BY TIM SABLIK

Good Data is Hard to Find

New challenges have emerged to the production of economic statistics.
How are Fed researchers and policymakers adjusting?

Fed officials frequently describe their monetary policy decisions as data dependent. As the central bank has navigated the recovery from the COVID-19 pandemic, a common refrain in its policy statements is that the Federal Open Market Committee (FOMC) will "carefully assess incoming data, the evolving outlook, and the balance of risks" when considering further adjustments.

"We are looking at the data to guide us in what we should do," Fed Chair Jerome Powell said at the press conference following the FOMC's meeting at the end of January.

The demand for data in economics as a whole has only grown in recent decades. A 2017 article in the *American Economic Review* found that the profession has become increasingly empirical since 1980, relying more on data analysis over theoretical models. This "empirical turn," as some economists have called it, has been facilitated by computerization, which has both increased the supply of data and aided in its analysis. At the same time, challenges around data quality and timeliness have emerged. How does the Fed ensure it's getting the best information to guide monetary policy?

SURVEYS TO THE RESCUE

For much of the 20th and 21st centuries, gold-standard U.S. economic data have been publicly produced. The federal government's entrance into the realm of data collection was driven by both public and private demand to better understand the industrializing economy. According to a 2019 article by Hugh Rockoff, an economic historian at Rutgers University, workers and employers wanted statistics on prices

in order to resolve mounting wage disputes in the late 19th and early 20th centuries. And lawmakers sought to better understand the ramifications of their policies as well as the evolution of the economy through the crises of the first half of the 20th century — two World Wars and the Great Depression.

The U.S. Bureau of Labor, later renamed the Bureau of Labor Statistics (BLS), was established in 1884 and produced its first indices of prices and wages in the 1890s. In 1918, the BLS conducted a national survey on the cost of living, releasing the results the following year. The BLS also started work on more frequent estimates of unemployment around the same time. Previously, national employment was measured only every decade as part of the census.

The newly formed Fed was an eager consumer of this new economic data.

"From its beginnings more than a century ago, the Federal Reserve has gone to great lengths to collect and rigorously analyze the best information to make sound decisions for the public we serve," Powell said in a 2019 speech.

The Fed was also a key early player in the dissemination of national economic data. According to a 2021 article by Diego Mendez-Carbajo and Genevieve Podleski of the St. Louis Fed, the Fed began publishing banking data the same year it opened its doors in 1914. In 1919, the same year the BLS released its first national cost of living estimates, the Fed Board of Governors began publishing monthly data on the manufacturing of several goods. In 1922, these data were collected into three monthly indexes capturing activity in manufacturing, mining, and agriculture. These measures of aggregate economic activity predate the concept

of gross domestic product, developed by economist Simon Kuznets in the 1930s, and are still updated and published today.

"The Federal Reserve System is an important producer of unique economic data and has recognized the value of sharing data with the public in an organic way that reflects its federated structure," says Mendez-Carbajo.

The government's rising interest in collecting better information about the economy coincided with advances in survey methodology. Robert Groves, director of the U.S. Census Bureau from 2009 to 2012 and currently interim president of Georgetown University, catalogued the history of survey research in a 2011 *Public Opinion Quarterly* article. The theory of probability sampling, or random sampling, developed in the 1930s offered researchers a means of using surveys to obtain bias-free inferences about a population.

Surveys provided, and continue to provide, the underlying data used in the calculation of many key economic indicators. Information about the labor force, including the unemployment rate and labor force participation rate, is collected via the monthly Current Population Survey (CPS) administered by the Census Bureau and the BLS. The Consumer Price Index, a commonly cited measure of inflation, is also computed using data gathered from surveys. In addition to households, the BLS also surveys businesses. Examples include measures of job openings and separations from the Job Openings and Labor Turnover Survey (JOLTS) and the Producer Price Index.

The Fed also uses surveys to collect national and regional economic information. For example, the Richmond

Fed launched its surveys of manufacturing and service sector activity in 1993 and continues to update them today. In addition, Fed policymakers look at the CFO Survey, which gathers insights from business leaders about the challenges and outlook for their own business and the overall economy. That survey, started by Duke University's Fuqua School of Business in 1996, has been conducted since 2020 by the Richmond and Atlanta Feds in partnership with Duke.

CRACKS EMERGE

In recent decades, however, researchers have faced mounting challenges to using surveys for data collection. One of the biggest is falling survey response rates. Early in the 20th century, most surveys were conducted face-to-face. From the 1960s to the 1990s, the proliferation of phones in households offered a new method for sampling large populations. While phones initially made it easier to reach survey participants, inventions like the answering machine and caller ID (which smartphones have made ubiquitous) made it easier for households and businesses to avoid such calls. The rise of phone and text scams may have also contributed to the growing unwillingness of individuals to respond to requests from unknown numbers. Finally, surveys may have become a victim of their own success. Between the 1980s and 2000s, the number and length of government and private surveys exploded. Some researchers suggest that this has led to survey fatigue among households, contributing to lower response rates.

The COVID-19 pandemic only intensified these trends. Even response rates from businesses, which had generally been more robust than household response rates, dropped sharply. In January 2020, the JOLTS response rate was 58 percent. In April 2020, it fell by about 10 percentage points and never recovered; as of September 2024, it was 33 percent. On the household side, the CPS response rate did recover after
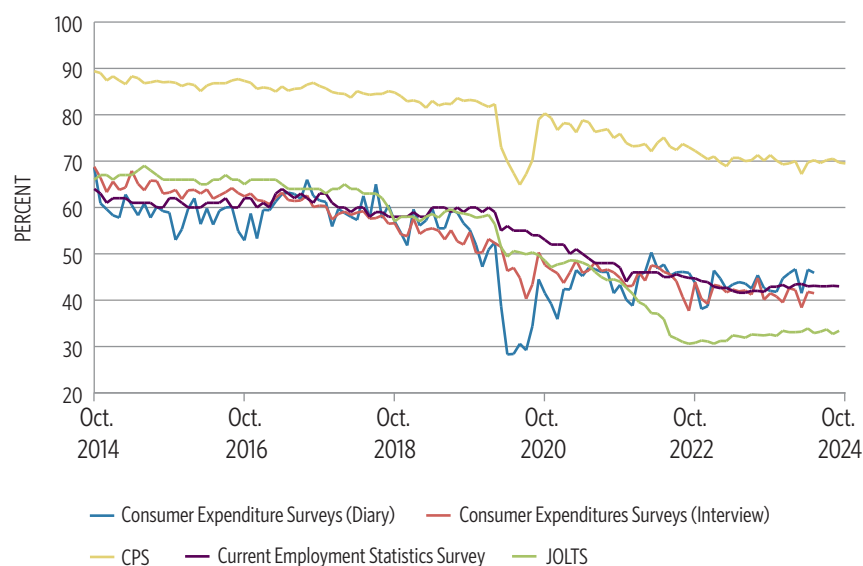
the initial COVID-19 shock, but it has continued a longer-running decline. It was nearly 70 percent in October 2024, roughly 20 percentage points lower than a decade earlier. (See chart.)

A 2015 article in the *Journal of Economic Perspectives* by Bruce Meyer of the University of Chicago, Wallace Mok of the Chinese University of Hong Kong, and James Sullivan of the University of Notre Dame highlighted other problems. The likelihood that survey respondents fail to answer each question, known as item nonresponse, has gone up. So has measurement error, which is when respondents provide inaccurate information. This is a particular problem for opt-in online surveys. Such surveys are typically cheaper and easier to produce, but they don't capture a true random sample, limiting the conclusions researchers can draw about the larger population. Work by Andrew Mercer, Courtney Kennedy, and Scott Keeter of Pew Research Center found that online survey participants who report being under the age of 30 are particularly likely to be what the researchers called "bogus respondents." In one opt-in survey, 12 percent of respondents ages

18 to 29 said they were licensed to operate a nuclear submarine.

These trends, alongside rising nonresponse rates, have increased worries about the introduction of bias into survey results. Researchers at the BLS and elsewhere track this issue carefully and have statistical methods of adjusting for lower response rates. Nevertheless, obtaining an adequate sample to produce unbiased insights even with these methods is becoming more difficult.

"Survey sponsors are finding it harder to obtain survey cooperation," says Jonathan Mendelson, a research statistician at the BLS. "This can increase the level of effort necessary to obtain interviews, which can potentially lead to increased data collection costs."

In 2023, the BLS announced plans to modernize the CPS to address falling response rates. This five-year plan includes careful testing of different surveying methods, culminating in the introduction of an online self-response mode by 2027. Such adjustments take time and resources, and according to a 2024 article from the Center for American Progress (a progressive think tank), the budget of the BLS

Survey Participation Wanes
BLS establishment and household survey response rates

SOURCE: U.S. Bureau of Labor Statistics

Legend:
— Consumer Expenditure Surveys (Diary)   — Consumer Expenditures Surveys (Interview)
— CPS   — Current Employment Statistics Survey   — JOLTS

has been shrinking in real terms since 2010. In the face of these financial constraints, BLS officials have said they might be forced to start shrinking the CPS sample. In October 2024, the BLS announced that such plans were on hold for now but that they could still happen in the future depending on the budget situation.

Researchers at the Fed have also grappled with constructing good survey samples amid declining response rates. Jason Kosakow, the Richmond Fed's survey director, published an article with Pierce Greenberg of Clemson University examining the effectiveness of different strategies for recruiting participants for the Richmond Fed business surveys via email. They found that a standard notification with no appeal to the benefits of taking the survey worked best, but conversion rates were still low — less than 2 percent. Kosakow is also working with researchers at the Richmond Fed to collect better information on Fifth District businesses using multiple data sources. This helps ensure that surveys are capturing a truly representative sample of regional business voices.

"The number one thing you need to do when creating a quality survey is have a good sample frame," says Kosakow. "You want it to be reflective of your population. And that's really hard to do, because people respond at different rates. So, one way to improve surveys is to use different technologies to find people or businesses who are less likely to respond, to mitigate these issues."

THE PROMISES AND PITFALLS OF BIG DATA

These challenges can increase the likelihood that preliminary economic indicators are subject to significant revisions later as new data become available. Last August, the BLS revised the number of jobs created from April 2023 to March 2024 down by more than 800,000. Such revisions pose a clear challenge for monetary policymakers trying to get a real-time picture of the economy to guide their decisions.

This has led Fed researchers to explore alternative data sources. In addition to helping survey-based research, the growing computerization of household and business activity has led to an explosion of new economic data. Often referred to as "big data," these datasets offer the potential to give researchers a much more granular and timelier snapshot of economic activity.

During the initial weeks and months of the COVID-19 pandemic, researchers across the Federal Reserve System turned to a variety of such nontraditional data sources to get a better understanding of what was happening to the economy. According to a 2022 book chapter by Tomaz Cajner, Laura Feiveson, Christopher Kurz, and Stacey Tevlin of the Fed Board of Governors, Fed researchers looked at employment data from payroll processors, retail sales from Fiserv card swipe data, restaurant reservations from OpenTable, and airport departures from the Transportation Security Administration, among other nontraditional data sources.

"Alternative data can help provide an additional signal that can either corroborate or question the indications coming from preliminary official statistics," says John O'Trakoun, a senior policy economist at the Richmond Fed. "In the case of high-frequency data, it can help provide a sneak peek of turning points or changes in momentum that the standard data would not be able to show until well after the fact."

Even outside of crises, Fed researchers are exploring how non-survey data might improve their ability to forecast changes in economic conditions. In a February article in *Economics Letters*, O'Trakoun and Adam Scavette of the Philadelphia Fed developed a new recession indicator based on the Sahm rule, which was created in 2019 by economist Claudia Sahm. The Sahm rule uses changes in the three-month moving average of the unemployment rate to predict the start of recessions. Rather than using the unemployment rate, which is based on responses to the CPS, O'Trakoun and Scavette used state claims for unemployment insurance. These are administrative data that are released weekly, while the survey-based unemployment rate is updated monthly. O'Trakoun and Scavette found that using these data improves the timeliness and accuracy of the Sahm recession indicator.

Alternative data sources can come with their own set of challenges, however, as highlighted by Cajner, Feiveson, Kurz, and Tevlin in their account of data lessons learned from the COVID-19 pandemic. They noted that a lot of big data are the byproduct of economic activity, meaning that they typically aren't collected to answer a particular research question. Therefore, it can require more work from researchers to understand the data well enough to extract useful insights about a larger population. Data collected by private companies are also typically not made freely available to the public, potentially making them expensive for researchers at policymaking institutions to access. Data owners may also place conditions on how the data can be used, limiting analysis. Finally, nontraditional data series may be new, making historical comparisons and seasonal adjustments difficult. This can make it hard to know how well these data series perform relative to traditional sources over the long run.

This latter challenge can apply to newer government statistics as well. The Business Formation Statistics data series was created by the Census Bureau in 2004. It provides information on filings for Employer Identification Numbers (EIN), a tax identification number used by businesses. Researchers at the Fed and in academia have explored using the Business Formation Statistics as an indicator of business and entrepreneurial activity, since individuals planning to start a new business often

file for an EIN. During the COVID-19 pandemic, there was a significant surge in EIN applications, suggesting an uptick in new business formation. As new businesses tend to grow faster than older ones, this presented the possibility for a wave of innovation and hiring. But subsequent research by Chen Yeh, a senior economist at the Richmond Fed, found that much of this new entry was concentrated in industries with low or even negative productivity growth, suggesting a modest impact on overall productivity. The short history of the Business Formation Statistics made it hard to discern in real time whether the COVID-19 episode was representative of past spikes in EIN filings.

All told, the trade-offs inherent to big data make it most likely to serve as a complement to surveys rather than a replacement.

"I don't think surveys are going to go away," says Mercer of Pew Research. "What we're going to see, and are already seeing, is increasing use of big data to improve the quality of survey estimates."

STAYING DATA DEPENDENT

In 2011, the FOMC introduced calendar-based forward guidance into its policy statement. The United States was in the midst of a slow recovery from the Great Recession, and the FOMC wanted to communicate its expectation that monetary policy would likely remain accommodative for at least a couple more years. Although this was intended to communicate the committee's expectations about future economic conditions and appropriate policy, some Fed watchers took it as a commitment to keep rates low for a prescribed period regardless of the data. In late 2012, the committee clarified this, changing the wording in the statements to more clearly indicate that future policy decisions would depend on economic data, not dates.

Fed policymakers have given little indication that they plan to deviate from this data-driven approach, despite the challenge of piecing together an accurate picture of the economy from various imperfect indicators. Members of the FOMC have spoken about how they weigh the strengths and weaknesses of each incoming data point, incorporating them into their own views of the economy. Meanwhile, researchers at the Fed and federal statistical agencies continue to explore new sources and methods for generating more accurate inputs to that process.

"Despite the many challenges, the future of economic measurement is bright," Fed Gov. Adriana Kugler said in a July 2024 speech at the National Association for Business Economics Foundation. "The statistical agencies have already proven their ability to innovate and adapt, even under tight resource constraints. And the wealth of private-sector data sources will only expand in the future." **EF**

READINGS

Cajner, Tomaz, Laura Feiveson, Christopher Kurz, and Stacey Tevlin. "Lessons Learned from the Use of Nontraditional Data during COVID-19." In Wendy Edelberg, Louise Sheiner, and David Wessel (eds.), *Recession Remedies: Lessons Learned from the U.S. Economic Policy Response to COVID-19*. The Hamilton Project and the Hutchins Center on Fiscal and Monetary Policy at Brookings, April 27, 2022.

Groves, Robert M. "Three Eras of Survey Research." *Public Opinion Quarterly*, December 2011, vol. 75, no. 5, pp. 861-871.

Mendez-Carbajo, Diego, and Genevieve M. Podleski. "Federal Reserve Economic Data: A History." *American Economist*, March 2021, vol. 66, no. 1, pp. 61-73.

Meyer, Bruce D., Wallace K. C. Mok, and James X. Sullivan. "Household Surveys in Crisis." *Journal of Economic Perspectives*, Fall 2015, vol. 29, no. 4, pp. 199-226.