**6. New Horizons: Measurement with Nontraditional Data**

Increasingly, nontraditional micro data and miscellaneous indicators collected or assembled elsewhere have become a key resource in R&S. In principle, such data may be part of an administrative program such as the income tax system, a survey conducted elsewhere for a different or overlapping purpose, remote sensor information, and a wide spectrum of other types of information generated within a nonstatistical framework, such as scanner data. Use of administrative data has a long history in R&S, most notably as an input in the construction of the Industrial Production Index and in the support for the SCF, as noted earlier. Federal Statistical Research Data Centers (FSRDC), a partnership between the Census Bureau, other government agencies and research organizations, has made remote access to a broader range of restricted-use government data. Under the leadership of Norman Morin in R&S, an FSRDC was established at the Board. Key movements elsewhere make it highly likely that this facility will grow in importance. Following the report of the 2016 U.S. Commission on Evidence-Based Policymaking created under the Evidence-Based Policymaking Commission Act (Public Law 114-140), in 2018 Congress passed the Foundations for Evidence-Based Policymaking Act (Public Law 115-435, often referred to as the “Evidence Act”). Among the most relevant aspects of the act for the work of R&S is the presumption that all data should be available, subject to statutory limitations and appropriate provisions for maintaining confidentiality, and the application for data access is standardized across all agencies. Linking or otherwise blending such data will be a central issue. Fortunately, there is a long experience with such work in R&S.

R&S has always applied its deep set of measurement skills in exploiting novel and informative data when analyzing the economic landscape. But over the past decade, measurement with nontraditional data has become more intentional and systematic in R&S, as it has in other parts of government and elsewhere. Indeed, such work may currently be the most rapidly developing area of statistics. In October of 2014, Christopher Kurz, Norman Morin, and John Stevens, along with Micheline Casey (head of the OCDO at the time) traveled out to the Bay Area in October of 2014 to visit tech companies and academic institutions.[[1]](#footnote-1) Around this time, R&S also had multiple tech companies visit the Board to continue exploring data-sharing and other collaboration to better inform economic measurement. The progress on this expanded measurement agenda (EMA) led David Wilcox to create a group by that name within the Industrial Output section in 2016, with the first manager of this group becoming Christopher Kurz, with the mission of systematically identifying novel data sources that can be used to develop timely and accurate measures of economic activity that are relevant to the FOMC’s dual mandate.

These new data sources serve two main purposes. First, they help the R&S staff to forecast - macroeconomic indicators with improved accuracy, at a higher frequency, or with greater disaggregation by, for example, geography or product type. Second, the new measures derived from these data themselves assist in the identification of the current cyclical position of the economy.For example, these and other series are used as inputs into a statistical filtering exercise to estimate latent factors associated with the economic cycle in real time; this “nowcasting” effort became part of the mission of the Current Macroeconomic Conditions Section, the creation of which in 2015 was another initiative under the leadership of David Wilcox. While the expanded measurement agenda formally resides within the Industrial Output section, the work with novel nontraditional data requires extensive collaboration throughout R&S and the Federal Reserve Board.

In light of the broader technical evolution in using such data and the discovery of new data sources. continuing outreach is an integral component of the EMA group’s responsibilities. The group stays actively connected to private, academic, and government institutions through participation in conferences, meetups and hosting visitors, as well as through close collaboration with the Federal Statistical Agencies and other policymaking institutions, such as foreign central banks. Inside the Board, the group has partnered with the Division of Monetary Affairs and the Division of International Finance.

In addition to active participation in organizing conferences elsewhere, such as events in 2018 and 2019 at the Bank of England, in 2019 the Federal Reserve Board hosted a two-day conference featuring new research on nontraditional data, machine learning, and natural language processing in macroeconomics. The event highlighted the importance of nontraditional data sources and new empirical methods for macroeconomic analysis, with a particular focus on policymaking.

The conference attracted about 80 in-person participants, including attendees from20 agencies and foreign central banks. Moreover, the success of the event has led to an international partnership and the expansion of the conference. Since 2019 the Board has coordinated on three more events, with the Sveriges Riksbank, Bank of Canada, and the Banca d’Italia. The events are also now jointly coordinated with the Central Bank Research Association (CEBRA), and the Economics with Nontraditional Data and Analytical Tools (ECONDAT) program. Notably, the event returns to the Board in 2023, with the 5th conference on “Nontraditional Data, Machine Learning, and Natural Language Processing in Macroeconomics” being held in November of 2023.

One of the earliest systematic efforts in R&S to leverage nontraditional data involved an attempt in 2016 to gain access to microdata from the payroll processing firm Automatic Data Processing (ADP). At that time ADP in collaboration with Moody’s was already using these data in a modeling exercise to predict payroll gains in the BLS CES data. However, many economists at the Board thought that a better utilization of ADP microdata would be to generate an independent measure of payroll employment gains. Following successful, if intense, work to obtain access to ADP weekly microdata in near real time , R&S staff were able to generate a new measure of payroll employment, which featured in a 2017 Tealbook.[[2]](#footnote-2) This effort and related work combining this estimate with CES estimate of payroll employment were reported in Cajner et al. (2018) and Cajner et al. (2019).

Another important early example of R&S working with nontraditional data was in the area of measuring consumption, using near real-time data from First Data (acquired by Fiserv in 2019). These data were successfully used for measuring consumption responses to different events, ranging from hurricanes (Aladangady et al., 2016), sales-tax holidays (Aladangady et al., 2017), and federal tax refunds with an earned income tax credit (Aladangady et al., 2018). Aladangady et al. (2019) gives a broad overview of the construction of a real-time, high-frequency, geographicly detailed measure of consumer spending from First Data transaction data.

In the area of measuring business formation, economists from R&S joined forces with Census and academic economists to construct a new dataset for measuring business formation in near real time (Bayard et al. (2018a, 2018b)). The new dataset was based on applications for Employer Identification Numbers (EINs) submitted in the United States, known as IRS Form SS-4 filings.

The experience with nontraditional data for economic measurement proved especially helpful in dealing with the economic uncertainty accompanying the COVID-19 crisis. At that time, economic activity was deteriorating at an unprecedented pace and many critical monetary and fiscal policy decisions had to be made before the release of the usual official statistics. Cajner et al. (2022) describes the work done in R&S to rapidly measure economic developments during that period, which relied importantly on high frequency indicators. (see table in the accompanying spreadsheet) The box “Monitoring Economic Activity with Nontraditional High-Frequency Indicators” in February 2021 Monetary Policy Report provides some additional examples.[[3]](#footnote-3)

Three main benefits that nontraditional data sources provided during the pandemic were timely measurement, granularity, and crisis-specific estimates. First, in terms of timely measurement during the pandemic, R&S took advantage of the progress made on the expanded measurement agenda in the years before the pandemic started. For example, the enormous employment losses during the pandemic became visible in official statistics only in the beginning of May 2020, when the BLS released its April 2020 employment situation report. However, R&S was able to monitor employment declines during the pandemic essentially in real-time with the weekly ADP-FRB data (see Cajner et al., 2020a and Cajner et al., 2020b)[[4]](#footnote-4). Similarly, First Data allowed R&S to measure in near real-time consumption declines, while Business Formation Statistics published by the Census provided crucial information about business starts, which surged surprisingly in the aftermath of the pandemic (see Decker and Haltiwanger, 2023). Relatedly, a team of authors from R&S worked on measuring in near real time business exit during the pandemic (see Crane et al., 2020).

Second, while official government statistics provide reliable information about aggregate statistics, they often lack granular detail, such as detailed measurement across geographic areas (states or counties) or by different breakdowns of other individual statistics (such as income). During the pandemic, the new methods developed within R&S allowed a granular assessment of events. For example, given that the first wave of the pandemic produced relatively more severe COVID-19 outbreaks in northeastern parts of the country, R&S was able to use an array of granular information in nontraditional data—including geographically detailed data on employment, spending, public transportation, and health indicators—to better understand links between health shocks and the responses of economic variables. This knowledge proved very important in predicting the evolution of the U.S. economy as the pandemic became widespread throughout the country. Similarly, essentially real-time analysis of the Paycheck Protection Program performed by a group of R&S and academic economists with the ADP microdata provided crucial understanding of how this program affected the evolution of employment (Autor et al., 2020).

Third, there were important measurement needs during the pandemic for which official government statistics provided little or no information. In particular, policymakers needed to pay attention to the influence of health-related indicators—such as COVID-19 cases, hospitalizations, deaths, and vaccinations—in understanding disruptions to the economy. With its accumulated experience and access to nontraditional data, R&S was quickly able to develop relevant indicators and provide regular weekly updates thereafter. This experience increases confidence that R&S can address other disruptions in the future that may differ from typical economic recessions.

Use of nontraditional data by R&S has proven worthwhile, but there are frequently costs and risks in doing so. Very many sources with seemingly useful data are commercial organizations that may face incentives that differ importantly from those of the Fed. Typically, data must be purchased, and the costs of nontraditional data can be substantial. Moreover, the costs may rise unexpectedly over time, to the point that the data become prohibitively expensive. In addition, because such data are most typically collected for a business purpose, variables may not be documented to the extent typical of data collected within a scientific framework, and the definitions may change over time as business needs change—or a series may stop altogether. R&S staff usually must invest considerable effort to make the data useable for its purposes, effort that is lost when the data can no longer be obtained. Frequently, such data are not available for a fully representative population and statistical methods must be developed to make them usable. Broader measurement expertise within R&S has been helpful in addressing this and other important methodological issues with nontraditional data. A final risk that has been raised by many users of such data is the risk that a data provider may work with their own data for the purpose of learning about the decision-making processes of a user.

Despite all the challenges of using nontraditional data, important opportunities remain for future work. Across the domain of official statistics, nontraditional data already have a role, and that role seems likely to increase. The ongoing discussions and collaborations within the government support sharing of methods and ideas that will benefit the Federal Reserve Board and the broader production of official statistics. In its most recent initiatives, the EMA group in R&S has been working to advance measurement efforts by using text analytics, natural language processing, machine learning, and large language models. While still in their initial stages, these efforts are already producing some promising results.

**References**

Aladangady, Aditya, Shifrah Aron-Dine, Wendy Dunn, Laura Feiveson, Paul Lengermann, and Claudia Sahm (2016). “The Effect of Hurricane Matthew on Consumer Spending,” FEDS Notes 2016-12-02.

Aladangady, Aditya, Shifrah Aron-Dine, Wendy Dunn, Laura Feiveson, Paul Lengermann, and Claudia Sahm (2017). “The Effect of Sales-Tax Holidays on Consumer Spending,” FEDS Notes 2017-03-24.

Aladangady, Aditya, Shifrah Aron-Dine, David Cashin, Wendy Dunn, Laura Feiveson, Paul Lengermann, Katherine Richard, and Claudia Sahm (2018). “High-frequency Spending Responses to the Earned Income Tax Credit,” FEDS Notes 2018-06-21.

Aladangady, Aditya, Shifrah Aron-Dine, Wendy Dunn, Laura Feiveson, Paul Lengermann, and Claudia Sahm (2019). “From Transactions Data to Economic Statistics: Constructing Real-time, High-frequency, Geographic Measures of Consumer Spending,” FEDS WP 2019-057.

Autor, David, David Cho, Leland D. Crane, Mita Goldar, Byron Lutz, Joshua Montes, William B. Peterman, David Ratner, Daniel Villar, and Ahu Yildirmaz (2020). “An Evaluation of the Paycheck Protection Program using Administrative Payroll Microdata,” MIT WP.

Bayard, Kimberly, Emin Dinlersoz, Timothy Dunne, John Haltiwanger, Javier Miranda, and John Stevens (2018a). “Measuring Early-Stage Business Formation.” FEDS Notes 2018-03-07.

Bayard, Kimberly, Emin Dinlersoz, Timothy Dunne, John Haltiwanger, Javier Miranda, and John Stevens (2018b). “Early-Stage Business Formation: An Analysis of Applications for Employer Identification Numbers.” FEDS WP 2018-015.

Cajner, Tomaz, Leland Crane, Ryan Decker, Adrian Hamins-Puertolas, Christopher Kurz, and Tyler Radler (2018). “Using Payroll Processor Microdata to Measure Aggregate Labor Market Activity,” FEDS WP 2018-005.

Cajner, Tomaz, Leland Crane, Ryan Decker, Adrian Hamins-Puertolas, and Christopher Kurz (2019). “Improving the Accuracy of Economic Measurement with Multiple Data Sources: The Case of Payroll Employment Data,” FEDS WP 2019-065.

Cajner, Tomaz, Leland D. Crane, Ryan A. Decker, John Grigsby, Adrian Hamins-Puertolas, Erik Hurst, Christopher Kurz, and Ahu Yildirmaz (2020a). “The U.S. Labor Market during the Beginning of the Pandemic Recession,” Brookings Papers on Economic Activity.

Cajner, Tomaz, Leland D. Crane, Ryan A. Decker, Adrian Hamins-Puertolas, and Christopher Kurz (2020b). “Tracking Labor Market Developments during the COVID-19 Pandemic: A Preliminary Assessment,” FEDS WP 2020-030.

Cajner, Tomaz, Laura J. Feiveson, Christopher J. Kurz, and Stacey Tevlin (2022). “Lessons Learned from the Use of Nontraditional Data during COVID-19,” in Recession Remedies.

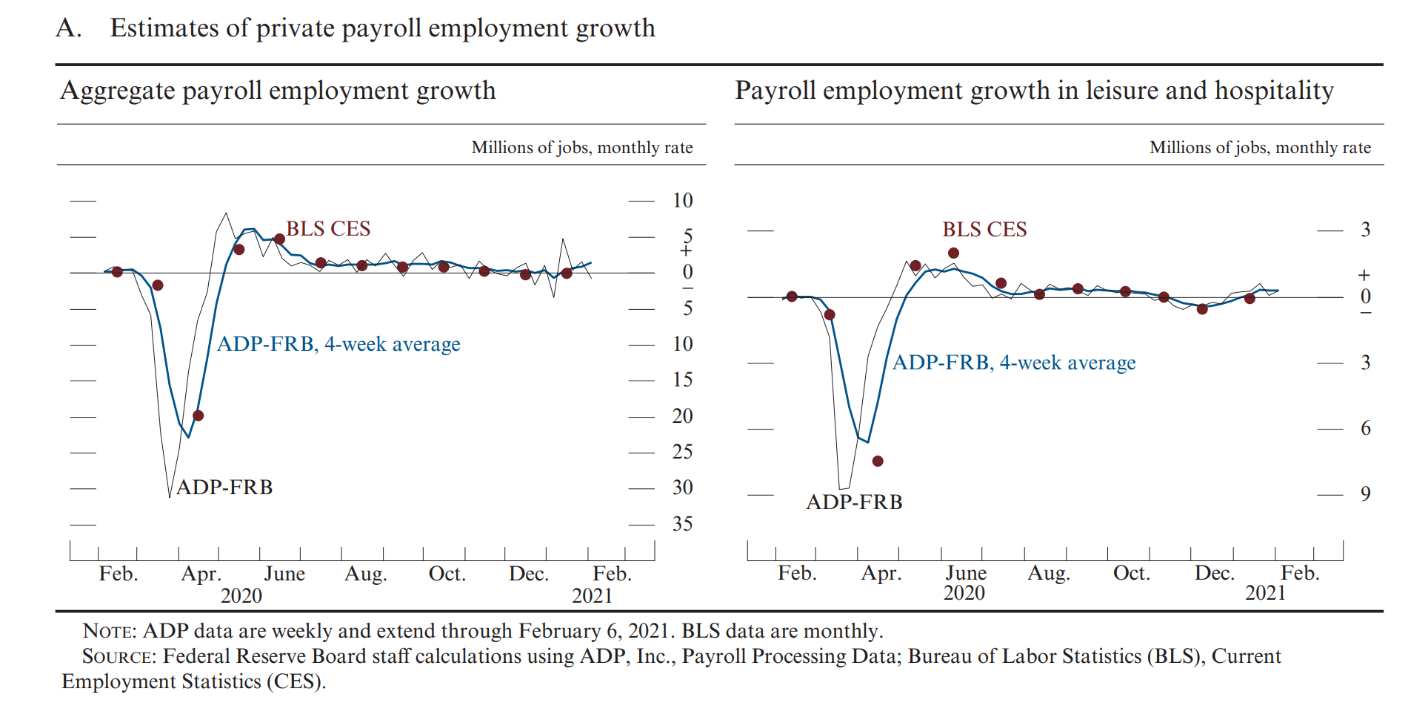
Crane, Leland D., Ryan A. Decker, Aaron Flaaen, Adrian Hamins-Puertolas, and Christopher Kurz (2020). “Business Exit during the COVID-19 Pandemic: Non-Traditional Measures in Historical Context,” FEDS WP 2020-089.

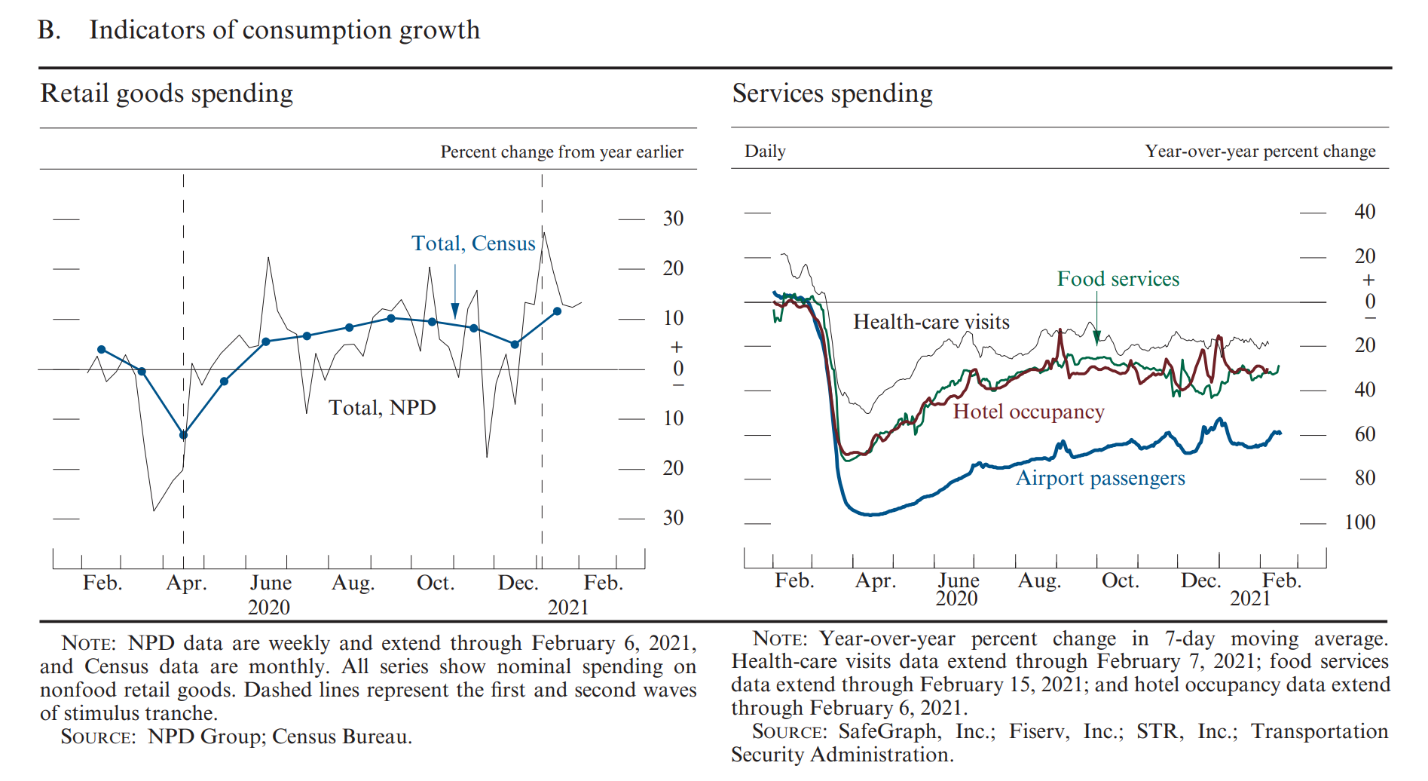
Decker, Ryan A., and John Haltiwanger (2023). “Surging Business Formation in the Pandemic: Causes and Consequences?,” Brookings Papers on Economic Activity.

**Possible Figures to accompany the text**

* **Charts from MPR, Feb 2021, to illustrate what we focused on during the pandemic**

(https://www.federalreserve.gov/monetarypolicy/files/20210219\_mprfullreport.pdf)





* **TB box from Dec 2017 (now publicly available) as an example of work done pre-pandemic**

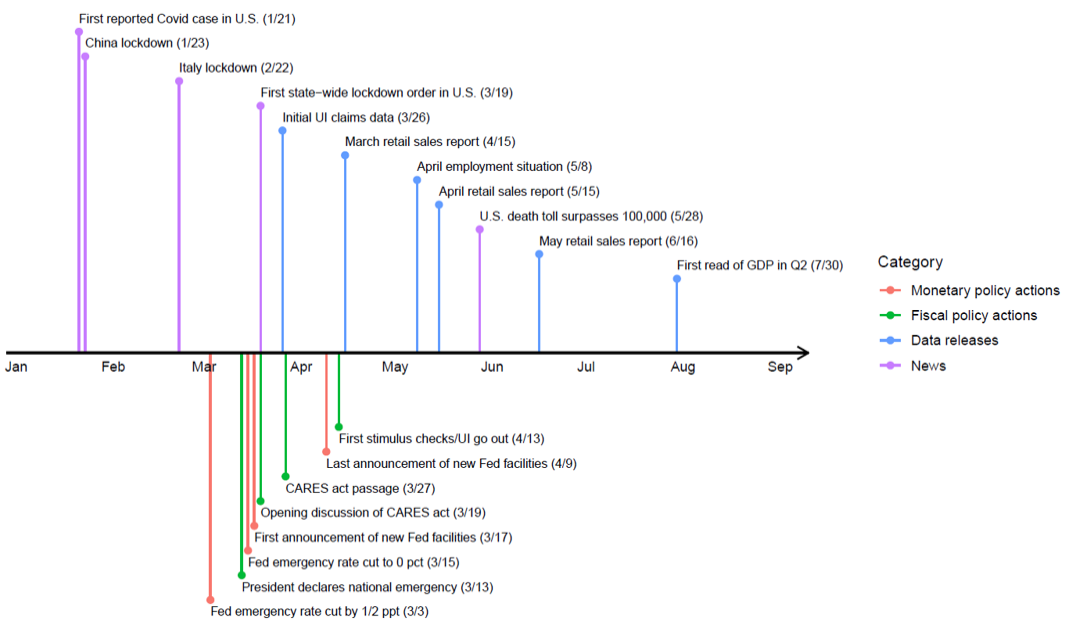
(https://www.federalreserve.gov/monetarypolicy/files/FOMC20171213tealbooka20171201.pdf)



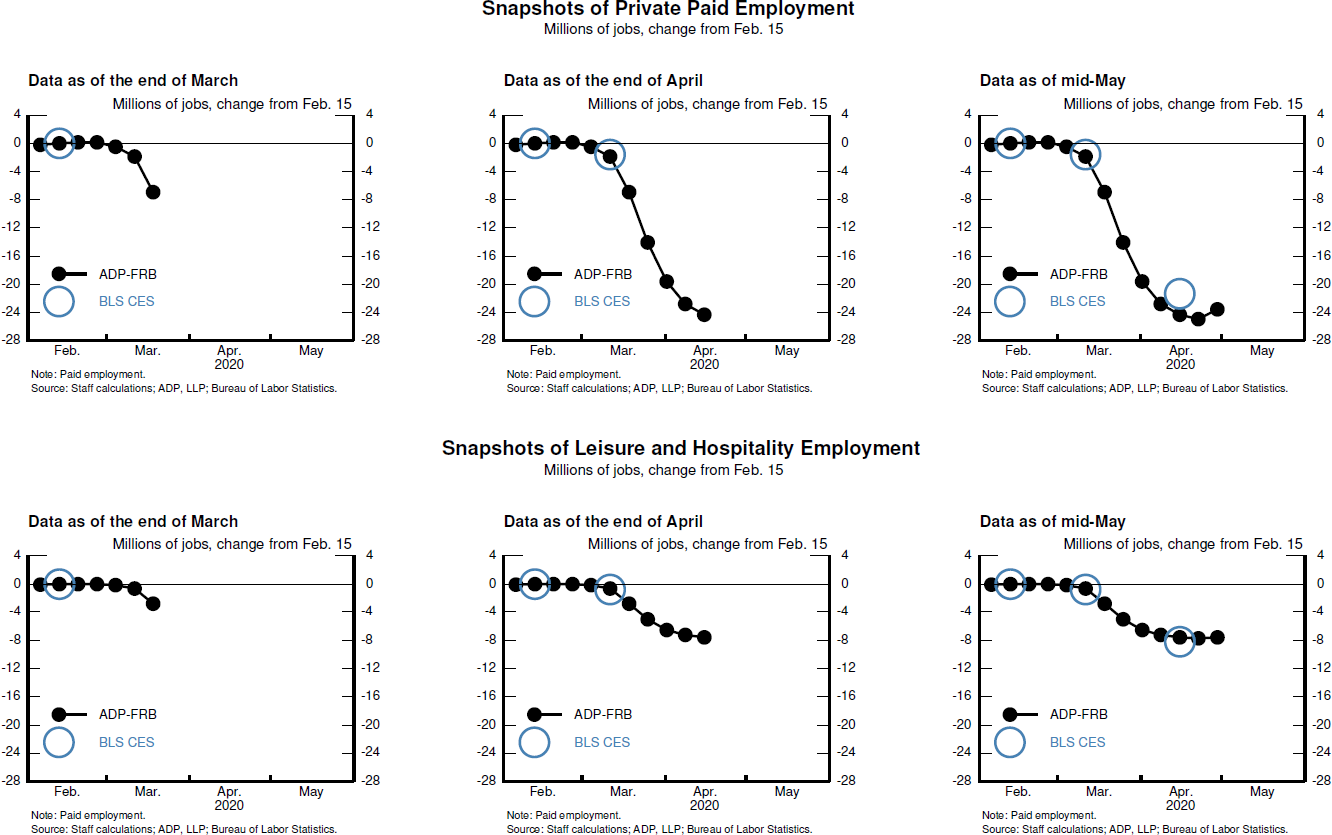
* **Figures from the paper by Cajner, Feiveson, Kurz, Tevlin to further illustrate what we did during the pandemic**

https://www.brookings.edu/wp-content/uploads/2022/04/RR-Chapter-9-Use-of-Nontraditional-Data.pdf

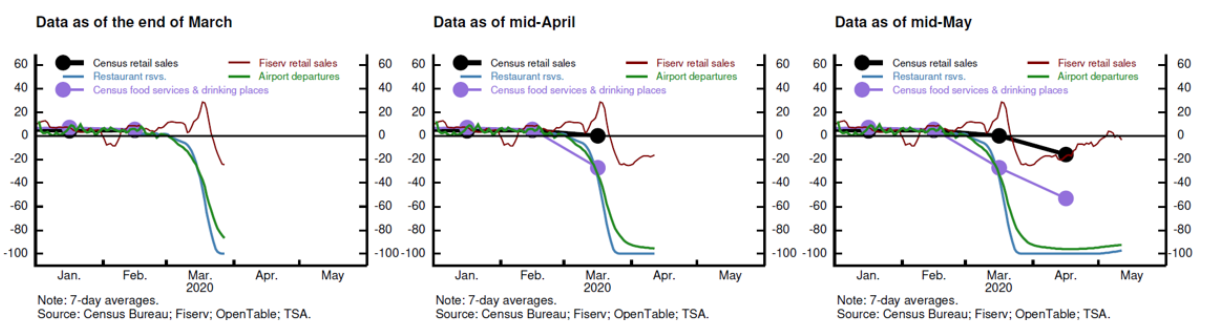
**Figure 1: Early Policy Responses to COVID-19**



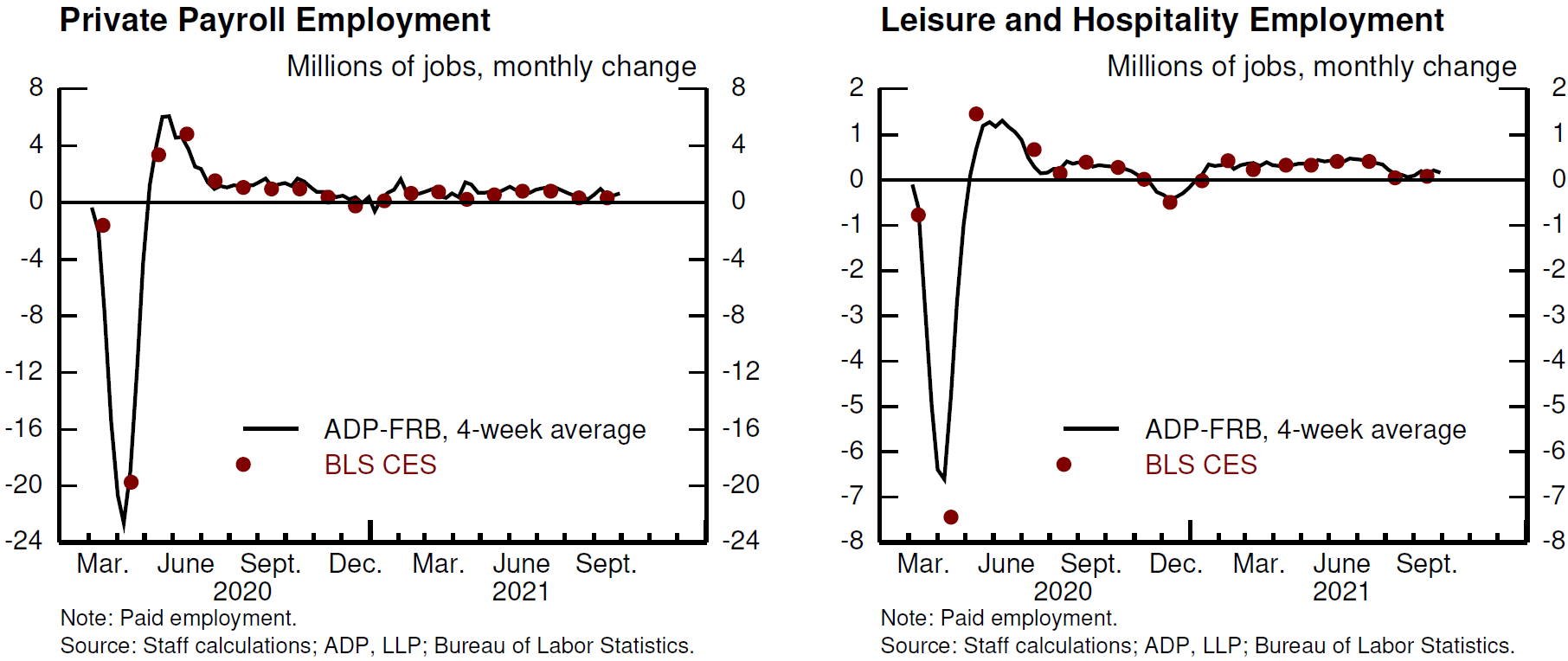
**Figure 2: Snapshots of Employment Data**



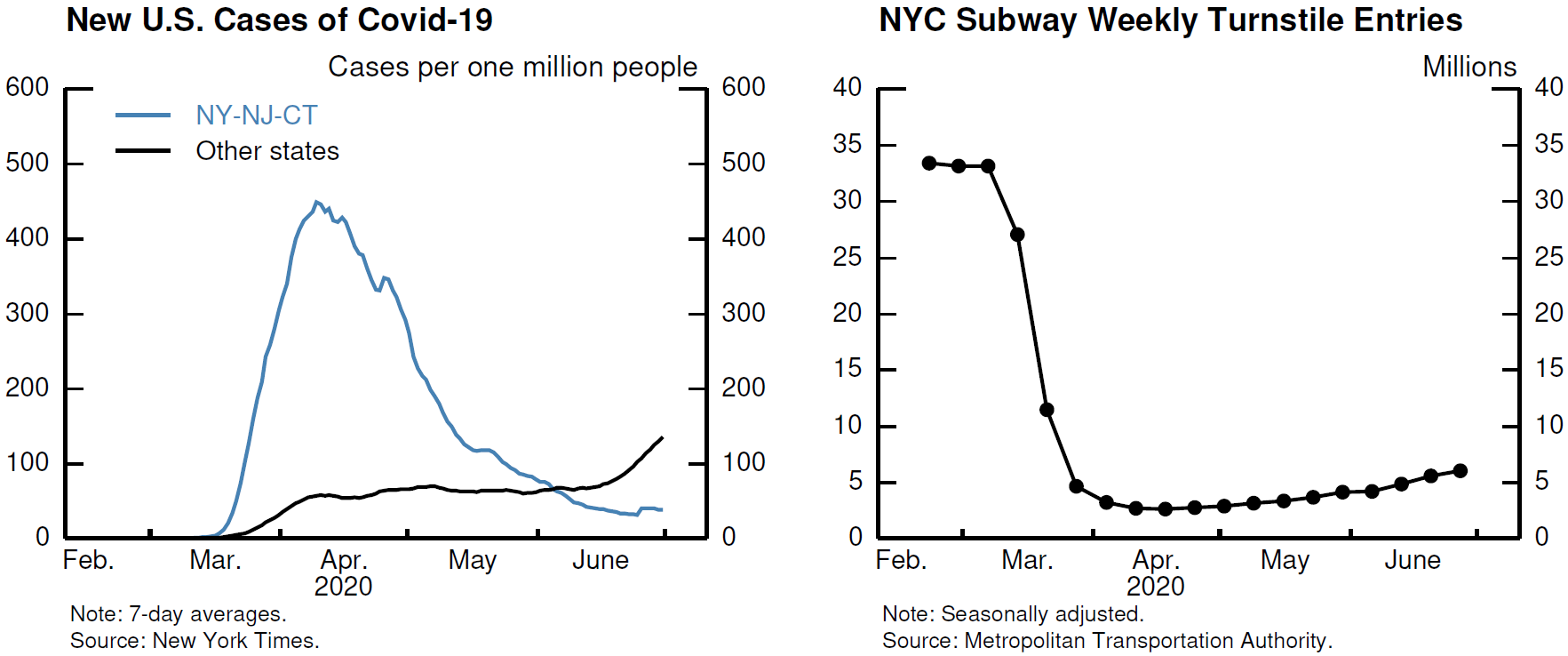
**Figure 3: Snapshots of Consumer Spending (percent change from same period in 2019)**



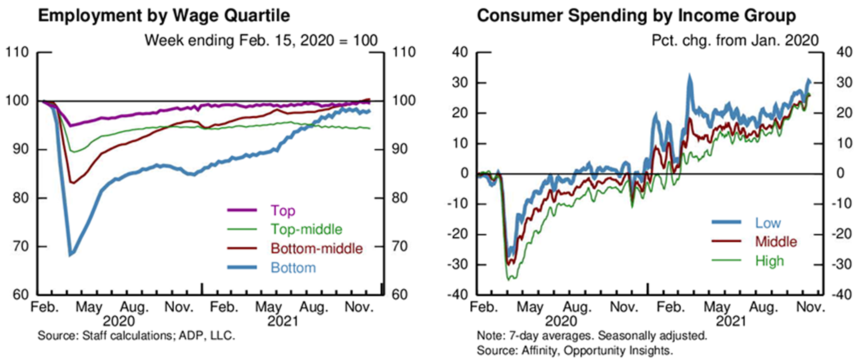
**Figure 4: ADP and BLS Measures of Employment**



**Figure 5: COVID-19 and Employment**

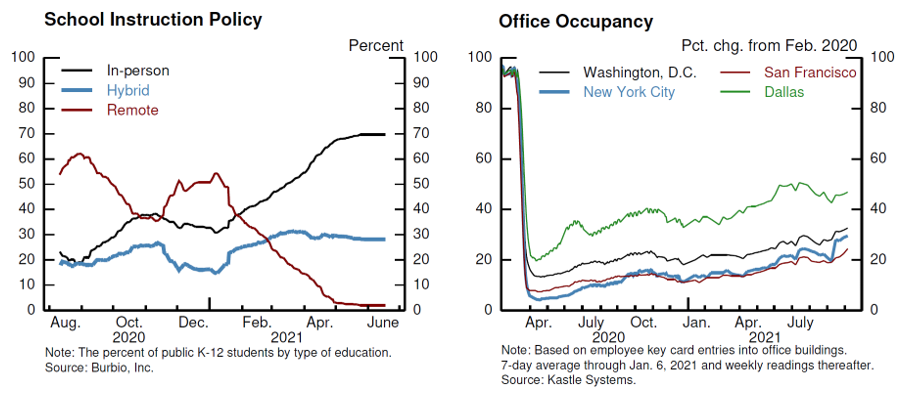


**Figure 6: Employment and Spending by Income**

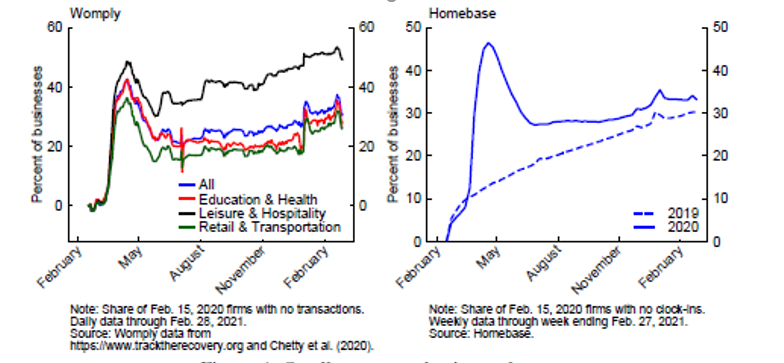
****

**Figure 6: Spending**

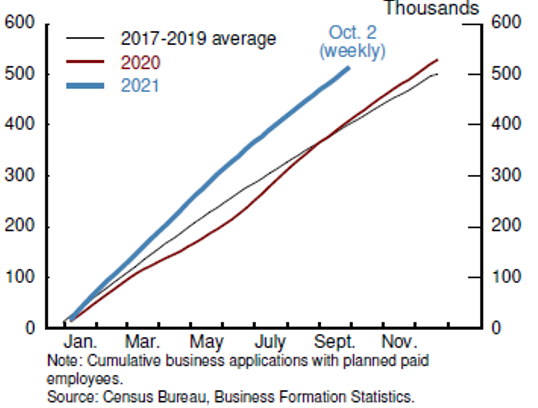
**Figure 8: School Instruction Policy and Office Occupancy**



**Figure 9:  Small Business Closures**



**Figure 10.  New Business Applications**



1. The visits were to Google, Intuit, Auction.com, Facebook, Amplab at UC Berkeley, Thumbtack, Visa, Salesforce, and Square. [↑](#footnote-ref-1)
2. Tealbook from December 2017 is now publicly available here, so we can safely talk about it:

   https://www.federalreserve.gov/monetarypolicy/files/FOMC20171213tealbooka20171201.pdf [↑](#footnote-ref-2)
3. See https://www.federalreserve.gov/monetarypolicy/files/20210219\_mprfullreport.pdf. [↑](#footnote-ref-3)
4. It is interesting to note here that production of weekly ADP-FRB employment data relied on a seasonal adjustment procedure developed decades earlier by R&S economist Bill Cleveland. In his time, R&S had been one of the leaders in the development of seasonal adjustment methods. [↑](#footnote-ref-4)