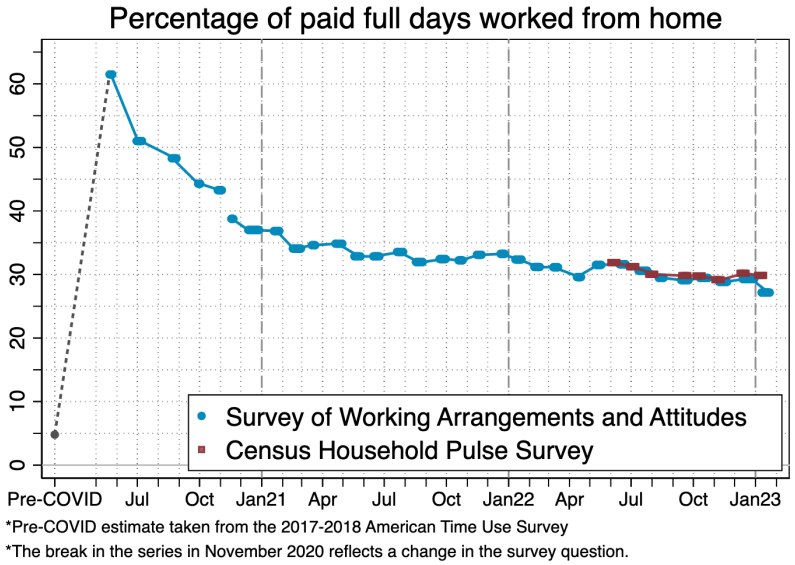
New data collections have been similarly valuable



Census Household Pulse

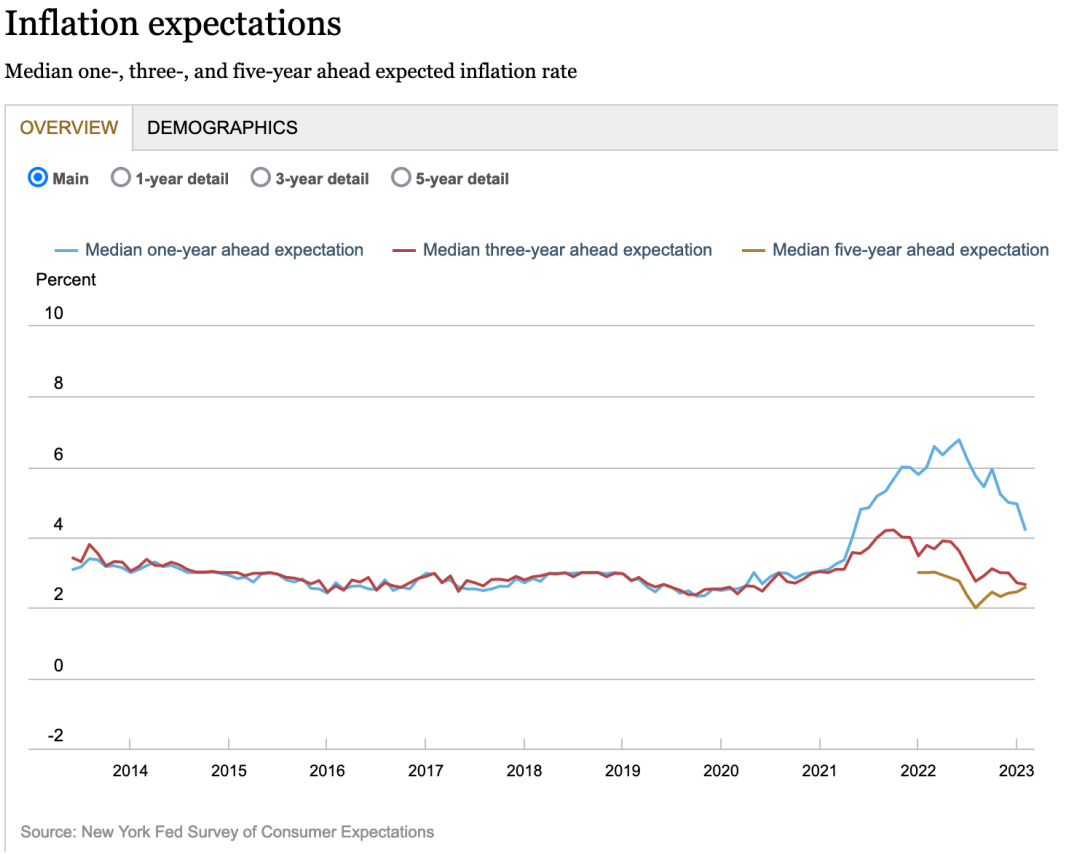


Census Small Business Pulse



WFH Research

The New York Fed's Survey of Consumer Expectations is yielding important insights about the risk of an inflationary spiral



Screenshot from New York Fed

The traditional data toolkit for economic policymakers

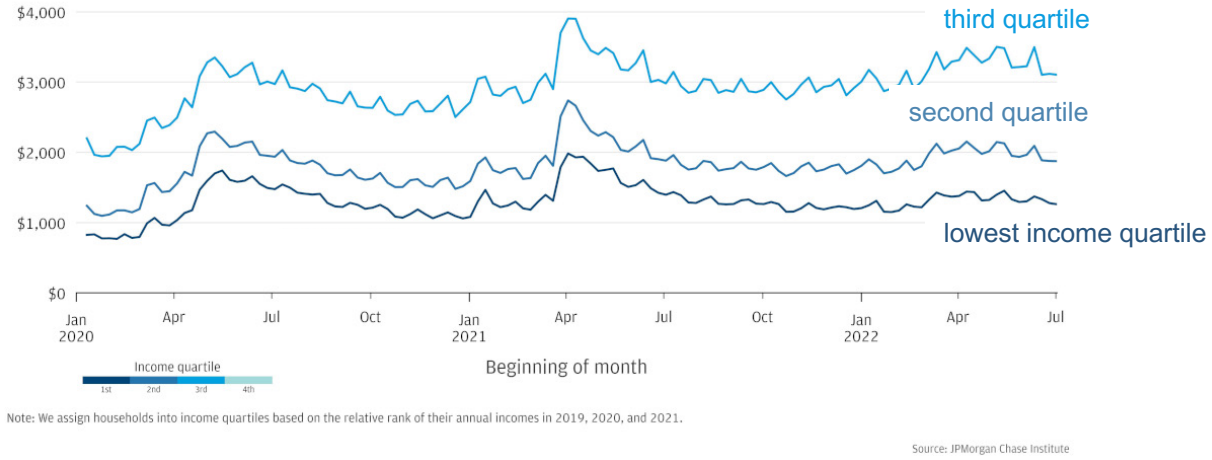
Important new additions to the economic data toolkit

Challenges associated with the new tools

Constructive next steps

Administrative sources are typically not designed to be representative

Data from the JPMorgan Chase Institute, show that median checking account balances were well above pre-COVID levels through mid-2022, even for the lowest income quartile (in this dataset)



To properly interpret, one needs to recognize that the lower-income customers of a large mainstream bank **are not the poorest Americans**

(According to the FDIC, about 5% of American households do not even have a bank account)

Administrative sources are typically not designed to capture economic concepts

Users of credit card records sometimes refer to new charges as “consumption”

But:

- People don’t typically buy cars or pay rent with credit cards
- Card charges may capture cash taken out; and the transfer of balances across accounts
- Health care consumption goes well beyond the out-of-pocket costs one might put on a card

Personal consumption expenditures	**17,360.4**
Goods	**5,939.6**
Durable goods	2,184.7
Motor vehicles and parts	723.0
Furnishings and durable household equipment	525.5
Recreational goods and vehicles	645.2
Other durable goods	291.0
Nondurable goods	3,754.9
Food and beverages purchased for off-premises consumption	1,277.4
Clothing and footwear	491.2
Gasoline and other energy goods	492.8
Other nondurable goods	1,493.6
Services	**11,420.8**
Household consumption expenditures (for services)	10,890.4
Housing and utilities	2,995.9
Health care	2,724.7
Transportation services	548.5
Recreation services	614.9
Food services and accommodations	1,253.3
Financial services and insurance	1,318.8
Other services	1,434.4

Components of personal consumption expenditures from the last GDP release

Representative issues and other issues lead to different studies yielding very different results

New data sources allowed for a burst of research exploring the effects of the Economic Impact Payments ("COVID checks") on consumer spending—Gelman and Stephens (2022) cite **10 important academic studies**

But the findings vary widely—from 75% of households saying they will use the EIPs to "mostly pay for expenses" to an estimated consumption response of about 10 cents for every dollar received

It is thus **unclear how informative the literature will be** to policy debates over future use of this type of support ¯_(ツ)_/¯

Other issues

Administrative data sets **may not be designed for longitudinal consistency**

For privately produced indicators, **the methodology used is often not transparent**

Some private providers **lack the training/experience** of staff at statistical agencies

> Dealing with seasonal pattern and other holiday issues, breaks in the sample, other types of data irregularities (e.g., a blip related to the move of Amazon Prime day) is complicated business!

Expense/legal restrictions/reputational risk means **unequal access** to (sometimes) indicators based on proprietary data and often the micro data

> "the ADP-FRB data have done a terrific job of tracking the employment gains seen in the BLS employment report,"
> "the ADP-FRB data are available on an ongoing basis *only to policymakers in the Federal Reserve System*"
>
> Cajner et al (2021)

More generally …

The **new abundance of indicators** based on different underlying populations, concepts, and methodologies raises the risk of **more confusion**

It also raises the risk of **more selective highlighting** of just the results that support the policy changes desired based on one's priors or self interest

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

The traditional data toolkit for economic policymakers

Important new additions to the economic data toolkit

Challenges associated with the new tools

Constructive next steps

A constructive next step for everyone

Increase the attention given to **interpretation, reconciliation, and accurate portrayal of new economic indicators and data sources** relative to that given to simply producing and casually disseminating them

Constructive next step for private actors

Private data producers can:

- Offer **more transparency** about what methods they are using
- Offer more **guidance about the limitations** of their measures

Those funding the production of new data and indicators (e.g., foundations supporting academic efforts) can:

- Ask producers to **adopt the high standards used by government statistical** agencies around transparency, methodologies, and other practices
- Ask producers to **release more data**

Constructive next steps for public actors (1)

Congress can pass **legislation ensuring that the statistical agencies have adequate funding** to maintain the quality of their existing statistics and innovate using new sources of data:

- More data production will require more staff
- Addressing falling response rates and privacy issues will require resources
- Proprietary administrative records can be very expensive

Other types of changes to the law would also be constructive—for example, passing a "data sync" proposal would enhance statistics by increasing the access of some agencies to tax data

Constructive next steps for public actors (2)

Statistical agencies considering changes that would restrict access (for example, to protect privacy) can **foster a constructive dialogue with users** about the costs and benefits of different approaches

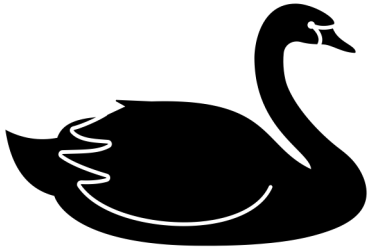
Regulatory agencies and other policymakers that have developed useful internal indicators to inform policymaking should **consider ongoing releases of the indicators they find important** to the public so as to foster a robust public conversation about policy decisions

For example, the ADP-FRB product mentioned earlier

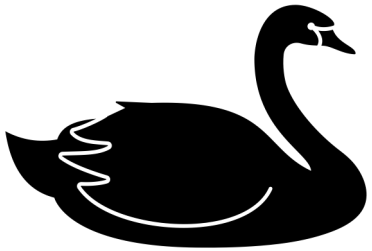
Improving the economic data tool kit is a high priority

Going into the 21st century, macroeconomic policymakers thought the economy had entered a “Great Moderation” and that financial crises only occurred in poorly run countries

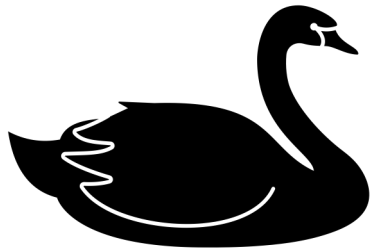
The past two decades have proved them wrong with 3 “Black Swan” crises:



The 2007-2009 financial crisis and Great Recession



The 2020 COVID and its economic fallout



The 2021 resurgence of inflation

Good economic data is **essential to informing policy that will reduce the odds of future crises** and **mitigate the damage** when they occur

Thank you!