

Blending Privately-Provided Payroll Data and Government Statistics

FCSM Blended Data Session – May 1, 2020

Tomaz Cajner – Leland Crane – Ryan Decker – Adrian Hamins-Puertolas – Chris Kurz

Federal Reserve Board of Governors

Opinions expressed herein are those of the authors alone and do not necessarily reflect the views of the Federal Reserve System. All results have been reviewed to ensure that no confidential data are disclosed.

1. Our data, processing, weighting, state-space model
2. Briefly: New weekly employment growth estimates for recent history. Suddenly more relevant

Introduction

How can one combine official and alternative data to improve the accuracy of economic measurement?

Our approach, for payroll employment:

- Build a payroll employment index using alternative, private-source microdata (ADP)
- Use official data for weighting, benchmarking
- Forecast official data, for validation
- State-space model: combine the information in alternative and official series (treat both as noisy indicators of true employment)

Government statistics are critical in our application: not about replacing them.

Alternative Data

Our data based on records from the payroll processor ADP (July 1999-present):

- ADP processes paychecks for 20 percent of private US workers
- Every pay period: client firm sends ADP data on the number of workers to pay
- We build estimates of aggregate monthly private employment, using methods similar to the official data

Strengths and weakness of our data:

- We observe all pay periods, not a once-a-month snapshots
- Not a probability sample, but approximately representative [More](#)
- Real-time measurement: we get data updates weekly, no lag in reporting

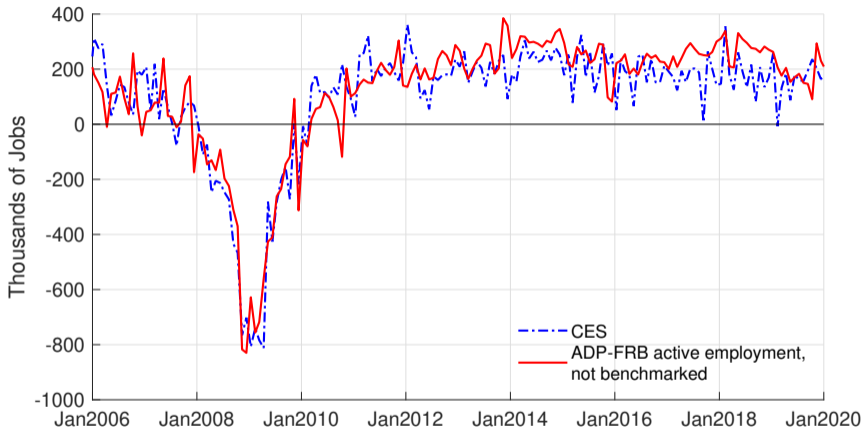
CES Survey

- Monthly employer survey
- Estimates available 1st week following the reference month
- Samples 20 percent (!) of employment
- **Significant** measurement error in real time

QCEW

- Census of employment
- Based on UI records
- Available with 6-9 month lag

CES and ADP-FRB Payroll Employment Gains (in thousands)



Plotted through **February**. ADP-FRB tracks the business cycle very well (even without benchmarking).

Uses of Official Data in Our Process

- Weights: make ADP more representative, by size and industry
- Benchmark ADP-FRB to QCEW when it arrives.
 - Makes sure our historical series are accurate.
 - Learn if there are persistent errors.
 - Similar to what the CES program does.
- Test if ADP-FRB can predict the monthly CES data
 - ADP-FRB data are statistically significant predictors, even after controlling for market expectations Regression Results
 - Improvement in RMSE is modest

What is the Best Use of ADP-FRB Data?

Do we really want to forecast CES?

CES has measurement error from:

- sampling error
- birth-death adjustment
- nonresponse
- reference period concept

Instead: Combine CES and ADP-FRB data within a state-space framework. Treat both as noisy signals of truth

State Space Model

Kalman filter assumptions:

- Unobserved state (true employment growth) follows an AR(1):

$$\Delta EMP_t^U = \mu(1 - \rho) + \rho \Delta EMP_{t-1}^U + \epsilon_t^U$$

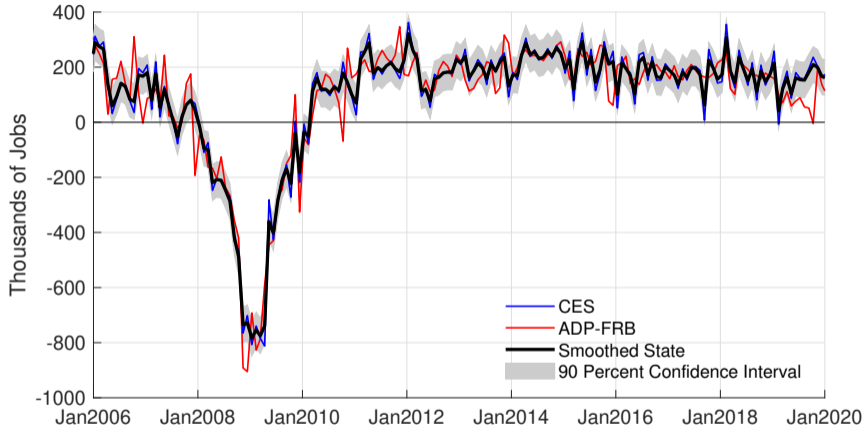
- CES and ADP-FRB are noisy signals of truth:

$$\begin{bmatrix} \Delta EMP_t^{CES} \\ \Delta EMP_t^{ADP-FRB} \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \Delta EMP_t^U + \begin{bmatrix} \epsilon_t^{CES} \\ \epsilon_t^{ADP-FRB} \end{bmatrix}$$

Feed in the data:

- Extract estimates of observation noise
- Extract estimates of true employment growth

State Space Estimate of Payroll Employment Gains (in thousands)



State estimate combines the noisy information in both series.

Interpretation of State Space Results

State-space estimates:

- Model puts roughly equal weight on CES, ADP-FRB (Kalman gains are similar)
- Robust to changes in assumptions: AR(1), random walk, correlation of observation noise
- Adding CPS series, adjusted to match CES scope, yields similar results (very low weight on CPS data)

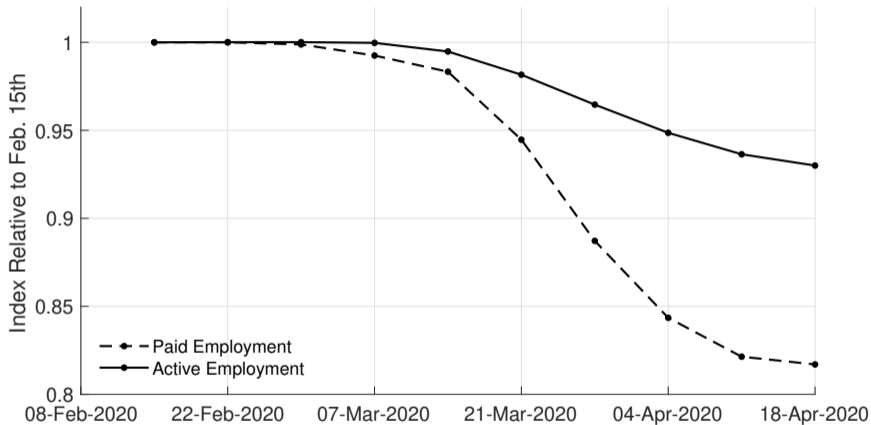
In a forecasting regression, the state-space estimate outperforms CES and ADP-FRB data (and the state-space estimate based on CES data only).

Regression Results

High Frequency Data: Weekly Impact of COVID-19

- The labor market has been changing *very fast*
- Our data are fundamentally weekly
- Will show our weekly employment indexes. Weighted to QCEW, seasonally adjusted.
- “Paid employment”: Number of workers who actually got paid
- “Active employment”: Number of workers the firm keeps in their payroll system. Includes those on vacation, temporary layoff

High Frequency Data: Weekly Impact of COVID-19



- Roughly 20% decline in paid employment. Active employment falls less
- Declines seem to start *before* lockdown orders, level off in recent weeks

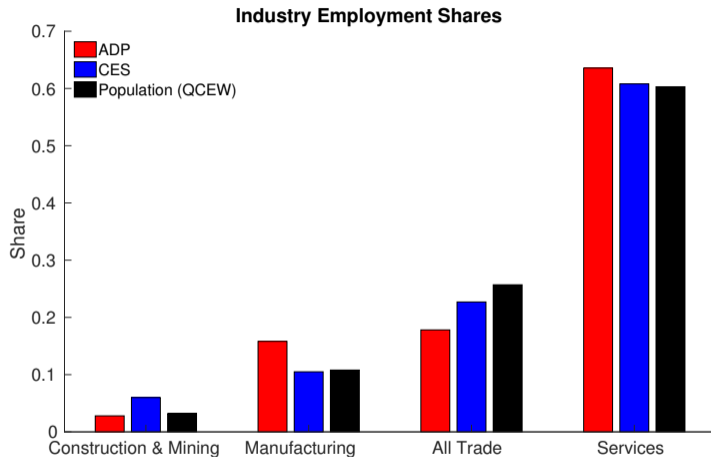
Conclusions

- Alternative data can improve the accuracy of payroll employment estimates
- Use official data for weighting, benchmarking, validation, signal combination
- Statistical agencies could potentially use data from payroll processors to increase their sample size
- More work on high frequency data coming next week. Should be posted at the Becker-Friedman Institute (<https://bfi.uchicago.edu/covid-19/>)

Source papers for this presentation:

- “Improving the Accuracy of Economic Measurement with Multiple Data Sources: The Case of Payroll Employment Data”
- “Tracking Labor Market Developments during the COVID-19 Pandemic: A Preliminary Assessment”

Representativeness



Representativeness

Pay frequency	ADP emp.	ADP estabs.	QCEW estabs.
Weekly	23.4	22.4	32.2
Biweekly	55.1	45.8	40.0
Semimonthly	17.5	20.6	18.5
Monthly	4.0	11.2	9.3

[Back](#)

Representativeness

Census Region	ADP emp.	ADP estabs.	QCEW emp.
Northeast	28.2	28.1	18.2
South	29.4	30.2	34.9
Midwest	20.2	16.6	20.1
West	22.2	25.2	26.8

[Back](#)

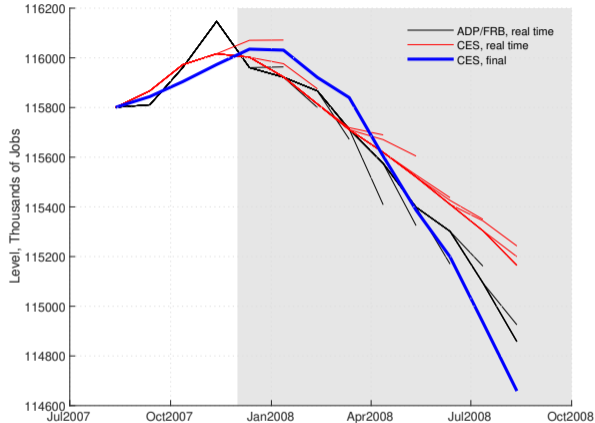
Benchmark Revisions

	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
ADP-FRB	-173	-451	12	709	283	-230	-1030	-853	-322	-623
CES	-137	-933	-391	229	481	340	105	-259	-151	136
CES No BD	645	-216	-55	561	972	975	874	638	737	1066

Notes: Units: Thousands of jobs. CES revisions are the post-benchmark (QCEW-based) March estimate less the pre-benchmark estimate. ADP-FRB revisions are calculated in a similar fashion. CES no BD are the CES benchmark revisions that would have occurred excluding net birth-death adjustment.

Source: <https://www.bls.gov/web/emp/sit/cesbmart.pdf>, authors' calculations.

Real-Time vs. Current Vintage Estimates



During Great Recession ADP-FRB data outperformed CES data.

[Back](#)

Forecasting Annual Employment Changes

	(1)	(2)	(3)	(4)	(5)
CES	1.126*** (0.0316)			1.104*** (0.142)	
CES excl. Birth-Death		1.154*** (0.0235)			0.927*** (0.0847)
ADP			0.976*** (0.0543)	0.0197 (0.121)	0.199** (0.0818)
Constant	-163.7* (76.93)	604.5*** (75.29)	-135.1 (172.8)	-163.6* (82.61)	452.5*** (79.37)
Observations	10	10	10	10	10
Adj. R-squared	0.989	0.993	0.965	0.988	0.994
RMSE	299.2	243.3	535.9	319.7	224.2

Notes: Dependent variable is benchmarked annual change in private nonfarm employment, March to March. Years 2008-2017. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively. Robust standard errors in parentheses.

Forecasting Monthly Employment Changes

[Back](#)

	(1)	(2)	(3)	(4)	(5)
ADP-FRB active employment			0.29** (0.11)	0.39*** (0.11)	0.16** (0.07)
Lagged private CES employment	0.82*** (0.07)	-0.13 (0.15)	-0.21 (0.14)	0.51*** (0.12)	
Lagged UR change	-156.73** (61.56)	-45.66 (52.17)	-43.05 (46.84)	-123.09** (58.02)	
Unemployment expectations	39.17*** (11.82)	30.95*** (11.01)	14.08 (12.29)	16.55 (12.74)	15.21 (10.88)
Initial UI claims	-3.10*** (0.74)	-0.91 (0.71)	-0.79 (0.72)	-2.52*** (0.83)	-0.56 (0.52)
CES employment expectations		1.15*** (0.16)	0.98*** (0.15)		
Private CES employment					0.97*** (0.07)
UR change					33.12 (36.03)
Constant	4.87 (9.36)	-17.77* (10.40)	-24.39** (11.58)	-7.48 (10.77)	-17.85** (8.98)
RMSE	99	84	80	92	58

Notes: Dependent variable is final print of CES private employment. ADP series are real-time vintage, as of 5 weeks after the start of the month (i.e., the week before or week of the Employment Situation release). Unemployment expectations are from the Michigan survey. CES employment expectations are eve-of-release median markets expectations. Lagged private CES employment refers to pre-Employment Situation release. Robust standard errors in parentheses. RSMEs are calculated in-sample. * $p < 0.10$, **

Forecasting Monthly Employment Changes with State Estimate

	(1) CES employment	(2) CES employment	(3) 3-month average CES employment
Constant	-28.14 (19.43)	-28.52 (18.78)	-17.05 (20.35)
ADP-CES State	1.43*** (0.49)	1.50*** (0.55)	1.69*** (0.44)
ADP-FRB Emp.	-0.18 (0.15)	-0.19 (0.16)	-0.30** (0.15)
CES Emp.	-0.18 (0.34)	-0.11 (0.55)	-0.41 (0.31)
CES State		-0.12 (0.68)	-0.04 (0.42)

Notes: The dependent variable in columns 1 and 2 is the fully revised change in CES private employment at time $t + 1$; in column 3 the dependent variable is the average of the fully revised change in CES private employment for $t + 1$, $t + 2$ and $t + 3$. ADP series are real-time vintage, as of 5 weeks after the start of the month. CES series appearing as independent variable or in state-space estimates are real-time vintage. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, ***