

Building Labor Market Statistics with Administrative Labor Market Data

Tomaz Cajner, Leland D. Crane, Ryan A. Decker, Adrian
Hamins-Puertolas, Christopher Kurz, Daniel Villar

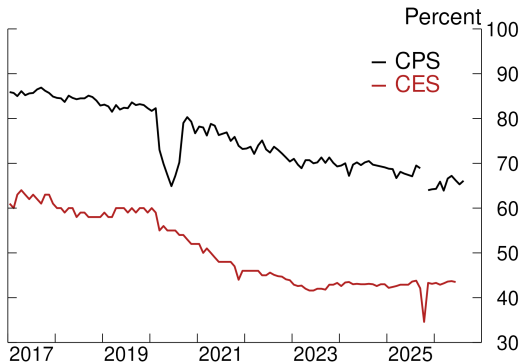
Federal Reserve Board

September 17, 2026

The analysis and conclusions set forth are those of the authors and do not indicate concurrence by other members of the research staff or the Board of Governors of the Federal Reserve System.

Labor market data: Challenges and opportunities

Figure: 3rd print response rates



- U.S. labor market data have been under stress
 - Declining survey response rates: Increased variance, potentially bias if response/nonresponse error is selected
 - Large benchmark revisions (1/2 percent in 2022, 2024, 2025) [History](#)
- Meanwhile: Increasing availability of private sector administrative data

What are the prospects for enhancing official labor statistics with private data?

Tradeoffs in official labor data

- Official employment data (CES, CPS, etc.) have key strengths...
 - Scientifically designed random probability samples optimized for representativeness
 - Large samples
 - Established, transparent, consistent methodologies
 - Long history for time series analysis
 - Updated with benchmarks, etc.; CES has a uniquely timely and high-quality benchmark source (QCEW)
- ...and limitations
 - Measurement error from sampling error, reporting error, nonresponse
 - Extremely difficult to capture business birth, death
 - *Monthly* population “truth” is never observed
 - Survey fatigue, stat agency resources, etc.

Tradeoffs in private sector data

- Private sector data (e.g., from payroll providers) may offer...
 - Large, complementary samples
 - Better timeliness, data frequency
 - Granularity (industry, geography)
- ...but are subject to
 - Selection on observables (fixable) and unobservables (not fixable)
 - Shorter time series history (seasonal adjustment? Business cycles?)
 - Population “truth” is never observed
 - Methodological consistency problems; holdup problems

Various such data sources including ADP, Paychex, UKG, Gusto, Homebase, Intuit, Revelio, etc.

Our decade with payroll processor data—producers and users

- **ADP**: Large payroll provider covering 1/5th of U.S. payroll employment with coverage broadly representative across industry, business size, and geography
- We had anonymized microdata access (about 2000 on), with new data received weekly (about 1 week delay), including:
 - “Active” employment: Number of employees on payroll database...
 - “Paid” employment: Number of paychecks issued (weekly, biweekly, monthly)...
 - Detailed NAICS industry, location...
 - At the payroll control level (for employment data) or paycheck level (for wages)
- We built various **payroll indexes** and **wage measures**

Upshot: These data helped us sense underlying labor market momentum across a variety of economic environments (pre-Covid, Covid, post-Covid), spot measurement issues in official data, and track high-frequency developments in near-real time.

Methodological challenges and hard choices

Start with raw data: ADP weekly file with paycheck history and both *active* and *paid* employment

- **How to deal with entry/exit?** Calculate weekly growth rates for *continuing* establishments; address entry/exit “bias” later
- **How to ensure industry/size representativeness?** Apply QCEW size-industry sampling weights
 - Cell delineation trades off detail against coverage; too narrow means sparse cells
 - Nothing can be done about representativeness re: unobservables; rely on benchmarking
 - Notable fraction of NAICS codes are missing; assume “missing at random” by size
- **Chain the growth rates into a weekly index**
- **How to seasonally adjust?** Adjust aggregate and sector-level series with Cleveland & Scott (2007). Weekly or monthly? Fixed or concurrent?

(More) Methodological challenges and hard choices

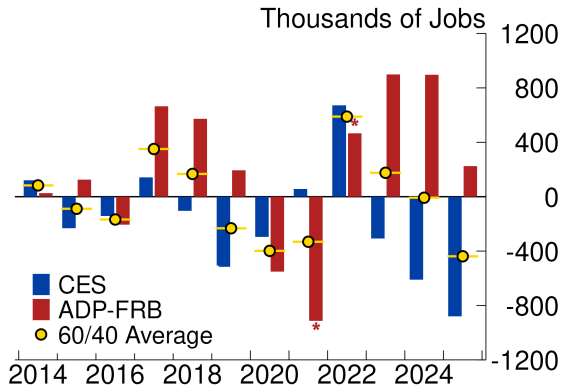
Now we have seasonally adjusted weekly payroll *indexes*—aggregate and sector level

- **How to translate into jobs?** Benchmark to *March* QCEW via CES
 - Big picture: **Benchmarking cleans up many messes**
- **What about birth/death?** CES uses net birth-death model
 - Other surveys (retail sales, quarterly services, pre-2000s CES) use benchmark-based bias adjustment
 - We use “forward benchmark:” project forward benchmark errors; some judgment involved
- **Active or paid?** We (usually) prefer *active* but some exceptions (more later)
 - Paid is closer to CES concept but can be (very) noisy
 - Active more closely tracks CES
- **Other critical housekeeping:**
 - Paycheck timing, frequency varies across businesses; snap to Saturday and adjust
 - Screen for outliers: Usually systematic, sometimes judgmental
 - Are payroll control units *firms* or *establishments*? Closer to establishments.

How to evaluate ADP-FRB payroll series?

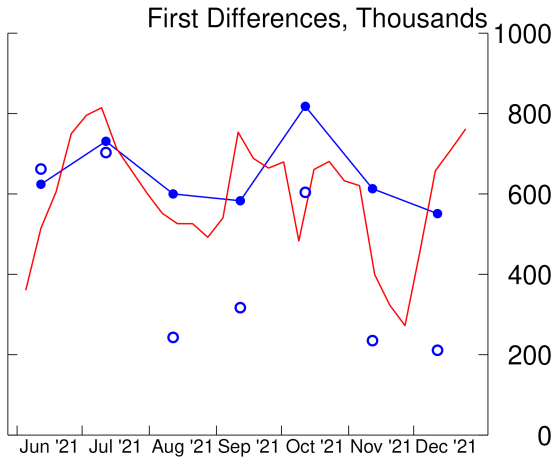
- In the CES universe “truth” is only observed once per year; monthly wiggles forever reflect initial sample/model estimates
- Therefore, an important test of alternative data is: Do they help predict annual benchmark employment levels?
- Other tests: Identifying periods of large CES mismeasurement; predicting monthly patterns; tracking rapid economic developments

ADP-FRB helps predict benchmark revisions



- ADP-FRB revisions *somewhat* larger than CES... but not dramatically so
- ADP-FRB revisions often in opposite direction from CES revisions—i.e., point the right direction
- 60/40 average is often a safe bet especially for recent large CES revisions

ADP-FRB in key labor market insights: Summer 2021



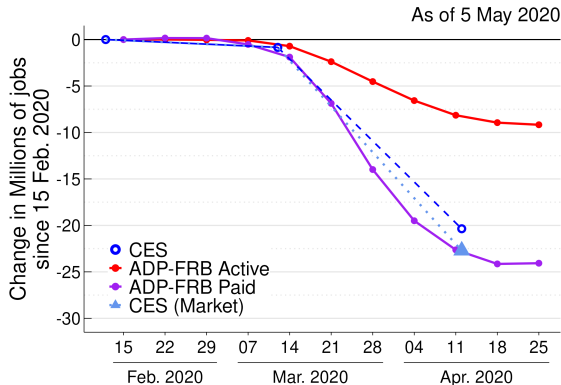
- CES (first vintage)
- CES (2023-02-03 vintage)
- ADP-FRB (2022-01-28 vintage)

- CES suggested rapid slowdown of labor market recovery in summer 2021...
- ...while ADP-FRB suggested continued strong job growth.
- CES revised up toward ADP-FRB.
- More broadly: ADP-FRB helpful even for monthly frequency and 1st-to-3rd-print CES tracking

Monthly frequency

2024-2025

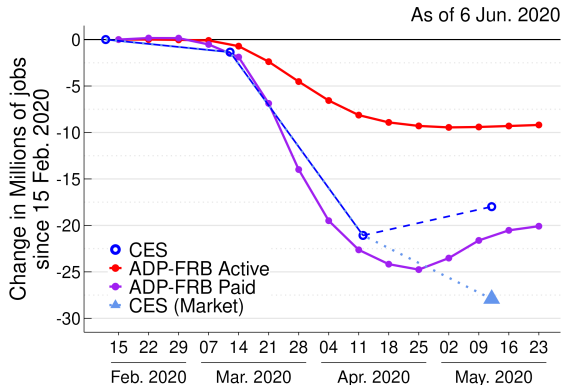
ADP-FRB in the early pandemic: The view from early May



- Pandemic: immediately obvious that *paid* was preferable to *active*
- By early May: the market forecasts saw the April payroll decline coming, as did ADP-FRB *paid*...
- ...which also suggested payrolls stabilized later in March.

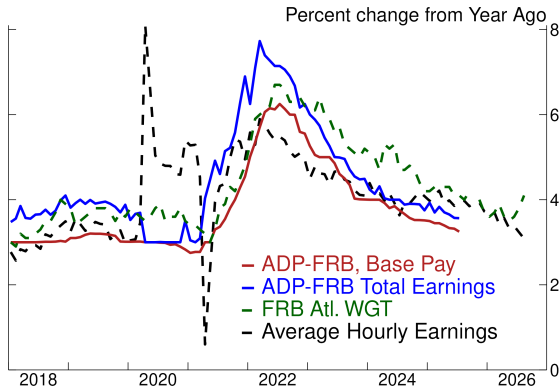
April

ADP-FRB in the early pandemic: The view from early June



- The market forecasts expected further decline in May, likely relying in part on (publicly available) alternative data sources...
- ...but CES payrolls stabilized, broadly consistent with ADP-FRB *paid*—which we observed several weeks in advance

ADP-FRB wage measures



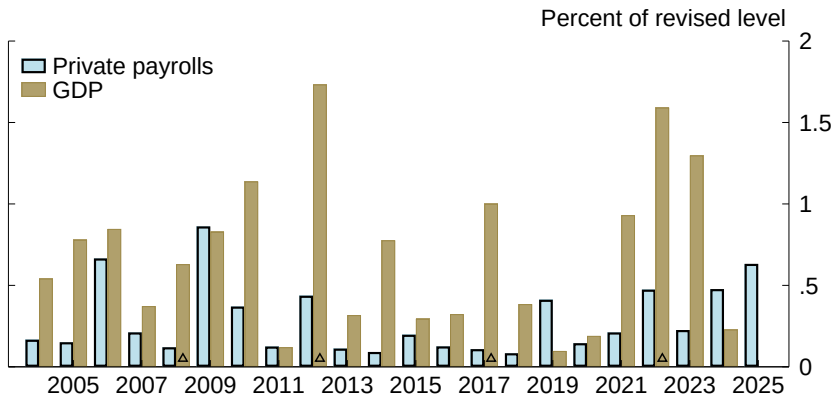
- ADP microdata permit Wage Growth Tracker (WGT)-style (median) measures that abstract from composition effects...
- ...which tracked official wage growth in the post-pandemic period.

Job switchers

Lessons learned

- **Private payroll processor data are useful** for measuring both high- and low-frequency employment changes—and wages—as a *complement* to official data
- **Producing statistics from the raw data is difficult**, requiring careful methodological design and judgments; stat agency staff are the most qualified in the world to do that
- **Usefulness of private payroll data depends on high-quality benchmarks.** The data will lead you far astray without them; and benchmarks buy data users some flexibility
- **The data are most useful in microdata form or close equivalent**, given difficult methodological considerations
- **Technology may enable more flexible data-sharing arrangements**, e.g., sending “agents” behind company firewalls to retrieve consistently constructed cell-level aggregates

Annual revisions, CES and GDP

[◀ Back](#)

Note: Absolute revision to level in March (for private payrolls) or full year (for GDP) as published in subsequent year. Δ denotes GDP comprehensive revision years.
Source: BLS Current Employment Statistics; BEA National Income and Product Accounts.
BEA data retrieved from ALFRED, Federal Reserve Bank of St. Louis.

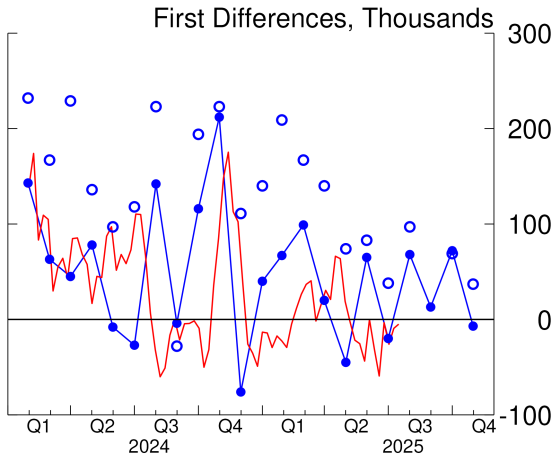
	(1)
ADP-FRB, centered	1.329***
	(0.278)
Observations	56
R^2	0.618

Standard errors in parentheses

* $p < .10$, ** $p < .05$, *** $p < .01$

Dependent variable is current vintage CES jobs gains (seasonally adjusted), centered by removing the April-March average from each interval. ADP-FRB is processed symmetrically. Robust standard errors in parentheses. Sample is 2021m1-2025m8. Source: BLS, ADP NER, authors' calculations

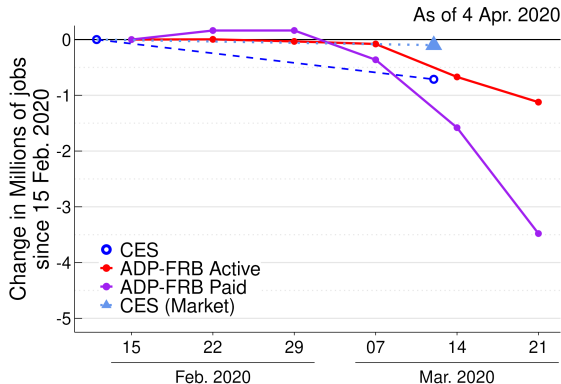
ADP-FRB in key labor market insights: 2024-2025

[◀ Back](#)

- CES suggested strong payrolls through 2024 into 2025...
- ...while ADP-FRB pointed weaker.
- CES revised down toward ADP-FRB.

○ CES (first vintage)
● CES (current vintage)
— ADP-FRB (2025-09-19 vintage)

ADP-FRB in the early pandemic: Early April

[◀ Back](#)

- As of the March jobs report, ADP-FRB *active* had signaled some job loss...
- ...while ADP-FRB *paid* showed steep collapse, imminent after March reference period...
- ...which gave us a unique view relative to market forecasts.

ADP-FRB wage measures: Switchers vs. stayers [◀ Back](#)

Figure: Wage growth for stayers and switchers

