

June 2023

## INSIDE THIS ISSUE

- The Amount and Structure of Taxpayer Wealth in Foreign Accounts
- Bank Liquidity and the Dynamics of the Fed's Balance Sheet
- Estimating the Macroeconomic Impacts of Fed Policies
- Dallas Initiatives Centered on Teacher Pay Show Strong Results
- Mergers in Consumer Packaged Goods and Consumer Prices

## Measuring the Productivity Impact of Generative AI

Customer support agents using an AI tool to guide their conversations saw a nearly 14 percent increase in productivity, with 35 percent improvements for the lowest skilled and least experienced workers, and zero or small negative effects on the most experienced/most able workers, [Erik Brynjolfsson](#), [Danielle Li](#), and [Lindsey R. Raymond](#) report in [Generative AI at Work](#) (NBER Working Paper 31161).

Using call data from roughly 5,000 agents working for a Fortune 500 software company, the researchers tracked the duration, quality, and outcome of customer support interactions as the company introduced a Generative Pre-trained Transformer (GPT) AI tool. The tool was rolled out to the agents gradually, mostly between November 2020 and February 2021. For a control group, the researchers also collected data from agents who did not receive the tool over 2020 and 2021. The AI tool was intended to support the work of human customer support agents, offering them potential responses to customer queries. The agents could choose to take those suggestions or ignore them and enter their own responses.

The researchers find that customer support agents utilizing the AI tool increased the number of customer issues resolved per hour by 13.8 percent. They attribute the increase to three factors: agents, who could participate in

improvements did not come at the expense of interaction quality.

The researchers divide the data by agents' length of tenure and pre-AI productivity, and find that the benefits of using the AI tool were

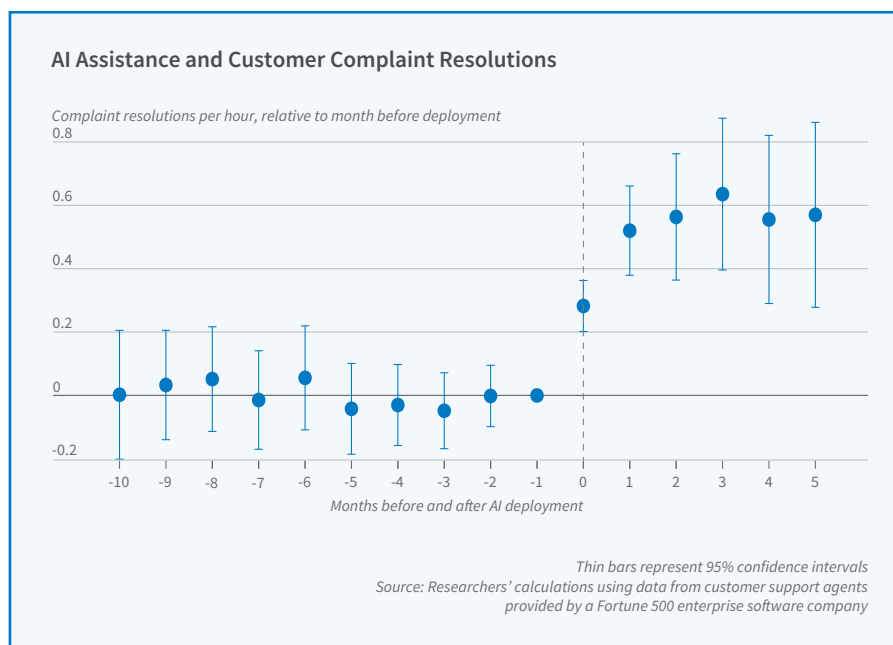
With AI assistance, customer service agents could handle more calls per hour and increase their resolution rate.

multiple chats at once, spent about 9 percent less time per chat, handled about 14 percent more chats per hour, and successfully resolved about 1.3 percent more chats overall. Measures of customer satisfaction showed no significant change, suggesting that the productivity

greatest among less experienced and lower skill workers, who saw gains of 35 percent, with little to no negative effects on top performing/most experienced workers. An agent using the AI tool who had just two months' tenure at the firm performed as well as an agent with

six months' tenure working without the tool. The researchers suggest that newer and lower skilled workers may have more to learn than higher skilled and more established workers, and that AI tools can help them adopt the skills and behavior of more experienced workers more quickly. Text analysis of agents' conversations supports this interpretation.

All agents changed their communication patterns after beginning to use the AI tool, but the change among lower-per-



forming agents was greater. This may be because the AI tool based its suggestions on the work style and outputs of the company's most productive agents, and therefore spread their pattern of behavior to newer and less skilled workers. For instance, the developers of the AI tool found that top performers were able to determine the underlying technical issue, based on a customer's description, twice as fast as lower performers. The AI tool, trained using the best

examples of resolved queries, learned to connect specific query phrases to useful diagnostic questions and potential solutions. The AI tool was also able to give more frequent feedback than a human manager. This gave new hires and lower performers the opportunity to improve faster than they would have without the tool, iterating with each call rather than only following managerial reviews.

The researchers also noticed that custom-

ers were more likely to express positive sentiments, and less likely to request help from a supervisor, when interacting with agents using AI assistance than when interacting with those who were not. Perhaps reflecting the improved tenor of the exchanges, attrition rates among agents with access to the AI tool were 8.6 percent lower than the comparable rates for agents without such access.

—Emma Salomon

## The Amount and Structure of Taxpayer Wealth in Foreign Accounts

Since 2015, the Foreign Account Tax Compliance Act (FATCA) has significantly expanded the amount of information available to the Internal Revenue Service (IRS) regarding wealth held by US taxpayers in foreign accounts. In *The Offshore World According to FATCA: New Evidence on the Foreign Wealth of US Households*, (NBER Working Paper 31055), Niels Johannesen, Daniel Reck, Max Risch, Joel Slemrod, John Guyton, and Patrick Langetieg document the size, distribution, and location of these funds.

FATCA requires foreign financial institutions (FFIs) to report to the IRS when any individual US taxpayer has an account worth more than \$50,000. The researchers examine the information returns filed by the FFIs—Form 8966—between 2015 and 2018. They exclude records containing no financial information, duplicate records, and a small number of observations with extremely large and suspect dollar values. The number of reported accounts grew substantially during this time period: for tax year (TY) 2015, their data include 178 countries, 27,000 institutions, and over 2 million accounts totaling \$1.6 trillion in assets; in TY2018, there were reports from 190 countries, 45,000 institutions, and about 4.6 million accounts totaling \$3.6 trillion in assets.

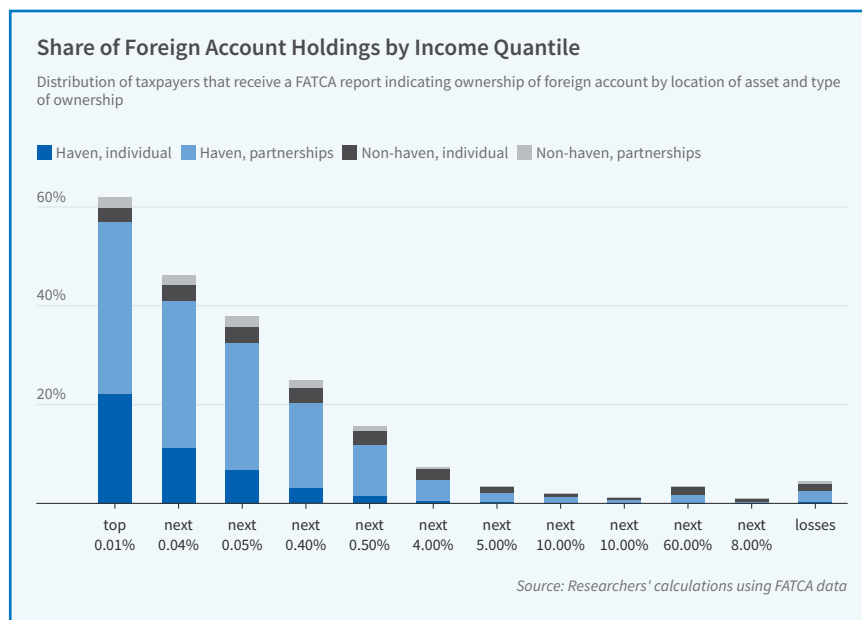
FFIs often report only partial owner information. Among forms with complete Taxpayer Identification Numbers (TINs), there were 790,000 distinct owners in TY2015, and 1.5

million distinct owners in TY2018. Moreover, there were many accounts without income information. The forms with available income information in 2018 showed \$13.2 billion in reported interest income, \$28.5 billion in dividends, \$274 billion in gross proceeds and redemptions, and \$208 billion in other income.

Most owners of offshore wealth are sophisticated global investors or US taxpayers with close ties to specific foreign countries, such

as recent immigrants or expats. Although the median account was not held in a tax haven, larger accounts were more likely to be held there and the researchers find that most of the wealth was owned by sophisticated investors and stored in tax havens and/or owned indirectly by a partnership. The 55 percent of accounts owned by identifiable individuals held 16 percent of total wealth reported on Form 8966, while the 1.4 percent of accounts owned by partnerships held 32 percent. Partnerships owned 52 percent of the \$1.9 trillion in total wealth that was reported in tax havens, but only 14 percent of wealth in other locations.

Partnerships are pass-through business entities, meaning income and losses are distributed to and taxed as part of the income of the partners. The researchers allocate reported foreign assets in proportion to each partner's share of the partnership's income; when the partnership is owned by another partnership, they repeat this process until the assets are allocated to an ultimate owner. The data show that 43 percent of assets held by partnerships were owned by US indi-



viduals, 10 percent by foreign individuals or entities, 8 percent by tax-exempt organizations, and 8 percent by trusts. Twenty percent had no identifiable beneficial owner. When they combine their information on the ownership

of foreign accounts, both directly and via partnerships, with taxpayer income on individual tax returns (IRS Form 1040), the researchers find that approximately 22 percent of those in the top 1 percent of the income distribution,

and 62 percent of those in the top 0.01 percent, held foreign accounts. Twenty-three percent of all reported foreign assets were held in tax havens by taxpayers in the top 0.01 percent. —Whitney Zhang

## Bank Liquidity and the Dynamics of the Fed’s Balance Sheet

In the wake of the global financial crisis of 2007–09, the Federal Reserve embarked on an ambitious program of quantitative easing (QE), purchasing large quantities of bonds held by commercial banks and the rest of the economy. QE gave banks an infusion of liquidity by exchanging their longer-term assets for liquid reserves. On the liability side, however, there was an expansion of demand deposits at banks, especially uninsured deposits such as checking accounts that depositors can withdraw from at any time. Banks also responded by originating off-balance-sheet liabilities in the form of lines of credit, which allow businesses to draw funds, up to a limit, at their discretion. Both demand deposits and credit lines are claims on the bank’s liquidity, and providing them without placing the bank at risk requires an appropriate stock of bank liquidity.

In 2017, the Fed reversed its QE policy and embarked on quantitative tightening (QT), shrinking its balance sheet by actively selling bonds for reserves, reducing Fed-provided liquidity in the banking system. While banks expanded their issuance of claims on liquidity during QE, they did not rein in such issuance when the Fed shifted to QT. Over time, as banks’ reserves diminished with QT, their ratio of outstanding claims on liquidity to reserves increased. In effect, the banks held constant the amount of withdrawals that depositors and firms could demand from them, even though their liquid reserves declined. This left the banks vulner-

able to episodes of liquidity stress, in which many claimants demand their money at one time and banks have too little liquid cash to satisfy everyone’s claims. Indeed, two such episodes occurred during the QT period, in

As the Fed expanded its balance sheet, banks provided more liquidity to households and firms, but when the Fed retrenched, banks did not reduce their provision of liquidity and became vulnerable.

September 2019 and in March 2020. In both cases, the Fed intervened in financial markets to provide additional liquidity.

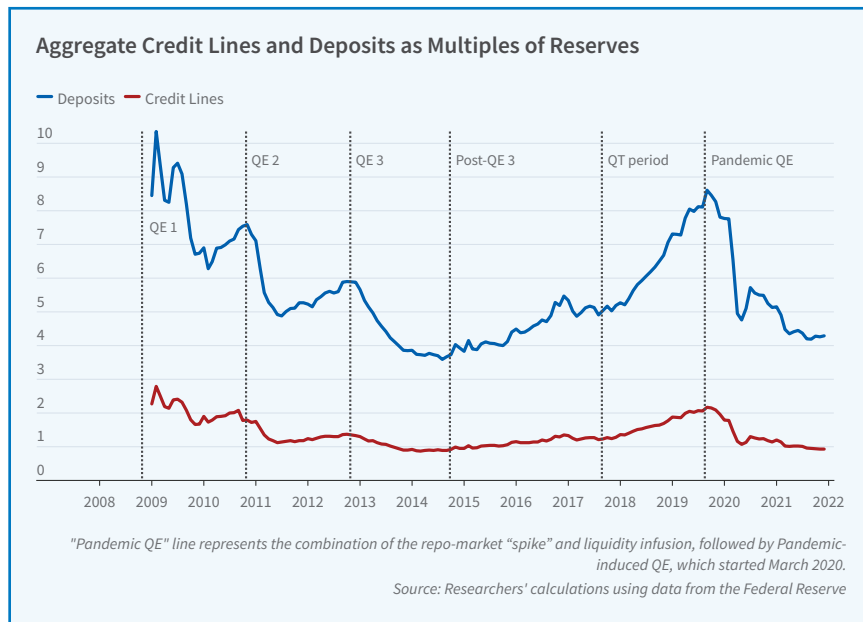
The researchers demonstrate, using bank-level data, that increases in banks’ central bank reserves during QE were associated with increases in the volume of demand deposits and credit lines. In contrast, they do

not find that the banks that provided customers with more claims on liquidity during QE reduced those claims when QT kicked in. This dynamic was particularly driven by less-well-capitalized banks, which maintained their provision of liquidity claims to customers even as their reserves declined. These banks may have been less risk averse on account of their weaker balance sheet positions.

The researchers conclude that QE and QT are not mirror images of each other. The two policies change banks’ liquidity-seeking behavior in different ways, and differently for different banks.

This asymmetry is also manifest in the recent failures of banks such as Silicon Valley Bank and Signature Bank that grew their uninsured deposit base during the pandemic QE but did not decrease it sufficiently during the monetary tightening phase of 2022. Such asymmetric bank behavior complicates the use of the Federal Reserve’s balance sheet as a tool for monetary stimulus.

—Shakked Noy



In [Liquidity Dependence and the Waxing and Waning of Central Bank Balance Sheets](#) (NBER Working Paper 31050), [Viral Acharya](#), [Rahul Chauhan](#), [Raghuram Rajan](#), and [Sascha Steffen](#) chronicle the evolution of both central bank balance sheet policy and banking sector issuance of liquidity claims over the last 15 years and develop tests for the causal pathways between the two.

# Estimating the Macroeconomic Impacts of Fed Policies

How monetary policy actions affect the trajectory of the macroeconomy is a key question for central bankers and macroeconomists. Researchers face formidable challenges in answering this question because of the endogeneity of the Federal Reserve's decisions. If these decisions were made randomly, researchers could simply examine the statistical correlation between the Fed's actions and subsequent economic outcomes to uncover the causal effects of decisions on outcomes. But of course, the Fed's decisions are not made randomly; they are responsive to both contemporaneous and anticipated economic conditions. This makes it difficult to disentangle their effects from the other factors that simultaneously influence both economic outcomes and Fed decisions. If inflation stays constant after the Fed raises interest rates, is this because the Fed's decision had no effect, or because the Fed raised rates in anticipation of inflation-increasing developments, implying that inflation would have increased absent the Fed's rate hike?

One way to solve this empirical challenge is what [Christina D. Romer](#) and [David H. Romer](#) term "the narrative approach." In [Does Monetary Policy Matter? The Narrative Approach after 35 Years](#) (NBER Working Paper 31170), they apply this methodology to historical data and draw inferences about the likely impact of recent monetary policy actions.

The narrative approach focuses on identifying and analyzing historical monetary policy decisions that were not influenced by contemporaneous or expected economic outcomes. Studying these episodes, and comparing macroeconomic conditions before and after the policy

changes, can yield causal evidence on the effects of policy changes.

To identify relevant policy changes, the researchers note that Fed decisions are

---

Empirical estimates of the effects of previous monetary contractions suggest that recent rate hikes will reduce employment, output, and inflation.

---

influenced by two factors: policymakers' assessments of current and future economic conditions, and their preferences over inflation and unemployment. Sometimes their preferences change in a way unrelated to contemporaneous economic activity, which results in a monetary policy change. For example, in January 1972, policymakers decided to prioritize a reduction

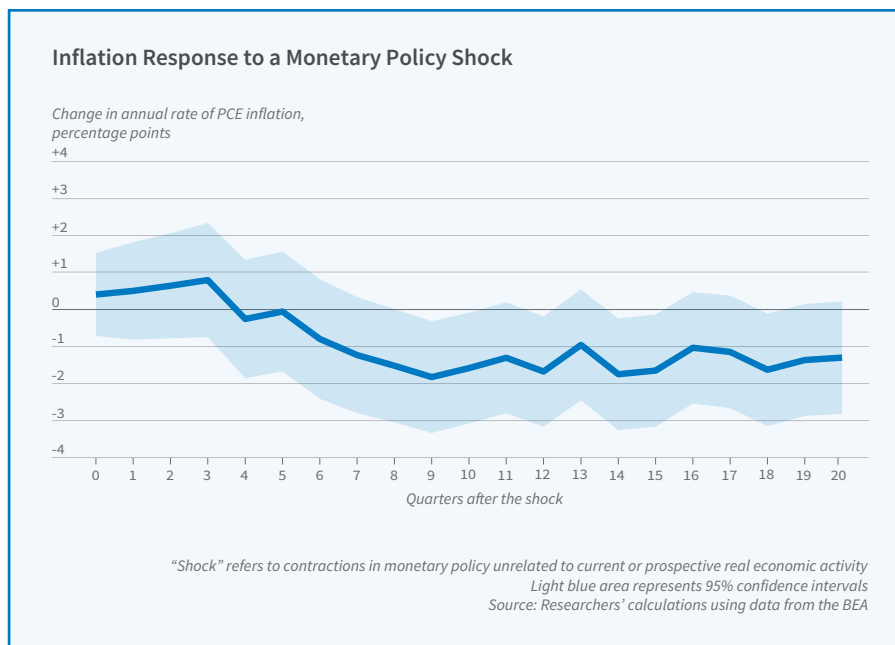
in the postwar period in the United States attributable to changing policymaker preferences, and reading these transcripts makes it possible to identify the reasoning behind monetary policy decisions. The researchers identify 10 monetary policy changes

in the postwar period in the United States attributable to changing policymaker preferences, and study the evolution of economic outcomes around these changes.

They find that the average effect of a contractionary monetary policy shock is an increase in the unemployment rate, with the effect peaking at a 1.6 percentage point increase about two years after the shock. Real GDP, meanwhile, falls, with the effect peaking at 4.4 percent after about two years. The estimates are highly statistically significant. Both effects dissipate by five years after the policy shock. Inflation, meanwhile, falls after a contractionary shock, though these effects are less precisely estimated. Five years after the policy tightening, the inflation rate is about 1.5 percentage points lower than it would have been otherwise.

Although the transcripts of the 2022 FOMC meetings will not be available until 2028, when the researchers apply their methodology to the FOMC documents from 2022 that are currently available, they conclude that there was likely a contractionary monetary shock in the third quarter of 2022. Their estimates imply that the associated interest rate hikes will reduce employment, output, and inflation, but that most of the effects will not occur until mid-2023 and after.

—Shakked Noy



in the then-stable unemployment rate at the expense of potentially higher inflation; they lowered interest rates. Conversely, in December 1988, policymakers decided the inflation rate was intolerably high and hiked rates.

The researchers identify such changes by reading the minutes and transcripts of Federal Open Market Committee (FOMC) meetings. Since these transcripts are released publicly only after five years have passed, meeting participants can speak freely about their expectations and prefer-

# Dallas Initiatives Centered on Teacher Pay Show Strong Results

Over the last decade, the Dallas Independent School District (ISD) has dramatically changed how it sets salaries. Jettisoning a typical pay scale tied to years of experience and academic credentials, Dallas began compensating educators based on a rigorous evaluation system. The result, according to two recent, related studies, is a marked improvement in student achievement.

In [The Effects of Comprehensive Educator Evaluation and Pay Reform on Achievement](#) (NBER Working Paper 31073), [Eric A. Hanushek](#), [Jin Luo](#), [Andrew J. Morgan](#), [Minh Nguyen](#), [Ben Ost](#), [Steven G. Rivkin](#), and [Ayman Shakeel](#) show that the new system increased the achievement levels of Dallas students above those of students enrolled in comparable schools with traditional pay scales elsewhere in Texas.

Under the Principal Excellence Initiative, introduced in 2013, and the Teacher Excellence Initiative, which followed two years later, the Dallas ISD pays educators based on their contributions to student achievement, supervisor observations, student or family feedback, and, in the case of principals, efforts to support teacher improvement. As an incentive to focus on disadvantaged students, principals are also judged on progress in reducing achievement gaps between their students and the district mean. Based on aggregate evaluation scores, educators are sorted into rating bins that are the primary factor in setting their salaries.

As their main control group, the researchers used schools drawn from the 20 largest Texas districts with at least 60 percent low-income students. The impact of the Dallas reforms became clear in 2016 when, after an initial period of high teacher turnover, math and to a lesser extent reading achievement scores rose. Prior to introduction of the new

salary scheme, scores in both the Dallas schools and the control group, which serve large shares of disadvantaged students, were substantially below the statewide mean; afterward, the stu-

Identifying effective teachers and principals and providing financial incentives for them to work in lagging schools raises achievement test scores in math and reading.

dents in Dallas rose to nearly the state mean level, while the control group did not improve.

Though initially disruptive, the teacher turnover resulted in a stronger staff, as teachers who left the system on average had lower evaluation scores than those who remained. The analysis attributes 15 percent of the improvement in math achievement to changes in the composition of the teaching staff. Other key contributing factors included higher incen-

The Accelerating Campus Excellence (ACE) program was launched with four schools in 2016 and expanded to nine schools in 2018. Teachers were paid effectiveness-

based stipends of up to \$10,000 and administrators up to \$13,000 for working at low-performing schools. Existing teachers who wanted to remain at the targeted schools underwent a rigorous screening process, resulting in less than 20 percent of them being retained.

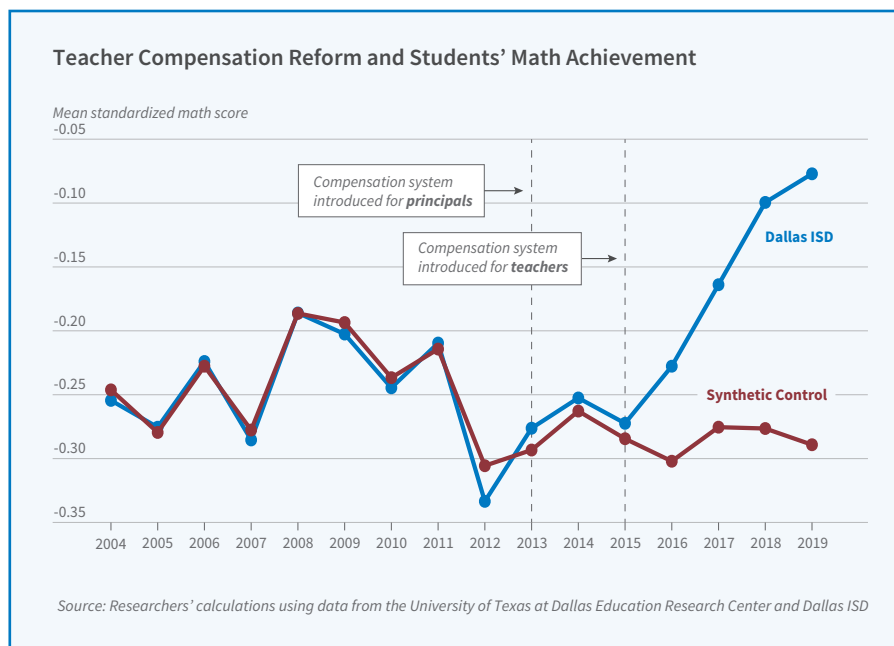
The program selected the worst performing schools based on test scores from 2014, two years before it was implemented. The researchers used as a control group the next lowest-performing schools in Dallas, a group exposed to similar conditions in the run-up to the program's implementation.

Overall, the program raised average achievement at the lowest-performing schools nearly to the districtwide average. Math scores saw greater gains, but reading increases were substantial as well. Students who attended targeted schools for two or more years continued to show large increases in achievement in middle school, suggesting last-

ing improvements in cognitive skills. The second wave of students showed similar results to those in the first, demonstrating that the program could be scaled up.

Ironically, the rewards system's success resulted in its undoing. When achievement scores at the targeted schools in the first wave approached the district average, stipends were largely removed. Consequently, the researchers write, "turnover jumped among the most effective teachers and test scores fell substantially."

—Steve Maas



tives, enhanced teacher support, and stronger school leadership.

In [Attracting and Retaining Highly Effective Educators in Hard-To-Staff Schools](#) (NBER Working Paper 31051), Morgan, Nguyen, Hanushek, Ost, and Rivkin report that Dallas obtained immediate and sustained improvements in student achievement in its lowest-ranked schools by using the new evaluation system to identify talented teachers and reward them with stipends if they worked at those schools.

# Mergers in Consumer Packaged Goods and Consumer Prices

Mergers can increase prices if the merging parties gain market power due to the deal. They can decrease prices if the union induces cost savings that the firms pass through to consumers. The regulatory agencies that review mergers must determine which scenario is more likely.

In *Merger Effects and Antitrust Enforcement: Evidence from US Retail* (NBER Working Paper 31123), Vivek Bhattacharya, Gastón Illanes, and David Stillerman study the results of 50 mergers. The deals examined were valued at \$280 million or more and involved producers of consumer packaged goods sold at grocery stores and mass merchandisers. These producers sell many products; these mergers impacted 126 product markets.

The researchers study the effect of mergers on the sale prices and quantities sold of individual products in the two years before and after completed mergers, both for the merging parties and for competitors. They rely on data from the NielsenIQ Retail Scanner Dataset.

After a merger, the average price of a product sold by the merging par-

ties decreased by 0.1 percent. Over the same period, prices for products sold by nonmerging firms rose by 2.1 percent. These averages mask substantial differences across mergers, consistent with the

Consumer packaged goods producer mergers led to disparate price changes. One quarter resulted in price drops of 2.3 percent or more, while another quarter led to increases of at least 5.3 percent.

notion that some deals lead to stronger exertion of market power while others lead to cost synergies. In 25 percent of the analyzed mergers, prices fell by at least 2.3 percent. Another 25 percent led to price increases of 5.3 percent or more.

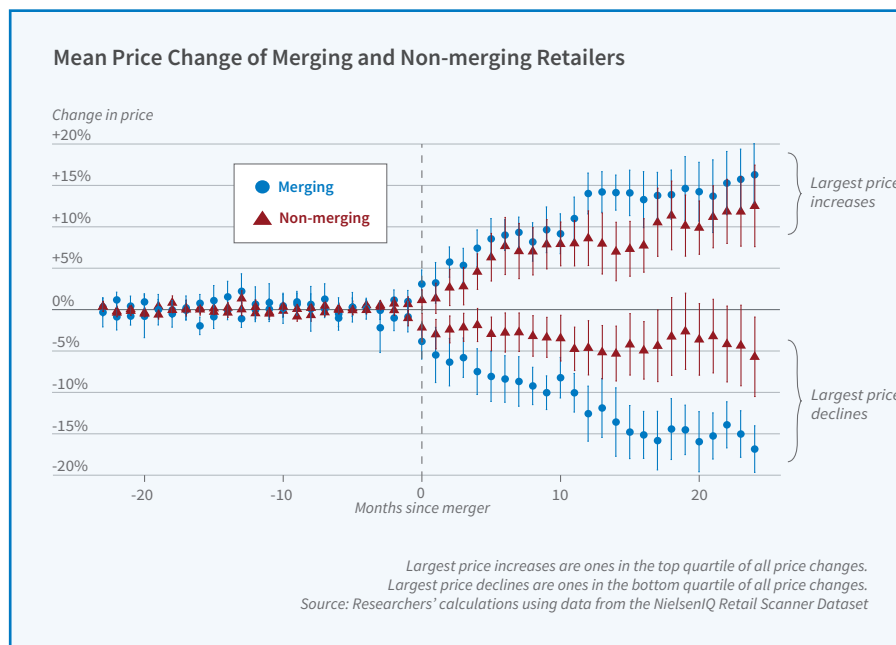
Aggregate quantities decreased by

2.3 percent on average after a merger, again with substantial variation: the first quartile saw a drop of 6.9 percent, compared with an increase of 3.1 percent for the third quartile. Significant quan-

tity declines following a merger correlate with narrower distribution networks and with reductions in product portfolios.

The researchers apply their findings to estimate the decision rule that agencies used in deciding whether to challenge proposed mergers. They find that during their study period, 2006 to 2017, US antitrust agencies challenged mergers with an expected sales-weighted average price increase of more than 8.6 percent. Reducing the challenge threshold to 5 percent would have blocked few procompetitive mergers while reducing the probability of approval for anti-competitive mergers. However, it would have tripled the number of proposed mergers that the agencies would have had to challenge.

—Linda Gorman



## NBER

The National Bureau of Economic Research is a private nonprofit research organization founded in 1920 and devoted to conducting and disseminating nonpartisan economic research. Its officers are:

James M. Poterba—President and Chief Executive Officer

John Lipsky—Chair  
Peter Blair Henry—Vice Chair  
Robert Mednick—Treasurer

The NBER Digest summarizes selected Working Papers recently produced as part of the NBER's program of research. Working Papers are intended to make preliminary research results available to encourage discussion and suggestions for revision. Neither the

Working Papers nor The Digest have been subject to peer review or review by the NBER Board of Directors.

The Digest is free and may be reproduced with appropriate attribution to the NBER. Please provide the NBER's Public Information Department ([ktasley@nber.org](mailto:ktasley@nber.org)) with copies of anything reproduced.

Requests for Digest subscriptions, changes of address, and cancellations may be sent to Digest, NBER, 1050 Massachusetts Avenue, Cambridge, MA 02138-5398 (please include the current mailing label), or emailed to [subs@nber.org](mailto:subs@nber.org). Print copies of the Digest are only mailed to subscribers in the US and Canada; those in other nations may request electronic subscriptions at [www.nber.org/drsunsubscribe/](http://www.nber.org/drsunsubscribe/).

Individual copies of NBER Working Papers are available online free of charge to affiliates of subscribing organizations, such as universities and colleges, and to employees of NBER corporate associates. All visitors to the NBER website receive three free downloads each year, after which there is a charge of \$5 per downloaded paper. To place an order, please email the NBER's Subscriptions Department at [subs@nber.org](mailto:subs@nber.org). A full subscription to the NBER Working Paper series entitles the subscriber to all new papers, recently more than 1,200 per year. The standard annual rate for a full digital subscription is \$2,995; the online academic rate is \$1,375. Partial subscriptions are also available; rates may be found at [nber.org/wpsubscribe.html](http://nber.org/wpsubscribe.html)