



Federal Reserve Board experience using transaction data

John Stevens

Prepared for the CNSTAT Workshop on “Scope, Components
and Key Characteristics of a 21st Century Data Infrastructure”
December 9, 2021

Any views expressed here are those of the presenter and not those of the Board of Governors of the Federal Reserve System or the Federal Reserve System.



How/why we use transaction data

- Fill in *knowledge gaps* with timeliness, higher frequency, broad coverage, greater granularity (geographic, industry, product, etc...)
- Create alternative estimates that *complement* official statistics or that *verify movements* in official statistics
- *Answer questions* that require more granular data
- Help with a *narrative* about economic developments



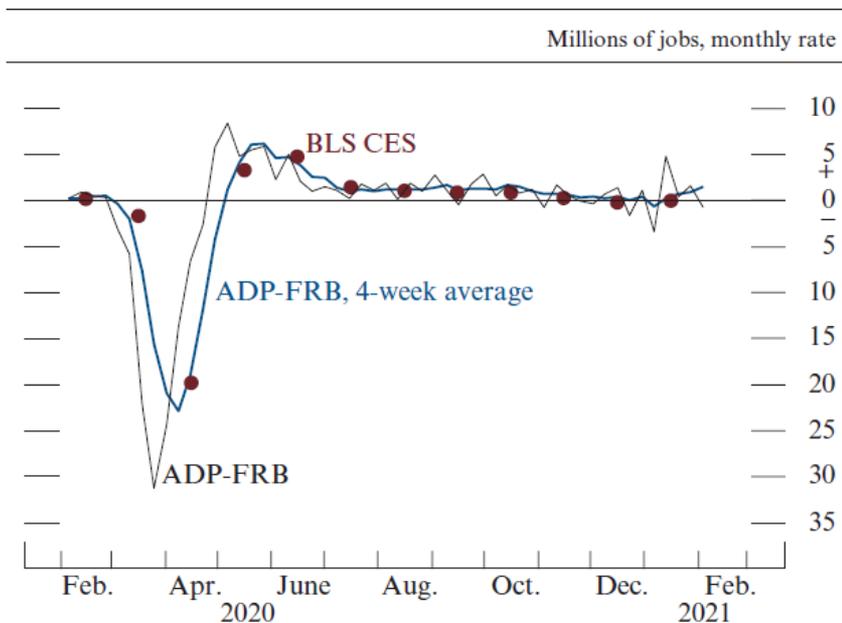
Value of transaction data

- Value of high-frequency data varies over time
 - At a FESAC meeting in mid-2020, I noted that the value of such data is increasing in the size of a shock or the arrival rate of shocks.
 - More recently, Domenico Giannone from Amazon, said at a “Big Data and Machine Learning Conference” (co-sponsored last month by the Board) that “high-frequency data are not so helpful in ‘normal times’ because of high-frequency noise”
- Suggests taking a longer perspective when assessing the costs and benefits of using such data
- Suggests value of combining data sources using statistical models to summarize movements and/or extract more signal
 - Example: The Fed’s Common Inflation Expectations index (which is a research series)

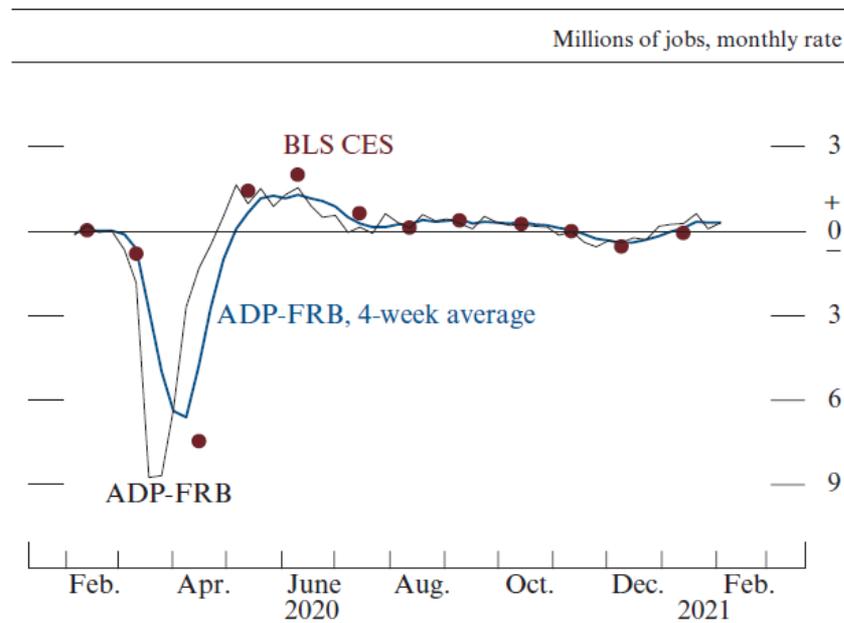
Example 1: Payroll Processor Data

A. Estimates of private payroll employment growth

Aggregate payroll employment growth



Payroll employment growth in leisure and hospitality



NOTE: ADP data are weekly and extend through February 6, 2021. BLS data are monthly.

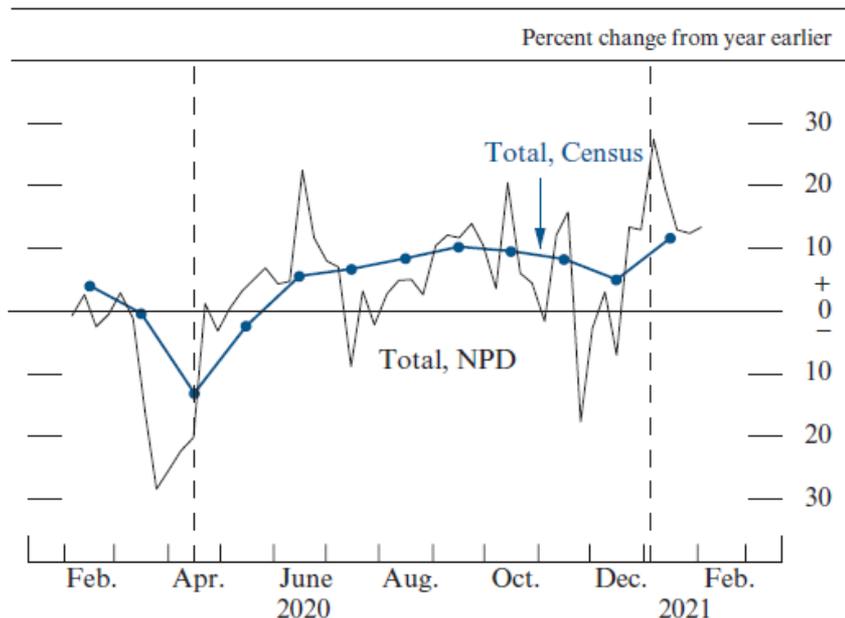
SOURCE: Federal Reserve Board staff calculations using ADP, Inc., Payroll Processing Data; Bureau of Labor Statistics (BLS), Current Employment Statistics (CES).

Chart is from: *Monetary Policy Report*, February 19, 2021, Board of Governors of the Federal Reserve System.

Example 2: Spending Indicators

B. Indicators of consumption growth

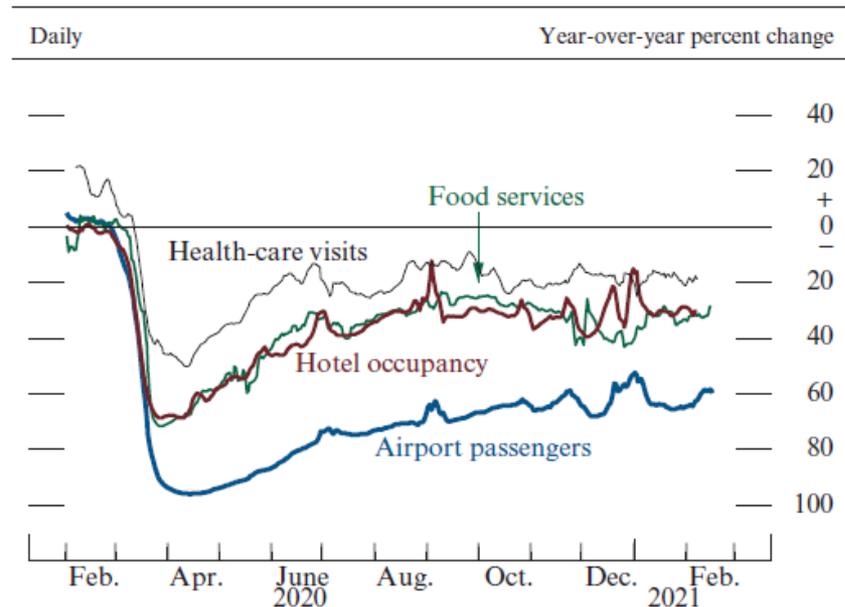
Retail goods spending



NOTE: NPD data are weekly and extend through February 6, 2021, and Census data are monthly. All series show nominal spending on nonfood retail goods. Dashed lines represent the first and second waves of stimulus tranche.

SOURCE: NPD Group; Census Bureau.

Services spending

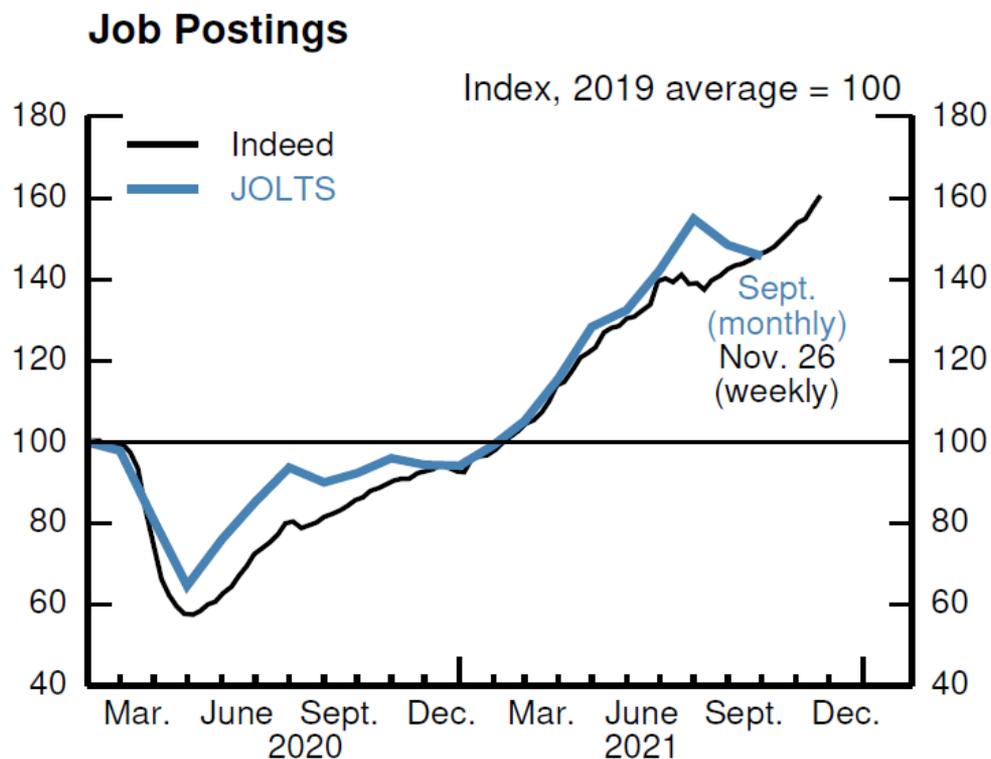


NOTE: Year-over-year percent change in 7-day moving average. Health-care visits data extend through February 7, 2021; food services data extend through February 15, 2021; and hotel occupancy data extend through February 6, 2021.

SOURCE: SafeGraph, Inc.; Fiserv, Inc.; STR, Inc.; Transportation Security Administration.

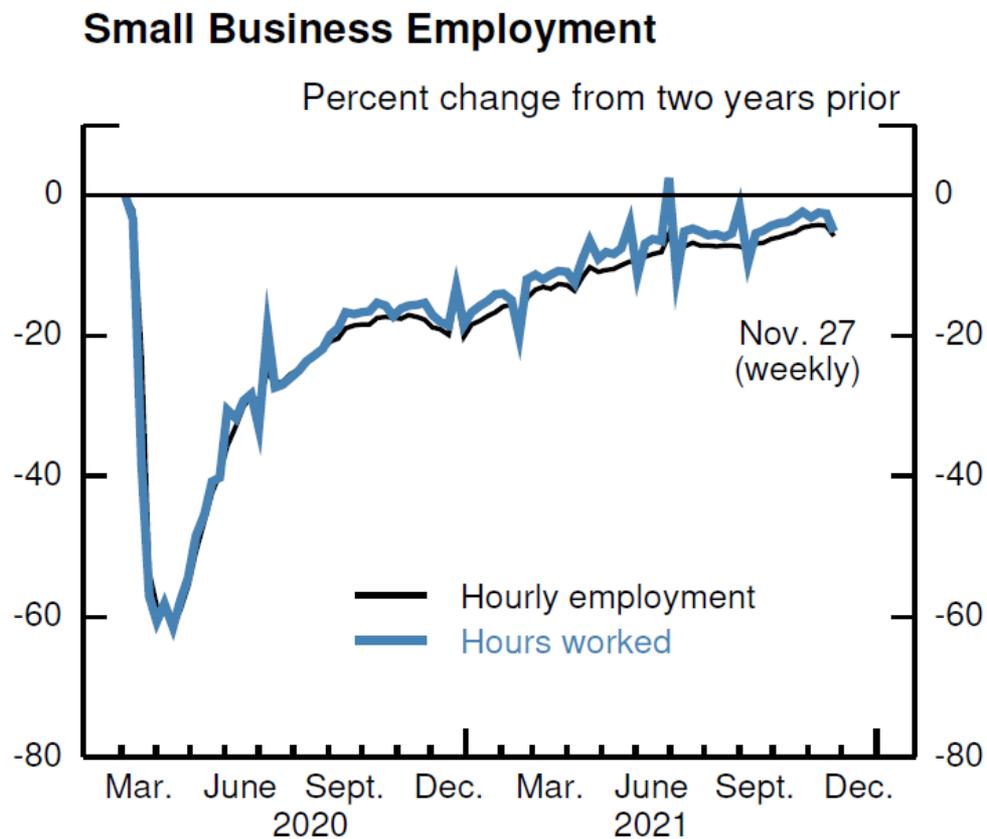
Chart is from: *Monetary Policy Report*, February 19, 2021, Board of Governors of the Federal Reserve System.

Example 3: Job Posting Indicators



Note: JOLTS is seasonally adjusted. Indeed is a 7-day average.
Source: Indeed, Inc.; Bureau of Labor Statistics.

Example 4: Employment Indicators



Source: Homebase. Data show their customer base of 60,000 small businesses, concentrated in leisure, hospitality, and retail industries.

Example 5: Search/Social Media

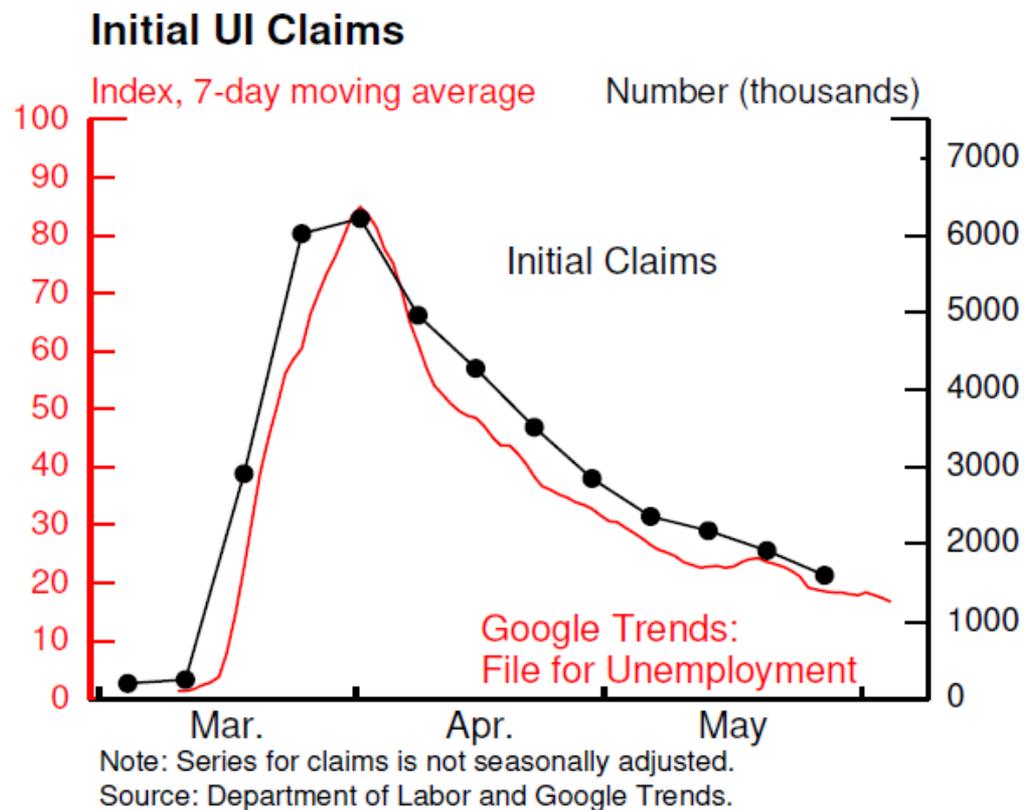
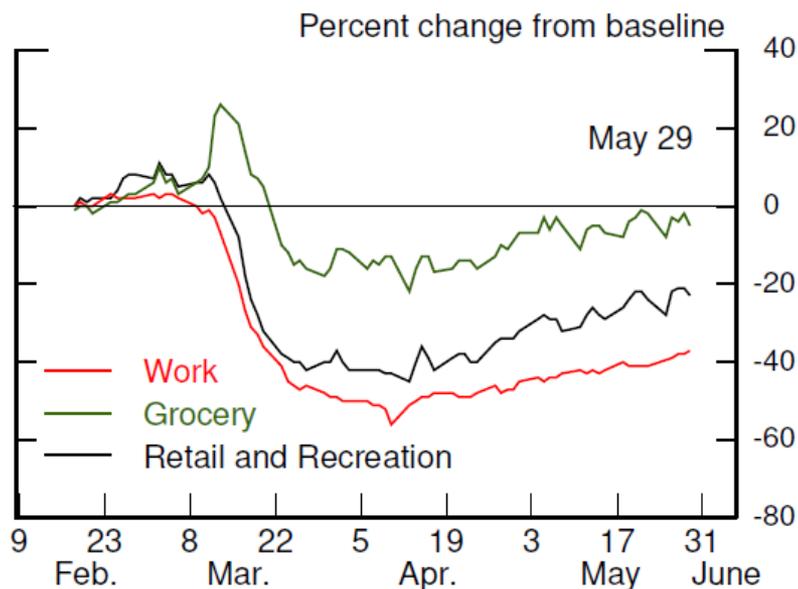


Chart from: Stevens, John. "The Potential for Alternative Data in Official Statistics," Federal Economic Statistics Advisory Committee Meeting, June 2020.

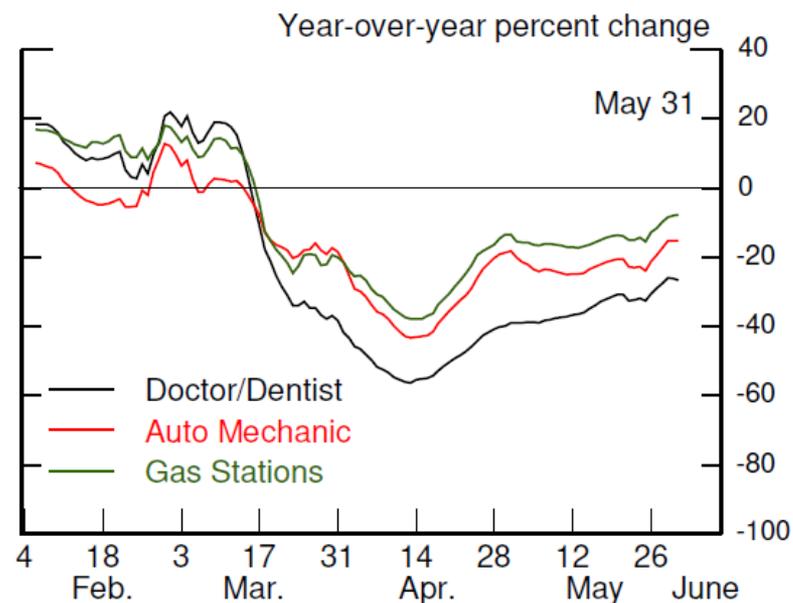
Example 6: Mobility Data

Google Mobility Index



Note: Baseline defined as median of Jan. 3 to Feb. 6 period.
Source: Google Mobility Reports.

Daily Visits to Select Locations



Source: SafeGraph, using cell phone geo-location data.

Charts from: Stevens, John. "The Potential for Alternative Data in Official Statistics," Federal Economic Statistics Advisory Committee Meeting, June 2020.



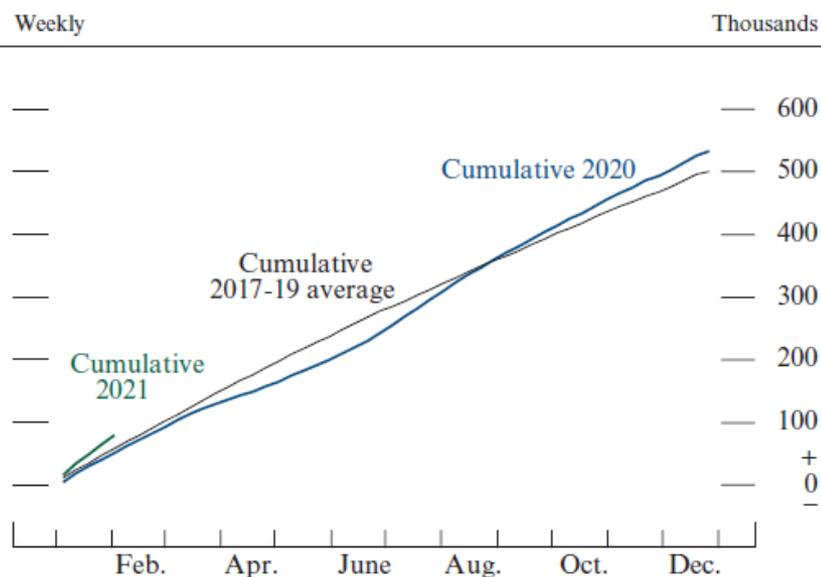
Additional Examples of Data Types...

7. Real estate vacancies
8. Prices for shipping
9. Prices for commodities
10. Production/capacity for various goods

An Aside on Non-Transaction Indicators

C. High-frequency indicators by official statistical agencies

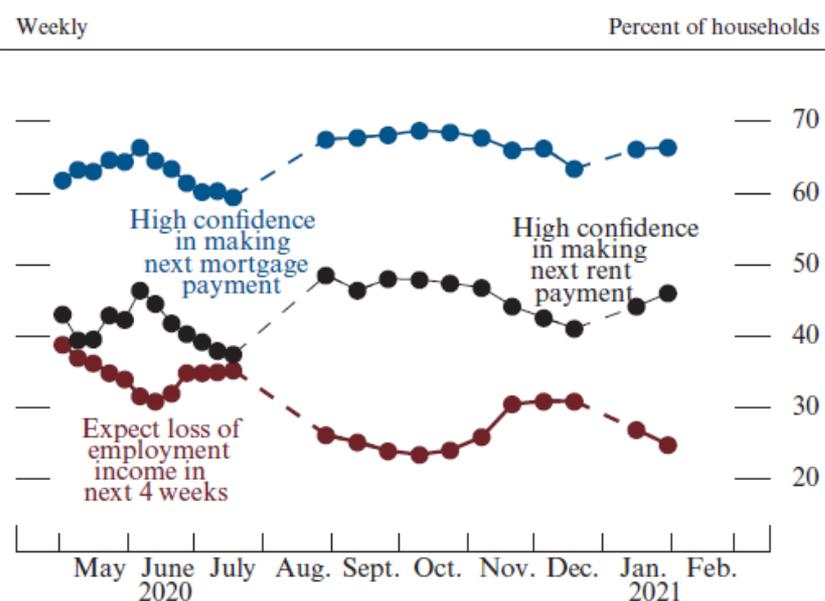
New business applications



NOTE: The cumulative 2021 data extend through February 6, 2021. The data are derived from Employer Identification Number applications with planned wages.

SOURCE: Business Formation Statistics, Census Bureau via Haver Analytics.

Household expectations



NOTE: Data extend through February 1, 2021. Dashed lines represent pauses in Household Pulse Survey data collection.

SOURCE: Household Pulse Survey, Census Bureau via Haver Analytics.

Chart is from: *Monetary Policy Report*, February 19, 2021, Board of Governors of the Federal Reserve System.



An Aside on Non-Transaction Indicators, continued

- FRBNY Survey of Consumer Expectations
- Atlanta Fed/Chicago Booth/Stanford Survey of Business Uncertainty (SBU)
- Many Federal Reserve Bank surveys for manufacturing, services, energy, agriculture (often of a diffusion index form)



Lessons Learned

- Acquisitions process takes a team
 - For us, we have a team of data librarians manage the acquisition process, including working with SMEs, our attorneys, procurement, and the vendor
 - Economists and other SMEs participate in data trials
- Every data provider is a little different
 - Need to partner with provider to understand their business, the way in which the data are generated and captured, and their unique sensitivities
 - Need to work closely with the provider to address measurement challenges
 - Can't be complacent, as new challenges will arise
 - Bespoke engagements don't scale
- Getting access to microdata is challenging
 - Second-best outcome is having transparency into the transformations and cleaning that is done to the underlying data
 - Third-best outcome is aggregated data with limited transparency (may be good enough for indicator usage but not for creating official statistics)



Lessons Learned, continued...

- Look for “choke points” in the flow of data
 - Organizations at those points have large volumes of data, and *possibly* more representative data
 - However, the amount of detail may be more limited than what you can get from other sources
- Be prepared to manage a multitude of challenges, including:
 - Adjusting for potential biases in the data
 - Adjusting for non-economic variability
 - Communicating findings (including the tension between wanting the same signal as corresponding official statistics and wanting a unique signal)
 - Long-term availability is harder to ensure than in-house data collections
 - Risk of hold-up in contract negotiations
 - Risk of reduced innovation (from not looking at additional data sources due to required cost and effort)