Expanding the frontier of economic statistics using big data

Paper by Dunn, English, Hood, Mason, and Quistorff

Discussion by Ryan Decker
Federal Reserve Board

Prepared for the June 2024 FESAC meeting
June 14, 2024

Without implication, this discussion draws on joint work and discussions with Tomaz Cajner, Leland Crane, Adrian Hamins-Puertolas, and Chris Kurz; and I thank Dave Byrne and Kurt Lewis for helpful conversations.

The analysis and conclusions set forth here are those of the author and do not indicate concurrence by members of the Federal Reserve staff or the Board of Governors.

Today’s discussion

1. Some challenges using nontraditional data
2. What is the best way for stat agencies to use private payroll data?

1. Challenges with using nontraditional data generally

- Hold-up problem/changes to terms of access
- Selection/representativeness (some can be addressed with weights as these authors do; but some unobservable selection)
- When data are a byproduct of business processes, changes to business processes -> changes to measurement concepts
- Short time series (e.g., how many recessions?)
- Industry codes: the private sector may handle this differently from official data (Barnatchez, Crane, and Decker 2017).
- Sample turnover (e.g., Crane et al. 2022)

(See Stevens 2021 and Abraham 2022 for thorough discussion)

Challenges with using private sector payroll data

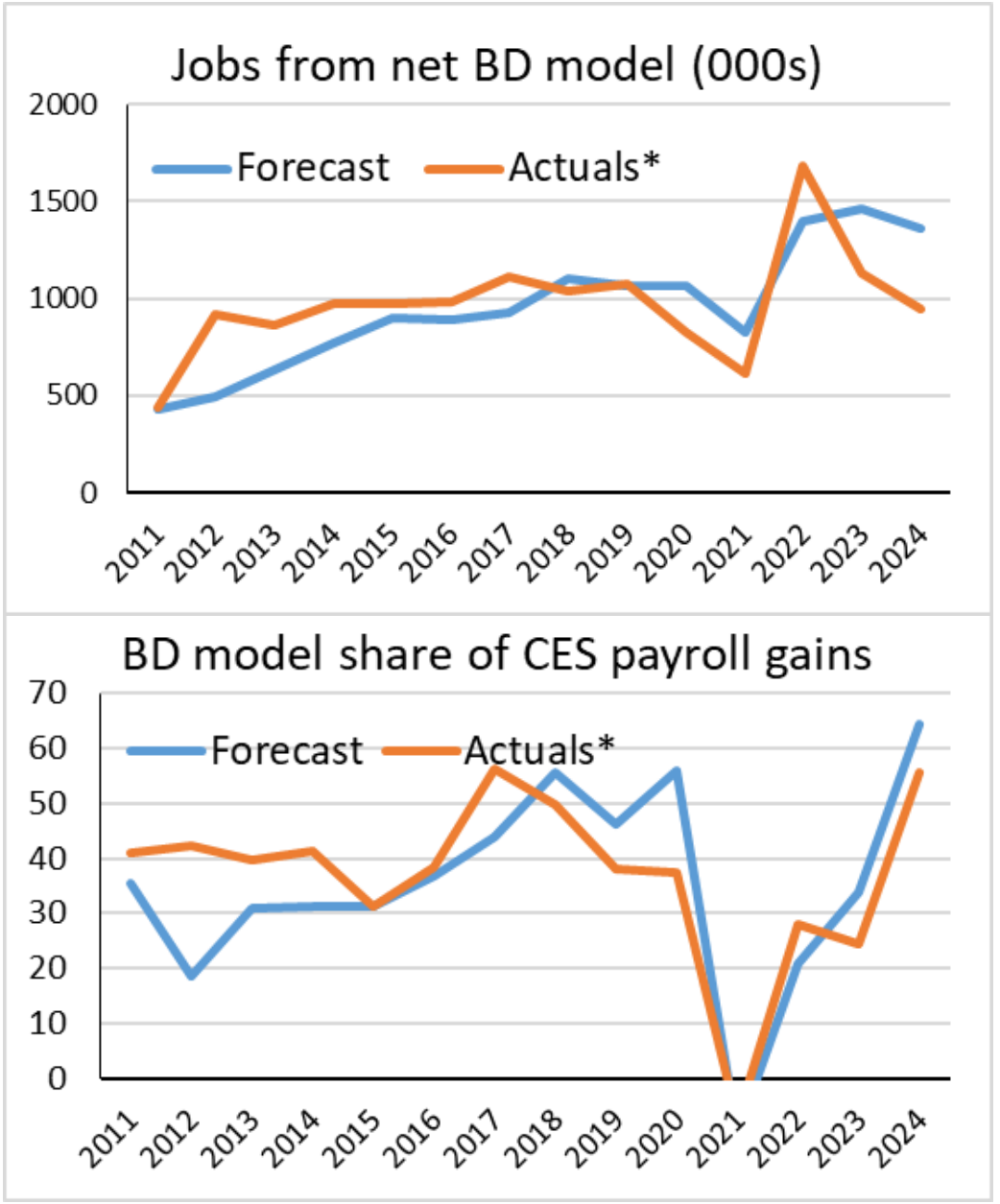
- Firm versus establishment distinction: What is the payroll unit? Matters for weighting, geography, etc.
- Paycheck/pay period timing (weekly, biweekly, monthly, quarterly); pay adjustments; bonuses and other irregular pay events.
- Seasonal adjustment: long enough time series? Adjust at granular level or aggregate?
- As with official business survey data, private payroll data require additional assumptions about **establishment birth/death** (next slide…)

Matters of birth and death

- Nontraditional data most useful when they can be made methodologically comparable to official data
- For payroll data, this requires both continuer-based measures and supplemental birth-death guesswork
  - …so any comparison leaves doubt about sample differences versus birth-death adjustment differences

Matters of birth and death

- Nontraditional data most useful when they can be made methodologically comparable to official data
- For payroll data, this requires both continuer-based measures and supplemental birth-death guesswork
  - …so any comparison leaves doubt about sample differences versus birth-death adjustment differences
- This is a challenge
  - ~1/3rd of CES reported jobs are supplied by the BD model (closer to 2/3 recently)
  - Additional birth-death components baked in via imputation step
- → There is no publicly available "continuers only" CES series for nontraditional data comparisons



* 2024 value is a simple BED-based forecast by me.

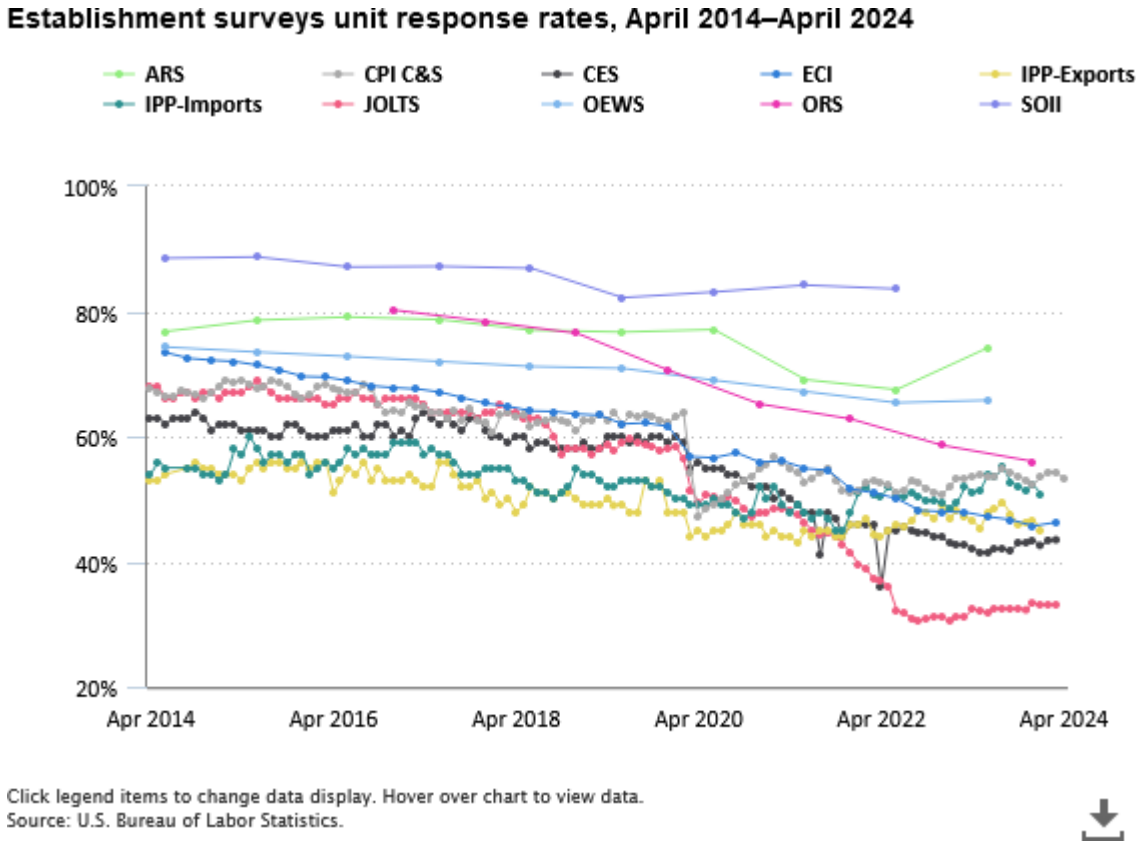
2. Stat agency usage of private payroll data?

2. Stat agency usage of private payroll data?

- This paper builds a private payroll product outside the stat agency firewall
  - Similar approach taken with other products (e.g., Cajner et al. 2018; Cajner et al. 2022 in 2022 NBER CRIW volume, "Big Data for 21st Century Economic Statistics.")
  - Other application for real-time geog detail: Natural disasters (e.g., Aladangady et al. 2016; Bayard et al. 2017)

2. Stat agency usage of private payroll data?

- This paper builds a private payroll product outside the stat agency firewall
  - Similar approach taken with other products (e.g., Cajner et al. 2018; Cajner et al. 2022 in 2022 NBER CRIW volume, "Big Data for 21st Century Economic Statistics.")
  - Other application for real-time geog detail: Natural disasters (e.g., Aladangady et al. 2016; Bayard et al. 2017)
- Alternative use: combine private payroll microdata with survey/admin microdata (see Abraham 2022)
  - Large, +biased QCEW revisions during pandemic (neg. correlated with response rates)*
  - Declining monthly survey response rates, high cost of increasing sample size (BLS 2015).
  - **Could private payroll companies contribute more to official labor market data collection?**
  - And… to what extent do private payroll providers already participate in survey and administrative data collection?



*Unpublished work by Seth Murray

Wrapping up

- Excellent paper; smart way to do something others are not doing; model-based approach side-steps some tricky issues like birth-death etc.
  - Particularly novel: Evaluate accuracy improvement relative to stat agency's own implicit standards
- Various challenges with private data – be careful folks!
- Nontraditional/private sector data most useful when users can compare to official data; hard to do with payroll data currently due to birth/death issues
- What are prospects for getting these data behind the firewall?

Extra slides and references

The challenge of birth and death

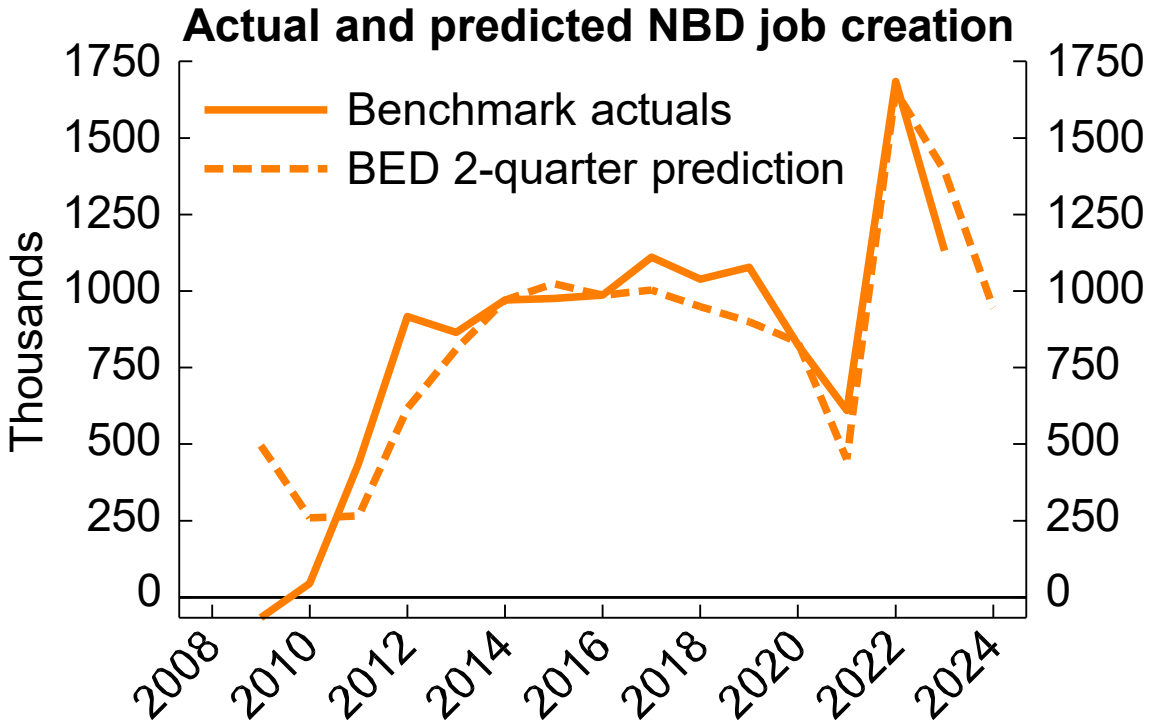
- Business birth and death present a challenge to measurement of economic activity from the business side
  - Can't instantly survey births, hard for deaths too
  - Private sector business registers typically can't distinguish sample entry/exit from true birth/death

This challenge is common in almost all business measurement settings!

- Ways to get around it
  - Find supplemental info – e.g., cell phone data (Crane et al. 2022), Facebook/Google data (Kurmann et al. 2024). May work for some industries, not all.
  - **More likely**: Just study continuers, perhaps with formal birth/death forecast (e.g., CES) or "bias adjustment" (e.g., old CES, many other business surveys, Cajner et al. 2018)

BED-based net birth-death forecast

- Obtain quarterly (NSA) net job gains from openings – closings from BED
- Obtain quarterly net birth-death model forecasts
- Regress 12-mo (to March) birth-death model actuals on cumulative BED and ratio of cumulative BED to cumulative NBD model forecasts, for first two quarters of the reference year (i.e., calendar Q2-Q3). Data for 2009-present.
- Preliminary result from in-sample (and non-real time) regressions has
  - Adj. R^2 = 73%, RMSE = 229k
  - 2024 prediction = 944k (versus NBD model forecast 1365k).
- Other specifications:
  - Model omitting BED/NBD ratio has adj R^2 = 24%, RMSE = 383k, 2024 prediction = 811k
  - Model omitting BED/NBD ratio but including NBD (in levels) has adj R^2 = 46%, RMSE = 325k, 2024 prediction = 1302.



Note: March reference month. Net birth-death model actuals, 12 months through March, from CES benchmark articles. Forecast using BED data and net birth-death model forecast using only the first two quarters of the benchmark year (i.e., Q2 and Q3).

References

- Abraham, Katharine. 2022. “Big data and official statistics.” Review of Income and Wealth 68 no. 4:835-1078.
- 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. At https://www.federalreserve.gov/econresdata/notes/feds-notes/2016/effect-of-hurricane-matthew-on-consumer-spending-20161202.html.
- Bayard, Kimberly, Ryan Decker, and Charles Gilbert. 2017. “Natural disasters and the measurement of Industrial Production: Hurricane Harvey, a case study.” FEDS Notes. At https://www.federalreserve.gov/econres/notes/feds-notes/natural-disasters-and-the-measurement-of-industrial-production-hurricane-harvey-a-case-study-20171011.html.
- Barnatchez, Keith, Leland Crane, and Ryan Decker. 2017. “An assessment of the National Establishment Time Series (NETS) database.” FEDS working paper #2017-110.
- Bureau of Labor Statistics. 2015. “A report to Congress on the Bureau of Labor Statistics’ Current Employment Statistics Methodology for Metropolitan Statistical Areas.” At https://www.bls.gov/sae/additional-resources/bls-report-to-congress-on-ces-methodology-for-metropolitan-statistical-areas-2015.pdf.
- Cajner, Tomaz, Leland Crane, Ryan Decker, Adrian Hamins-Puertolas, and Christopher Kurz. 2022. “Improving the accuracy of economic measurement with multiple data sources: The case of payroll employment data.” In Katharine G. Abraham, Ron S. Jarmin, Brian Moyer, and Matthew D. Shapiro, eds., Big Data for 21st Century Economic Statistics, NBER Book Series Studies in Income and Wealth, University of Chicago Press.
- Crane, Leland, Aaron Flaaen, Adrian Hamins-Puertolas, and Christopher Kurz. 2022. “Business exit during the COVID-19 pandemic: Non-traditional measures in historical context.” Journal of Macroeconomics 72 (June).
- Kurmann, Andre, Etienne Lale, and Lien Ta. 2024. “Measuring small business dynamics and employment with private-sector real-time data.” Working paper.
- Stevens, John. 2021. “Federal Reserve Board experience using transaction data.” Presentation for the CNSTAT Workshop on “Scope, Components and Key Characteristics of a 21st Century Data Infrastructure,” December 9, 2021. At https://www.nationalacademies.org/event/12-09-2021/the-scope-components-and-key-characteristics-of-a-21st-century-data-infrastructure-workshop-1a.