Business Economics
https://doi.org/10.1057/s11369-021-00250-3

NABE
National Association for Business Economics

ORIGINAL ARTICLE

Blending data to understand the economic impact of COVID-19

Scott A. Brave[1] · Rebecca Hutchinson[2] · Christopher J. Kurz[3] · Michael Stepner[4] · Ron Jarmin[2]

This is a U.S. government work and not under copyright protection in the U.S.; foreign copyright protection may apply 2022

Abstract
The data demands during the pandemic heightened the need to blend information from numerous sources to get a more timely and granular picture of economic developments. Ongoing efforts include the Chicago Fed's weekly retail sales estimate, the Census Bureau's work on higher-frequency state-level retail sales data, the Federal Reserve Board's computations of business closures and weekly payrolls, and the academic Opportunity Insights team's estimates of spending, business revenues and employment by income and ZIP code.

Keywords Blended data · Pandemic · Private data · Public data · Weekly data

1 Ron Jarmin

This is an immensely important topic. Blending data to understand the world is a topic in and of itself, and I think especially important during something like the COVID-19 pandemic. I speak from experience about what happens to our data collection infrastructure during those times, with the Census and other things we do at the Census Bureau.

A lot of people have applied some very creative and clever ways to blend different data sources together to build something more than what could have happened if you only had access to one data source. We're going to get started with Scott Brave, from the Federal Reserve Bank of Chicago.

2 Scott Brave

Today I'd like to talk about a project that we've been working on at the Chicago Fed for about the last two years now. It really started when the pandemic started–blending economic statistics and Big Data into something that we call the Chicago Fed Advance Retail Trade Summary.

For us the COVID-19 pandemic really heightened the need for timely measures of consumer spending. If we think back to February and March 2020, just how quickly things changed during that period and how the consumer was front and center at that time. It really highlighted some weaknesses for us on data collection, and really brought about this project.

At the Chicago Fed we developed a new weekly index for this purpose that combines high-frequency Big Data with the U.S. Census Bureau's Monthly Retail Trade Survey. What I'm going to talk to you today is how this index works, and then some of the applications that we've come up with.

What do we do? We take high-frequency data from five private companies and one federal agency to construct a weekly measure of retail and food services sales, excluding

Based on the session *Blending Data to Help Understand the Economic Effect of COVID19*, held at the NABE TEC Conference, November 8, 2021. All views expressed are those of the authors, and do not necessarily reflect those of the U.S. Census Bureau, the Federal Reserve Bank of Chicago, or the Board of Governors of the Federal Reserve.

✉ Christopher J. Kurz
christopher.j.kurz@frb.gov

Scott A. Brave
scott.a.brave@chi.frb.org

Rebecca Hutchinson
Rebecca.J.Hutchinson@census.gov

Michael Stepner
michael.stepner@utoronto.ca

Ron Jarmin
ron.s.jarmin@census.gov

[1] Federal Reserve Bank of Chicago, Chicago, USA

[2] US Census Bureau, Washington, DC, USA

[3] Board of Governors of the Federal Reserve System, Washington, DC, USA

[4] University of Toronto, Toronto, ON, Canada

Published online: 23 February 2022

automotive spending. It is, most importantly, benchmarked to the Census Bureau's Monthly Retail Trade Survey.

We use a statistical framework and mixed-frequency dynamic-factor model to constrain this latent measure to match the latest Census Bureau data, the latest Monthly Retail Trade Survey. This allows us to create something that is both timely and available more frequently. This is something we publish publicly twice a month, so twice as often as the Monthly Retail Trade Survey. The index covers the period from January 2018, so it's mostly using very recent data, with four weekly values per month.

We're using a bit of a non-standard calendar to allow us to replicate the Census structure. We adjust for that with the way we seasonally adjust our data. The most important thing that we found that using this high-frequency data really allows us to accurately predict the Advance Monthly Retail Trade Survey, or MARTS.

We've been back-testing this model and comparing it against consensus nowcasts since February 2020. We found that out-of-sample this model is producing forecasts that are roughly about 50% more accurate than the consensus forecast. I don't have a whole lot of time to go over the details today, but all those details are available in a Chicago Fed working paper (Brave *et al.* 2021a).

We call this measure CARTS, a little bit of a play on words comparing it to MARTS, the Census Bureau measure. It's really short for "Chicago Fed Advance Retail Trade Summary." CARTS contains, like I mentioned, a weekly index of retail trade summarizing data on credit and debit card transactions, retail foot traffic, gasoline consumption, and consumer sentiment.

This is all available on our public website: ChicagoFed.org/CARTS. But you can also find it on the St. Louis Fed's FRED Database, and in Haver Analytics Survey Database as well. And there's also a nice little non-technical introduction to CARTS and everything that it contains. It's most recently published in the *Chicago Fed Insights Blogpost* (Brave *et al*. 2021b).

Let me just give you a basic sense of how this works, rather than going into the nuts and bolts of a dynamic-factor model. CARTS is a single common factor that's matching the Monthly Retail Trade Survey on a monthly basis. We're starting with the monthly Census Bureau data. We're adding to that weekly data on gasoline consumption from the EIA, daily data that's aggregated up to the weekly level from three sources of credit and debit card transactions, two that are consumer-facing, Consumer Edge and Facteus, and one that is business-facing, Womply.

We're adding to that retail foot traffic data that's daily and aggregated up to the weekly frequency from SafeGraph; and, finally, daily surveys of consumer senstiment that are run by the polling for Morning Consult, one of the last bit of information to go into this model. All of that is combined together in our dynamic-factor model into what we call CARTS.

If you go to our website and take a look at what we're producing, you're going to see a picture that looks like this (Fig. 1). This is retail and food services sales excluding automotive spending, expressed as billions of dollars on a seasonally adjusted basis.

What you'll notice from this figure is the importance of the benchmarking. The line with the Xs: that's actually the Census Bureau data, that's the Monthly Retail Trade Survey. You can see the weekly index line is basically intersecting the Census Bureau data. That's by construction.

We benchmark to the Census Bureau data so that the monthly frequency of our measure averages out to it. The interesting thing that comes about, though, in this process is you can see a fair amount of weekly variation that you wouldn't necessarily see in the monthly data. I think that's most visible in the pandemic period, where we actually pick up a few weeks of stockpiling effect early on, prior to the

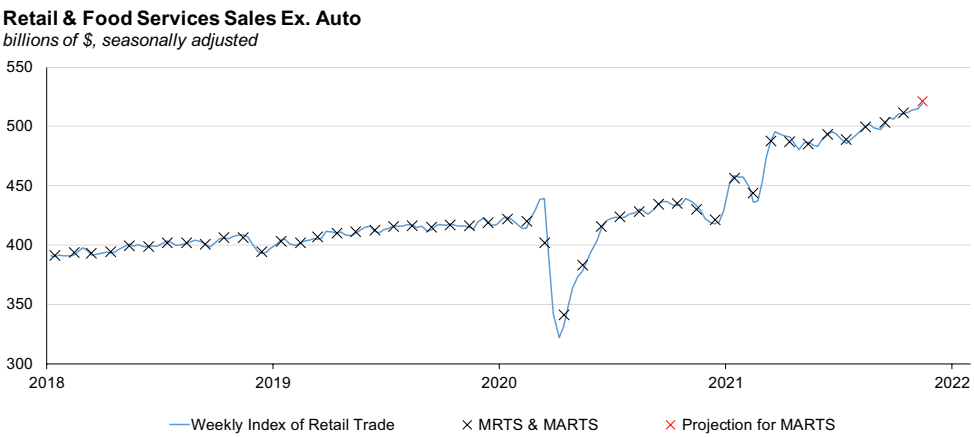


Fig. 1 Retail and Food Services Sales Ex. Auto. Billions of $, seasonally adjusted

shutdowns; and we also see an early turning point when the recovery first started.

The other aspect of what we're producing, if you look at the far righthand side of the graph, is that little X. Because we are using higher-frequency data, we're actually able to project what we think the upcoming MARTS value, or Advance Monthly Retail Trade Survey, is going to be.

We can also take a deeper dive into the data and look at our weekly units of retail trade, to see what's explaining the week-to-week variation in the index. It's about a 50/50 split between the contributions of the lagged monthly Census data and the high-frequency Big Data.

I mentioned that one of the things that we've looked at is the ability of this index to actually nowcast, or forecast the Advance Monthly Retail Trade Survey. We've been producing this index publicly now since June, so we have four releases of MARTS to compare our projections against other projections that are out there (Fig. 2).

Our forecasts have done pretty well over the last four months. In fact, in a mean absolute error sense, they've been about 55% more accurate than the consensus forecast. So, overall, we think this is proving to be a very useful measure. You can go to the website and actually see in tabular form our most recent projections (https://www.chicagofed.org/publications/carts/index). We're hoping that it's as useful for everyone as it certainly has been for us.

3 Ron Jarmin

I haven't thought of it, but there's some pun in there about "cart before the horse" or something like that Now we're going to go with Rebecca Hutchinson from the Census Bureau.

4 Rebecca Hutchinson

Today I will be presenting on the monthly state retail sales, a Census Bureau experimental data product that was developed and released last year in response to the need for more timely sub-national data during the pandemic, especially in the retail sector, which was so impacted by the pandemic.

These state-level retail data are some of the most-requested data from our data users, and we were all really excited to finally release the first version of these data last September. Previously, state-level retail sales information was only available once every five years as part of the Economic Census. The high cost of survey collection and respondent burden has made survey-based state-level retail sales difficult and costly to obtain. These data are modeled, so they required no new data collections.

We are publishing year-over-year percentage changes for each state and the District of Columbia for total retail sales, excluding non-store retailers, as well as for some of the retail subsectors. We are not publishing those non-store retailers which are primarily e-commerce-type businesses. The best method and data sources needed to attribute e-commerce sales to state geographies is a bigger challenge that we're still working through.

Let's take a quick look at some of the data. On this slide (Fig. 3) you can see the year-over-year percentage changes for total retail excluding nonstore retailers for each state in April and July of 2020 and 2021. States darkly shaded have positive year-over-year percentage changes, and those in lighter shades have negative changes.

In April 2020 when much of the country was impacted with those pandemic-related closures, you can see that much of the country had negative year-over-year percentage changes lightly shaded. And in July 2020, in the upper

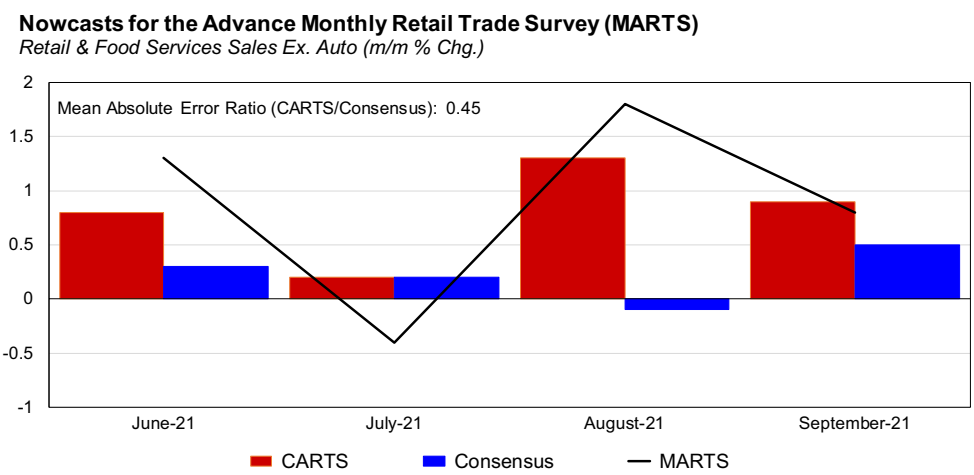


Fig. 2 Nowcasts for the advance monthly retail trade survey (MARTS). Retail and food services sales ex. auto (m/m % Chg.)

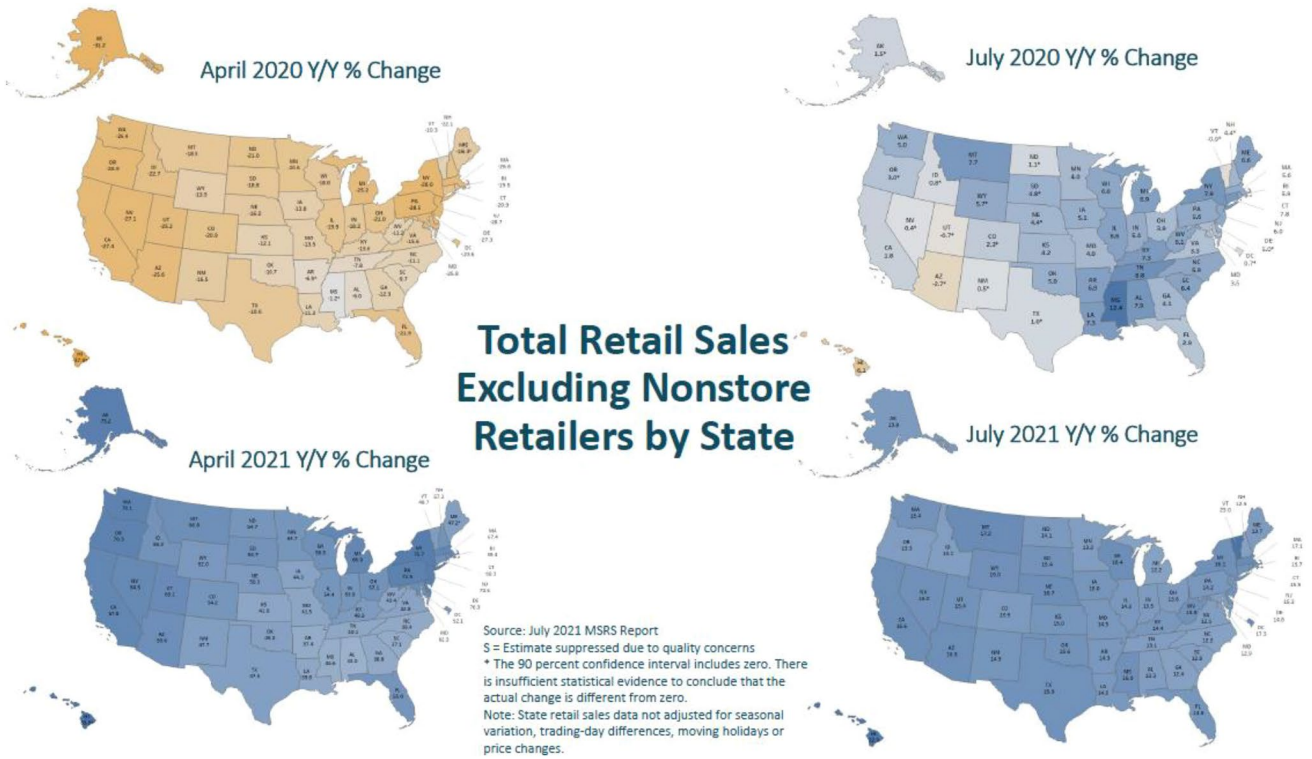


Fig. 3 Total retail sales excluding nonstore retailers by state

right, there was some improvement, with many states having positive and significant year-over-year changes.

In that bottom row you can see that both April and July 2021 every state has positive year-over-year changes when compared to 2020. It is interesting to note that many states with positive year-over-year changes in July 2020 also have year-over-year growth in July 2021.

During those early months of the pandemic, most of retail saw negative year-over-year changes. However, food and beverage stores had a more positive April 2020 (Fig. 4), because we were all home cooking and baking bread. Almost every state has positive year-over-year changes in sales in April 2020.

Flash forward to 2021, in that bottom row, and you can see that, not surprisingly, sales in April 2021 were down year-over-year for most states with restaurants reopening. But it is interesting to look at a state like New York that was up 18% in April 2020 and in April 2021 built on that growth and was up 4.7% compared to 2020. The same thing happens in Texas in July, where it was up 13.8% in July 2020 and it's up 5.9% in July 2021.

I just want to touch briefly on how we created these data. Early in the pandemic a small group of us were a tasked with assessing quickly the data we had that could be used for these estimates, and developing a methodology. Data that we had access to that would be useful to this work included administrative data, in the form of payroll; national- and company-level retail sales data, from the Monthly Retail Trade Survey; and point-of-sale retail sales data, from a third-party provider, in this case it's MPD, for a small number of retailers. We used national tabulations of this third-party data for non-response in our Monthly Retail Trade Survey, but it is excellent for state-level estimates. However, this data requires retailer consent because it is available at that store-level, down to the ZIP Code.

Getting more retailers quickly was not an option. We had to be a little creative to get more of this data. If retailers are grouped together, the third-party provider does not need retailer consent to share. We gave the provider with lists of retailers by mixed codes, and they create monthly totals by states for each of the groups. We gave up that store-level granularity, but this was a really great way to get a large amount of valuable data quickly.

With this data on the figure, a blended approach to the estimates was the natural choice. Jenny Thompson, a senior mathematical statistician at the Census Bureau, and her methodology team developed a composite estimator that is a weighted average of synthetic and hybrid estimates. The weight used is a ratio of the variance of the synthetic estimate to the total variance of both estimates.

The synthetic estimates you can see on the left side of Fig. 5 make use of survey data and that payroll data. These

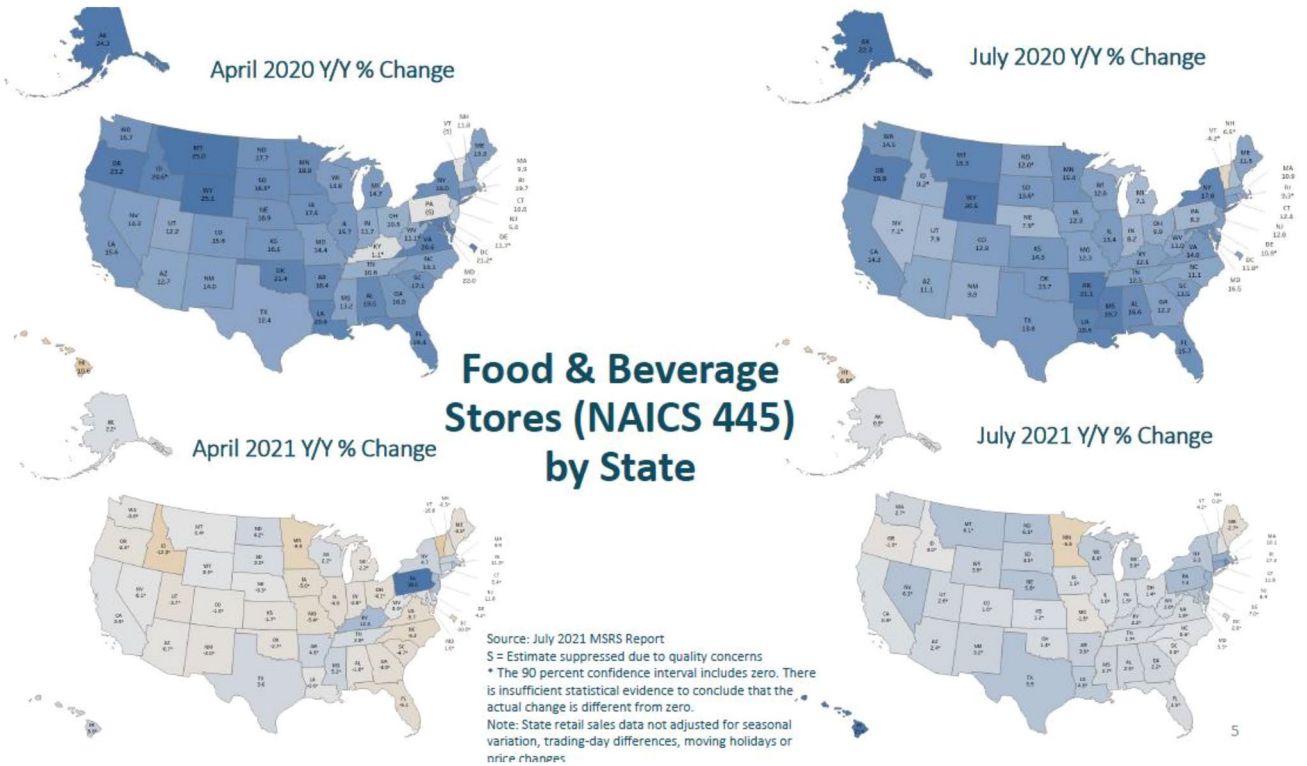
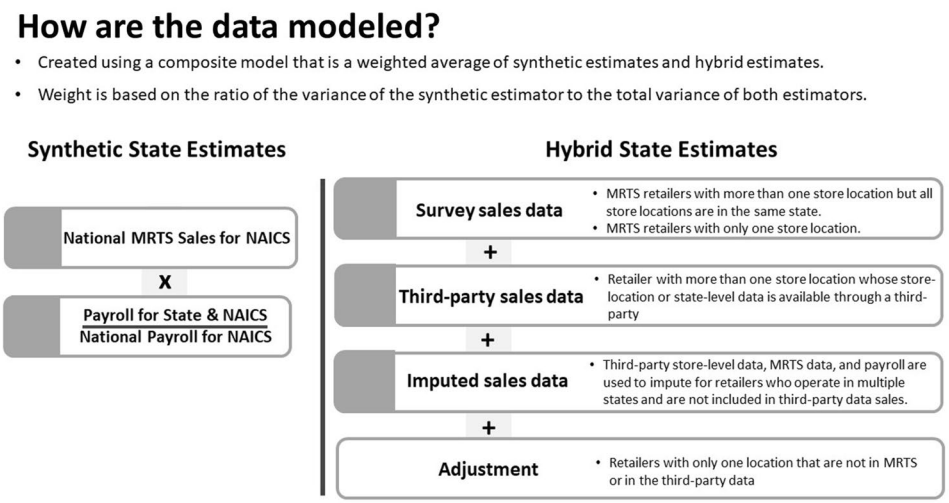


Fig. 4 Food and beverage stores (NAICS 445) by state

Fig. 5 Modeling the data



are simple to calculate using a ratio of the state payroll to the national payroll multiplied against the national Monthly Retail Trade Survey estimates. The hybrid estimates make use of all the data that were on the previous slide.

If a retailer on the Monthly Retail Trade Survey operates many store locations, but only in one state, or only has a single location, we use their survey data. The point-of-sale data that we have at either the store-level or at the aggregate level are then added in.

For those retailers with more than one location that we don't have survey data for, or third-party data, we impute for their sales using either the store-level point-of-sale data or a combination of survey data and payroll data. We currently do not have a good method for imputing retailers with only a single location. As a final step we apply an adjustment

factor to bring the data up to the national level to account for these locations.

These data are being improved on an ongoing basis, and we have a few research efforts underway. One requirement we have had to date with our third-party data are that we have to know all of the retailers included in the data. Because so many data sources with valuable retail information are out there, where this isn't possible we are exploring ways to look at this data and incorporate it in these models.

Currently we are working to see if, rather than only using a payroll ratio in the synthetic estimate, could we model that ratio using a variety of sources, including payment processor data, state sales tax data, and other data of that nature? We are also seeing if we can extend this methodology to other sectors of the economy, including services.

5 Ron Jarmin

Next up is Chris Kurz from the Federal Reserve Board.

6 Christopher Kurz

I'm going to be talking about blending data to better understand the economic impact of COVID-19. This work primarily goes through three papers that I've been jointly working on (Cajner *et al.* 2020a, b; Cajner, Crane, Decker, Hamins-Puertolas, Yildirmaz, and Kurz. 2020; Crane *et al.* 2021), and some ongoing research.

I don't really have to say that Big Data has played a critical role in tracking the 2020 pandemic, but I just want to put that out there as the start for this discussion. It's important to know that many new things came out, new data sources, some produced by the statistical agencies, like some things we've just seen, in addition to some of the work done by the regional Feds, location tracking information, credit card spending information, employment trackers.

A lot of this nontraditional data have come to the forefront, but it's still important to know that these official statistics in the background are the gold standard for tracking the economy. And the way I view thinking about blending data are, "How do we combine this information to get a better idea of what's happening in the economy?".

And I think the main takeaway point I'd like to focus on here is really that these nontraditional indicators are going to require a frame of reference. I'm going to tackle this question by really thinking about it through business exits and employment data.

The first one was thinking about a possible business exit surge in 2020. When we were asking this question mid-last year we really didn't know what was happening in the economy. We knew that the BLS data, the BED that would help us answer this question, wouldn't be out until late 2021, and the Census data weren't going to be out, maybe, until 2023. That might also depend on Ron.

I'm also asking the question, "What did blended employment data tell us?" And I think it's important to think about that: The framework we had for blending employment data sort of got called into question in the middle of the pandemic.

I'll first review a couple of small business exit trackers (Fig. 6) We had some credit card transaction data that was really valuable to look at firms that were no longer receiving transactions. That was telling us about 35% of these small businesses were gone, or exited.

And the same thing you're seeing in the Homebase data, which is a clock-in, clock-out tracker. But I think this sort

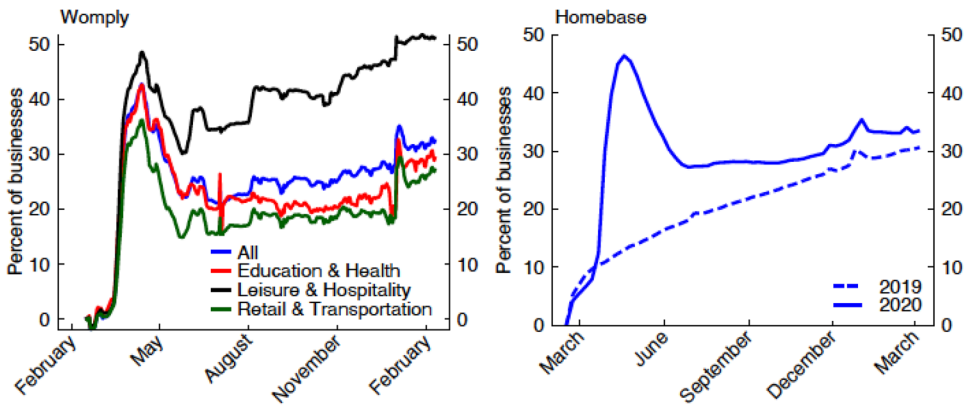


Fig. 6 Exit in small business trackers

- Womply: credit card transaction processor
- Homebase: clock in/out tracker

of contextualization that I'm pointing to is really important. You need to compare the 2020 moves to the previous year, 2019, to understand the magnitude of these differences.

So one of the most important things we did in the start of this project was, "Let's look at the BLS and Census traditional data to establish statistics, better understand what annual firm exit rates are," so that we can then compare that to what we were seeing in the current data in 2020.

Some stylized facts that you should take with you for the rest of your career should be that annual firm exit rates are about seven and a half percent; quarterly it's about eight and a half percent. Temporary closure is actually really common, about 2.0% of establishments per quarter temporarily close.

If you look at the latest BED data, we're thinking about 600,000 total deaths and about 150,000 to 200,000 excess deaths in 2020. This is something that we've only been able to conclude recently. We'll show how other data sources point us in that direction.

We also looked at the ADP microdata to look at shutdown based on the length of time that establishments or businesses within that data were not issuing pay. And, interesting, in August and September of last year we were seeing that the count of businesses that were closed was returning to normal. Being able to say that in August of last year was actually quite surprising, because at that time we weren't sure what was happening in the overall economy.

We also leveraged SafeGraph data to infer closure information based on, let's say, a 65% drop in foot traffic. And from that information we're able to conclude a lot earlier than that latest BED data that a little less than 200,000 establishments died in excess of normal in 2020.

Thinking about the SafeGraph data I just mentioned, we were able to use that location data to talk about restaurants that are temporarily closed (Fig. 7). You can see how that jumped up in March 2020, and the rate of temperature closure declined thereafter. Cumulatively, we were looking at the percent of restaurants that were permanently closed, and that went up to something like thirteen and a half by February of 2021.

I'm now going to turn my attention to another case for blended data. This is going to be similar to the CARTS we saw earlier, where you're taking different information sources and kind of fitting, let's say, an unobserved state. This is a common filter approach to employment information.

We're taking information from ADP and combining that with information with the CES, so basically creating our own data series from the underlying ADP microdata, and combining that with the published BLS data. Those two pieces of information were combined to measure what the underlying state was.

The noisy information for both series were being combined in the context that the point estimate for a given month of, let's say, private payroll employment from the BLS usually has a confidence interval around 110,000 employees.

Combining those two information sources gives us a more certain, let's say, point estimate of what's happening in private payrolls and improves the current estimate, like I said, but also improves, our understanding of what the next three months are likely to be. The Kalman Filter smooth state actually gives you a better idea of what employment's going to be in the next three months than either independent source (Fig. 8).

Importantly, though, we ran into the pandemic, where this idea of being able to combine two pieces of information together might have run into the fact that we were actually seeing large employment declines in extremely fast fashion.

Therefore we were completely reliant on the high-frequency aspect of the ADP/FRB microdata (Fig. 9). You can see the dramatic employment declines that we were only aware of having the nontraditional data in hand. The question is: Would this sort of need for timeliness become the norm, and reduce the importance of blending both of those data sources in real time?

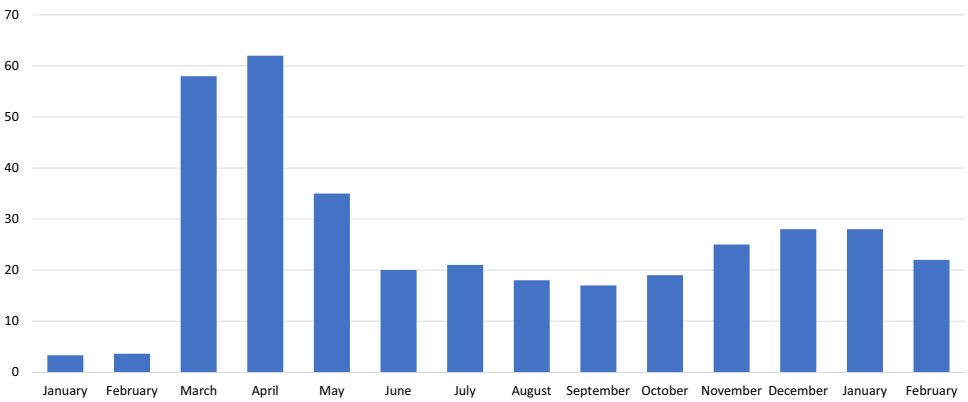


Fig. 7 Percent of restaurants temporarily closed

Source: Author's calculations from Safegraph microdata.

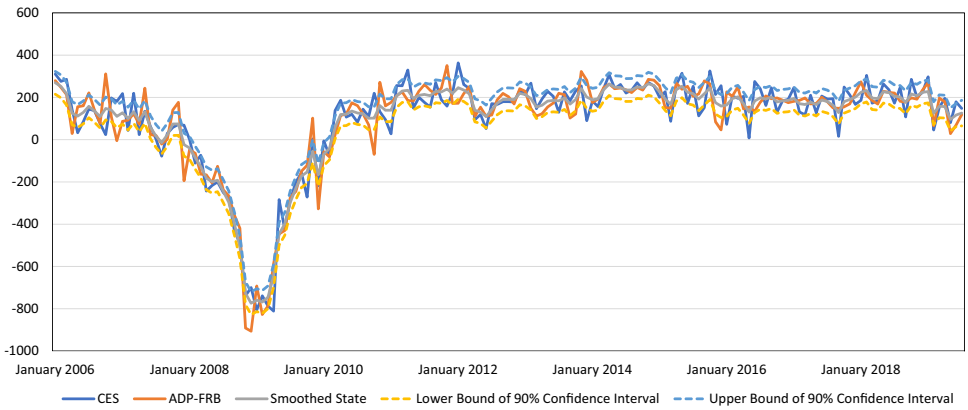


Fig. 8 Best case for blended data? Kalman filter approach

Source: BLS, ADP micro data, internal calculations.

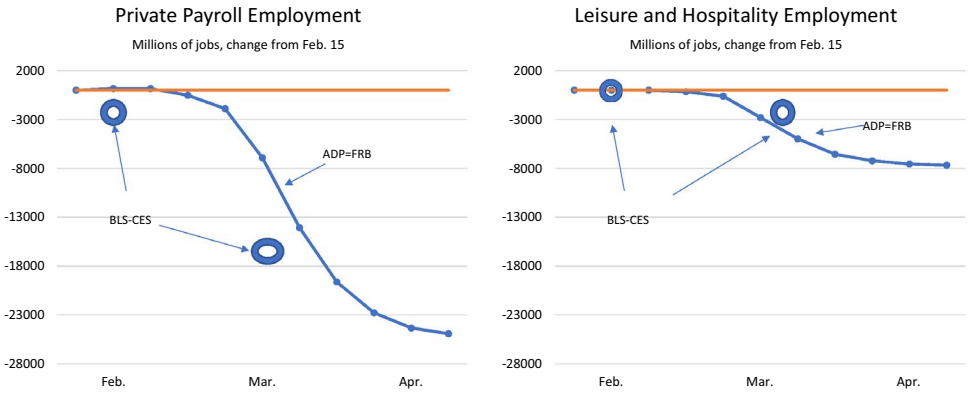


Fig. 9 Payroll employment during the pandemic

Source: FRB Staff calculations, ADP-LLP, Bureau of Labor Statistics

Source: FRB Staff calculations, ADP-LLP, Bureau of Labor Statistics

I would say probably not. In most cases it takes a lot longer to hit peak changes in a downturn. During the Great Recession it took something like 16 months to reach our largest employment decline, whereas in the pandemic recession, it only took two months: things moved a lot faster and sort of moved the pendulum more towards nontraditional data.

My conclusions are: Blending data are important and necessary; it allows for the proper weighting and benchmarking. But you have to understand the context, and that's where blending data are really important. You can either do it from external sources to aid in understanding what these traditional metrics look like by comparing them to the actual exit rates we see in the historical BED or BEDS data.

Or you do it from within: Compare current measures to what we've seen in the past, so you can contextualize what those changes have been.

One takeaway point from this two-threaded research has been that last year's business exit was elevated, but it wasn't as dramatic as a lot of us expected. That's a shocking conclusion.

Then we see the high-frequency employment data are invaluable, but it also raised ideas about this methodology, about how we combine things in real time, and moved us more towards thinking that the nontraditional data were more important within the context of the pandemic.

I would just walk away from this presentation saying that blending data are one of the most important things to provide context and the ability to understand nontraditional data sources.

7 Ron Jarmin

Our last paper is by Michael Stepner from the University of Toronto talking about some of the work from the Opportunity Insights team.

8 Michael Stepner

I'm going to be presenting a large blend of datasets that I put together in collaboration with Raj Chetty, John Friedman, Nathan Hendren, and a really fantastic team of young economists working at Opportunity Insights. There are so many more papers that have been using transaction data, private sector data, to analyze the economic impacts of COVID-19 over the course of the last 18 months.

What we really focused on within our team was, I think, two important things. First was taking private sector data and converting it into public statistics, so that we have in common with all of the folks here. But traditionally many of the researchers will sign a confidentiality agreement and then analyze private sector data.

Our challenge here was to take this private sector data, make it public in a way that's useful both for our own research and for the research of others, and then also combine, blend, this data on spending, employment, job postings, business revenue, and other outcomes, and use that to form a more complete picture of the chain of macroeconomic events than any study that focuses on one set of outcomes alone.

In this project we combined data on spending, revenues, employment, job postings, and education. We're blending both within categories and across categories. I want to give a high-level overview of two stories that we can learn about what has transpired over the last 18 months.

We are taking raw data and performing many of the same procedures that my colleagues have described. We're cleaning the data; smoothing out seasonal fluctuations; excluding small cells; and combining data from multiple companies to protect the privacy of companies and their customers; and then benchmarking to the gold standard, the national statistics, which might be less frequent or less granular, but guarantee us a representative sample.

From this benchmarked, processed data we can tell two stories. Let me start from the first one: What happened at the start of the pandemic? Here (Fig. 10), by blending data on multiple segments of the economy, we can see consumer spending falling, especially among high-income Americans. We can tell a story of contagion: How that fall in consumer spending led to business revenues falling, especially among those small businesses serving high-income Americans; and then employment falling, especially so among low-income workers in high-income areas.

I'm going to walk through that story, and then conclude by talking about what's happening today. At the start of the pandemic, folks in the top-income quartile were responsible for the preponderance of the decline in aggregate spending.

In the first weeks, 40% of the decline in aggregate spending came from the top-income quartile. By the middle of the summer, and into the fall, they were responsible for 50% of the total spending decline. If you look at the bottom of the income distribution, only about 12% at the beginning, and 5% in the middle of the pandemic of that aggregate spending decline was being driven by the changes in behavior of low-income households.

We can see that translating into business revenue. Look, say, at a map of San Francisco (Fig. 11). If you're familiar with San Francisco you'll see the most affluent areas of San Francisco had the largest declines in small business revenue

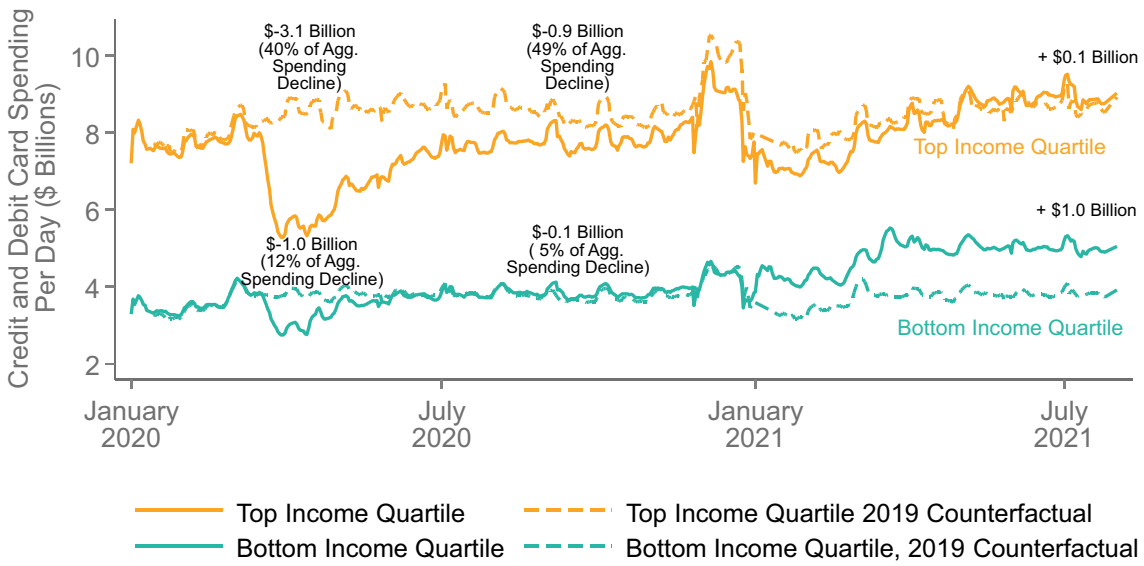


Fig. 10 Consumer spending by income quartile

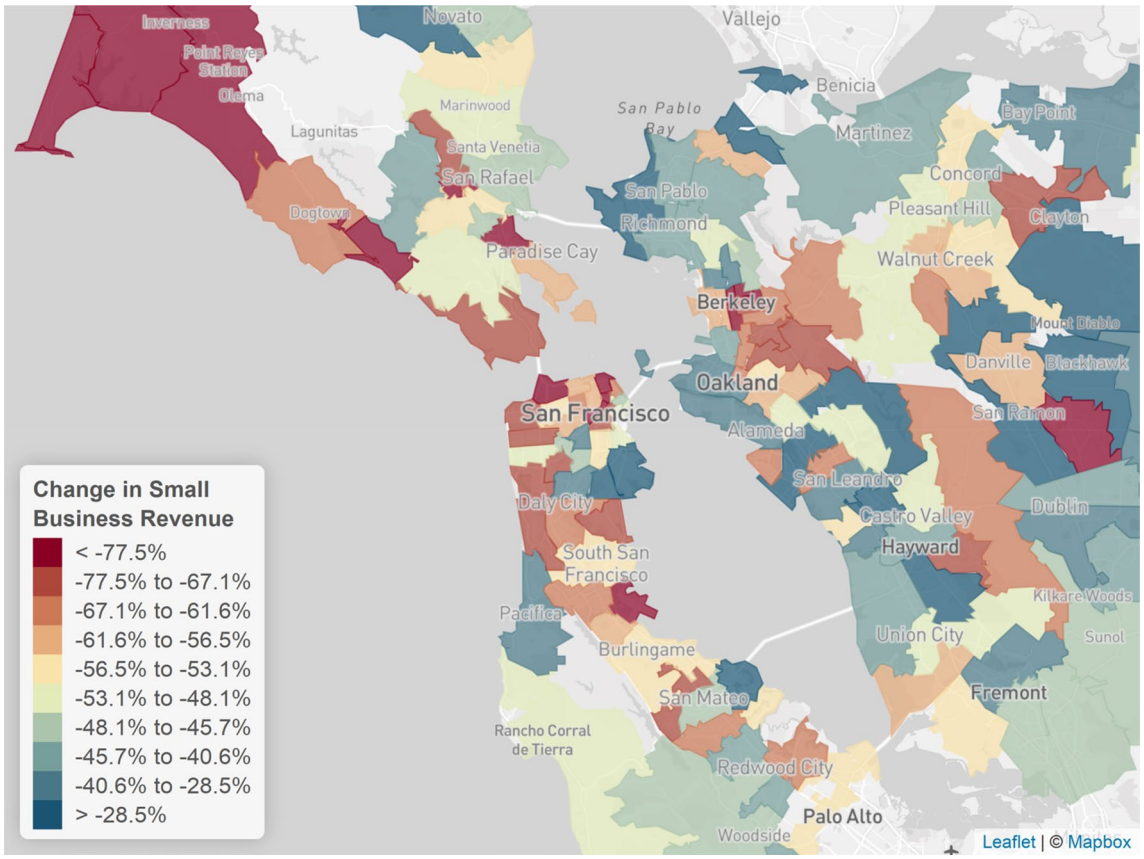


Fig. 11 Changes in small business revenues from January to April by ZIP code. San Francisco

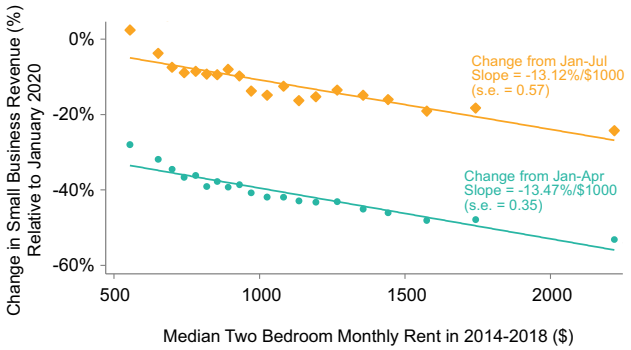


Fig. 12 Changes in small business revenues vs. rent, by ZIP code

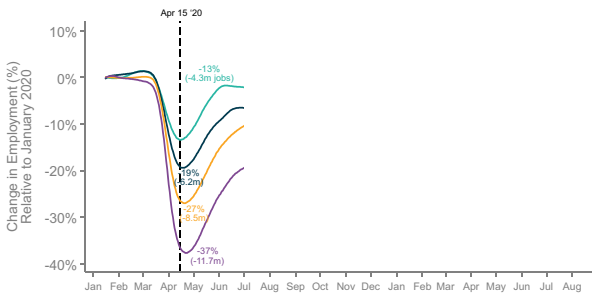


Fig. 13 Employment changes by income quartile

at the start of the pandemic, whether that's Downtown San Francisco, the Silicon Valley, or the area around Berkeley.

This is just an example. But if we go broader, outside of San Francisco, and you compare on the X axis, you look at monthly rents in an area, the idea being that areas with high monthly rent are areas with a large number of people with a fairly substantial income, those tend to be high-rent areas.

You see this consistent pattern, with larger declines in small business revenue in the higher-rent areas of the United States (Fig. 12). That was true in April, and we can see that continued to be true in July. Everywhere had a recovery, but you still had this differential loss.

That translated into employment losses (Fig. 13). Those employment losses were concentrated in the lowest income quartile. The bottom-income quartile that lost 40% of their employment at the trough of the pandemic, while the top-income quartile also declined, but by only 13%. And once again those losses were concentrated in the areas that were most affluent.

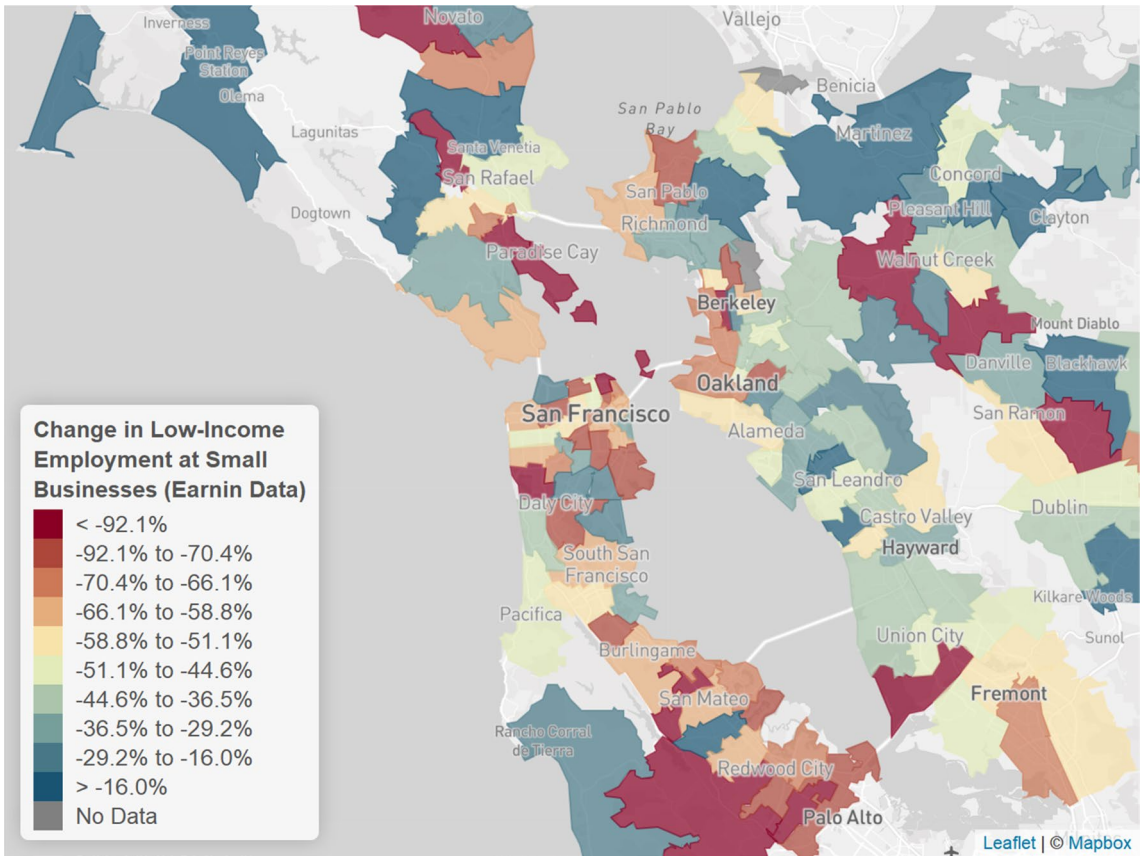


Fig. 14 Changes in low-wage (bottom quartile) employment rates by ZIP code. San Francisco

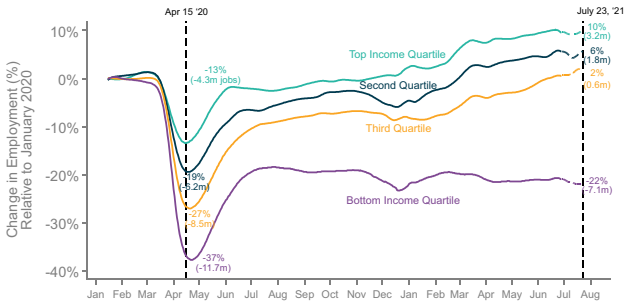


Fig. 15 Employment changes by income quartile

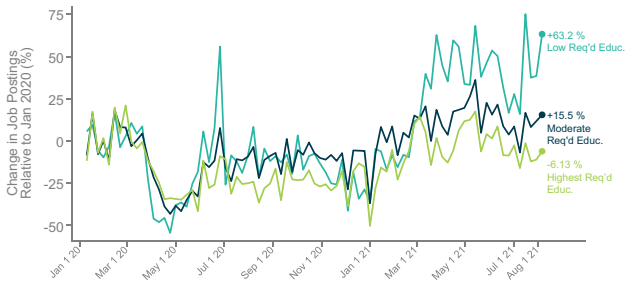


Fig. 16 Job postings by required education level

Within a map it's those same areas in San Francisco, the Bay Area, around Berkeley, that had the largest declines in low-income employment (Fig. 14).

Let's conclude by talking about what's happening right now. Here we tell a story of really a bifurcated recovery where consumer spending has recovered to pre-pandemic levels, and frequently above. Employment has fully recovered for the top three-quarters of the wage distribution, but employment has totally stagnated at the bottom of the wage distribution (Fig. 15). We have more than seven million missing low-wage jobs. We can ask why hasn't employment recovered at the bottom.

Blended data tells us that labor demand is unlikely to be the cause. We can actually see a surge in job postings, especially for low-skill jobs. If we look to data on job postings collected by Burning Glass Technologies (Fig. 16), we see jobs with low required education had much higher job postings than in a typical year by the summer of 2021.

The question is, "Well, what's going on on the labor supply side?" There's been research using both our data and others showing that this high-level of unemployment

insurance has not really had a big effect on the labor supply. That leaves a couple open questions about the extent of labor market mismatch: how the needs of jobs and employers have diverged from the supply of labor for that; or this change in preferences, the story of the Great Resignation that is now so prevalent.

What I want you to leave with is that this blended data can tell a very detailed story of propagation through the macroeconomy, both in our own analyses and in the other work that folks have done using our data. So let me conclude there, pass it back.

9 Ron Jarmin

Let me take the chair's prerogative here. Obviously, the title of the session was using blended data to understand the COVID-19 pandemic.

But we could generalize a little bit. And let's say how do you all see these being useful in maybe sort of more run-of-the-mill business cycle, or other shocks to the economy, maybe a large-scale natural disaster or something like that?

To speak to some of the difficulties of working with these data, how important it would be to incorporate them into the regular measurement infrastructure, and how useful they could be to help businesses and policymakers react to shocks?

10 Michael Stepner

I think that alternative data sources, and blended data, provide two differentiations from a traditional survey database. The first is the timeliness, and the second is the granularity. In a normal business cycle, I suspect the timeliness at a national level is much less important. Things are pretty stable at a national level and don't fluctuate wildly like during COVID-19.

I think that that's where the granularity can really shine. If we look at our data, we can see these sharp spikes in early 2021 in Texas and Louisiana, when we had that cold snap that led to the shutdown of the Texas economy. If a hurricane hits Louisiana or the Eastern Seaboard, you can pick that up and really understand, in almost real time, how that's affecting the economy.

I think, going forward, with those types of granular changes, which are difficult to pick up in surveys, we will know more and more about what is happening in small areas, and to specific subgroups of people, whether this is hurting a specific sector or high-income people versus low-income people, we'll get a much quicker appreciation of how regional or sectoral changes are affecting the economy.

11 Ron Jarmin

Let me just follow that up. Obviously the COVID-19 pandemic was a big deal that everyone was paying attention to, and I think folks were willing to share some data.

But let's think about a more mundane time, like the hurricane on the Gulf Coast, where it's not something that's got everyone's undivided attention, just a relatively small portion of the population. To have these sorts of datasets in place in time to look at that, what are some of the challenges that you all have encountered in gaining access, and then making sense of, the data?

12 Rebecca Hutchinson

I think for us one of the biggest challenges with these datasets is the cost; they're pretty expensive to acquire. Other issues we ran into with getting, especially the most granular data, down to the retailer, is you do need retailer consent. A lot of retailers don't want to provide that.

I know as we look to expand this concept to other sectors of the economy, with retail you're pretty lucky with the kind of data that's available out there. Finding similar data sources for the service sector, which is a pretty diverse part the economy, is a challenge we are trying to work through right now.

13 Ron Jarmin

Somebody had a question in the chat that's very related to this. It's about the transparency that you need. We're using data that maybe not everybody gets to see how the sausage was made. How do we ensure the data users, especially decisionmakers, whether they be central bankers, whether they be CEOs, or folks running NGOs, or just average folks: How do we ensure that they understand the transparency? And then how do you ensure that you have access to this over time?

14 Christopher Kurz

That's just a huge problem. It's a problem, from our perspective as, let's say, someone in the middle of providing this data, or understanding it, and I think it's even more of a problem for just your average consumer, who might be downloading it from a website.

I know the Board, different Reserve banks, and statistical agencies have had a lot of difficulties trying to break



through that and better understand the underlying microdata. I think it's imperative for the organizations to realize this and work with these actors to know that, for example, the Census Bureau really needs to be able to look under the hood of something that they're going to be able to try to provide to the public. Getting everybody on that same page is really important. And I think only through good user agreements, understanding how to protect confidentiality, and provide it in public methodology are we going to get to that point.

15 Scott Brave

I would second that. This was the biggest hurdle for us to jump, and we would have been able to start making this public a year ago but there were a lot of data sources that we had to eliminate because of these concerns. Something changed right away in the middle of the process.

For us to get down to the five that we got down to, these were the companies that were willing to work with us to help us understand what was going on, and willing to give that kind of information, and allow us to make these kinds of things public, even if we're not publishing the data, which we don't.

I think that that's probably unrealistic to expect of them always, if they're making money off the data. But the methodology and understanding how all the sausage is made was very important to us. It took us a while to find partners that were willing to do that, but I think once they understood where we were coming from and what we were trying to get out of this, and really the policy uses of what we're trying to do with the bigger picture, we were able to make headway on that.

16 Christopher Kurz

I just want to jump back to that first point and say that I think we opened the door to a lot more granularity in the past year. I would say timeliness let us know that something was happening, but the granularity helped us understand what was happening.

I'd actually say that the verdict is really going to be out for, unfortunately, the next downturn for us to understand which of those is going to be more important because the timeliness was so valuable. If you really jump back to the Great Recession, how many months did it even take the NBER to call that recession, and really understand what was happening in the economy? When we have a "normal", and I want to put quotes around that, downturn; when we hit that sort of peak, then we're going to really understand how some of this investment into non-traditional data are really going to pay off.

17 Ron Jarmin

Central banks and statistical agencies can't pay every company in the economy for their data. What benefits do you think that these companies could get from allowing more access to the data for these more public uses of their data?

18 Michael Stepner

I would say we're in early days for the use of private sector data for public sector statistics. At the beginning of COVID-19 incentives were aligned: The economy was in freefall; they had valuable information that they typically use just for their own internal business decisions; they wanted to make that available to policymakers, to journalists, to everyone trying to understand the pandemic. Because if that helped policymakers diagnose what was going on and assist the economy, that was in their own business interests. Everyone's incentives were aligned.

I think, going forward, I expect we will see more partnerships between large companies in the banking sector, in the tech sector, who see the value of their data for the public good. And I think corporate social responsibility will play into that.

But the idea that they have data which is valuable for American policymaking, which they can share without compromising the privacy of their users or their customers or their own business prospects: I think that message will start to lead to a greater expansion of partnerships. And hopefully lead to a new era in the construction of national statistics, where we have more reliable, more granular data, even more so than what we are showing in these early days.

19 Ron Jarmin

Well, I share your hope for that optimistic picture of the world.

References

Brave, Scott A., Ross Cole, Michael Fogarty, Ezra Karger, and Liam Puknys. 2001. Introducing CARTS: A new index tracking national retail spending. *Chicago Fed Insights* June 12. https://www.chicagofed.org/publications/blogs/chicago-fed-insights/2021/introducing-carts.

Brave, Scott A., Michael Fogarty, Daniel Aaronson, Ezra Karger, and Spencer Krane. 2001. *Tracking U.S. consumers in real time with a new weekly index of retail trade.* Working Paper 2021–2025, Federal Reserve Bank of Chicago. https://www.chicagofed.org/publications/working-papers/2021/2021-05.

Cajner, Tomaz, Leland D. Crane, Ryan A. Decker, John Grigsby, Adrian Hamins-Puertolas, Eric Hurst, Christopher Kurz, and Ahu Yildirmaz. 2020. *The U.S. labor market during the beginning of the pandemic recession.* National Bureau of Economic Research Working Paper 27159. https://www.nber.org/papers/w27159.

Cajner, Tomaz, Leland D. Crane, Ryan A. Decker, Adrian Hamins-Puertolas, Ahu Yildirmaz, and Christopher Kurz. 2020. *Improving the accuracy of economic measurement with multiple data sources: The care of payroll employment data.* https://www.nber.org/system/files/chapters/c14272/c14272.pdf

Crane, Leland D., Ryan A. Decker, Aaron Flaaen, Adrian Hamins-Puertolas, and Christopher Kurz. 2021. *Business exit during the COVID-19 pandemic: Non- traditional measures in historical context*. Finance and Economics Discussion Series 2020–089r1. Washington: Board of Governors of the Federal Reserve System, https://doi.org/10.17016/FEDS.2020.089r1

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Scott Brave Scott A. Brave is the lead consumer spending economist for decision intelligence company Morning Consult, where he leverages the company's high-frequency proprietary data in forecasting and predictive analytics to help clients make better business decisions. Previously, Brave led the releases of economic activity and financial conditions indexes as a senior economist in the Economic Research Department of the Federal Reserve Bank of Chicago. His research expertise has focused on the use of high-frequency data in forecasting applications, mixed-frequency econometric modeling, and sports economics, with publication in journals including the *International Journal of Forecasting*, *International Journal of Central Banking*, *Growth and Change: A Journal of Urban and Regional Policy*, *STATA Journal*, *Journal of Sports Economics*, and *Journal of Sports Analytics,* as well as Chicago Fed research publications, such as *Economic Perspectives* and the *Chicago Fed Letter*. Brave earned his bachelor's degree in economics with honors from the University of Chicago and an MBA with concentrations in economics, statistics, and finance from the University of Chicago Booth School of Business.

Rebecca Hutchinson is the Economic Indicator Division's Big Data Lead at the Department of Commerce's Census Bureau. She leads agile teams that explore the use of alternative data sources including satellite imagery, payment processor data, scanner data, administrative data, and web-scraped financial filings to enhance the measurement of the nation's economy, either through the improvement of existing data products or the creation of new products. Her recent work has focused on using alternative data sources to create a Monthly State Retail Sales data product. Additionally, she serves as a champion for the Bureau's data science training programs, which seek to teach important data science skillsets to the existing workforce.

Christopher Kurz is the section chief of the Industrial Output section in the Division of Research and Statistics at the Federal Reserve Board. His responsibilities at the Board of Governors have ranged from forecasting industrial production and motor vehicle sales to coordinating the staff forecast of GDP. Christopher's research interests are similarly broad and cover international trade, international finance, nontraditional data, and economic history. Research projects of Christopher's have employed microdata to quantify the effects of international trade on employment volatility at the firm level and back tested systemic risk measures using historical banking data. Over the past five years, Chris has been working on the Federal Reserve Board's Expanded Measurement Agenda (EMA). The EMA is an attempt to identify novel data sources and leverage new techniques that can be used to develop timely and accurate measures of economic activity that are relevant to the formation of monetary policy.

Michael Stepner is an assistant professor at the University of Toronto and a research principal at Opportunity Insights. His research examines the relationship between health and economic inequality, with a focus on how public policy can improve the health and financial security of low-income populations. He also serves as the network leader for Health Trends and Inequalities research at the NBER Center for Aging and Health Research, and as the executive director of the Opportunity Insights Economic Tracker. He received his PhD from MIT in 2019, and his dissertation research was awarded the top dissertation award from the National Academy of Social Insurance.

Ron Jarmin is the U.S. Census Bureau's deputy director and chief operating officer and has been since January 2019. He performed the nonexclusive functions and duties of the director from July 2017 to January 2019 and from January 2021 to November 2021 and previously served as the associate director for economic programs. He led the team for the 2017 Economic Census, overseeing a move to 100 percent Internet data collection and leveraging enterprise investments to minimize system, application, and dissemination costs. Data products from the economic census provide the foundation for key measures of economic performance, including the nation's gross domestic product. From 2011 to 2016, Dr. Jarmin served as the assistant director for research and methodology. He oversaw a broad research program in statistics, survey methodology, and economics aimed at improving economic and social measurement within the federal statistical system. Since beginning his career at the Census Bureau in 1992, he has also served as the chief economist, chief of the Center for Economic Studies, and a research economist. Dr. Jarmin holds a Ph.D. in economics from the University of Oregon. An elected fellow of the American Statistical Association, he has published papers in the areas of industrial organization, business dynamics, entrepreneurship, technology and firm performance, urban economics, data access, and statistical disclosure avoidance.

