

Examining the Utility of Educational Administrative Records for Research and Improving Survey Operations: A Pilot Project at the U.S. Census Bureau

2018 Federal Committee on Statistical Methodology
Research and Policy Conference
March 8, 2018

Nikolas Pharris-Ciurej, PhD
Center for Administrative Records Research and Applications (CARRA)
U.S. Census Bureau

Disclaimer: The views expressed are those of the presenter and not necessarily those of the U.S. Census Bureau.

Pilot Project Background

- CARRA developed data sharing agreements to obtain administrative school records
 - Two California school districts
 - State of Oregon
 - North Carolina (expected)
- Able to link school records to other U.S. Census Bureau administrative records and survey data

Potential Benefits

1. Answer research questions using novel linked data
2. Assess accuracy of educational attainment and school based program participation items on surveys
3. Population estimates for hard to count youth populations
4. Assess coverage of school-aged children in surveys and decennial censuses
5. Impute or replace educational attainment and school based program participation items on surveys
6. Serve as a frame for sample surveys

Data Usage Case Studies

- Multiple projects at the U.S. Census Bureau are using school records from this pilot project
- Highlight data usage with case studies
 1. Measuring National School Lunch Program (NSLP) free or reduced-price lunch eligibility
 2. Examining the career paths and earnings of former teachers

Case Study 1: Measuring NSLP Eligibility

- NSLP is one of the oldest and largest student development programs in the United States
 - Founded in 1946
 - Presently provides free or reduced-priced meals to over 30 million students
- Eligibility requirements
 - Free lunch: Household income less than 1.3 times the poverty level
 - Reduced-price lunch: Household income 1.3 to 1.85 times the poverty level

Utilizing Free or Reduced-Price Lunch Data

- Free or reduced-price lunch (FRL) enrollment is widely used as a proxy for student poverty
- Education research
 - Numerous studies have employed FRL enrollment
 - Particularly common in research using school records
- Educational finance and reporting requirements
 - *No Child Left Behind, Every Student Succeeds Act*, state accountability
 - Title I, and weighted per student funding formulae

Adequacy of FRL Data to Capture Poverty and Socioeconomic Status

- Obscures variation at the top and bottom of the income distribution
- FRL measures may underestimate the association between family background and education outcomes
- USDA Community Eligibility Provision may exacerbate measurement problems
- Prior research has not been able to assess the measurement properties of FRL enrollment

Research Questions

- What is the relationship between IRS-reported annual household income and free or reduced-price lunch program enrollment?
- What does school-reported free or reduced-price lunch program enrollment tell us about the relationship between family socioeconomic background and student test scores?

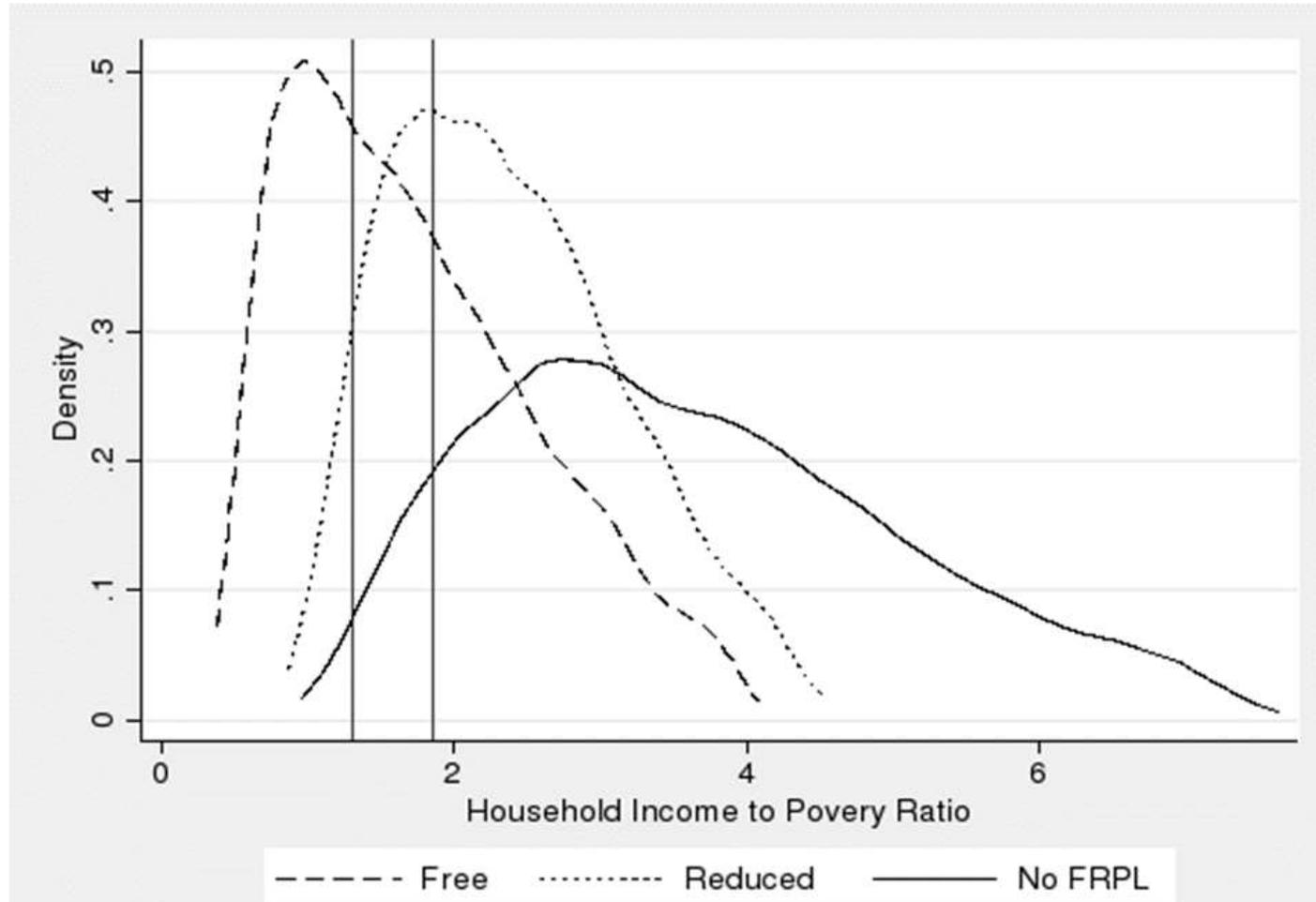
Data Sources

- Student-level administrative data from a California school district
 - 8th graders from 2008-09 through 2013-14
- Internal Revenue Service (IRS) 1040 tax data from 2000 through 2014
- Data linkage at the U.S. Census Bureau
 - Unique Protected Identification Keys (PIK) were placed on records, using personally identifying information
 - PIK allows records linkage
 - PIK rate: 94% for school records and ~99% for IRS 1040 forms

Data: Measures

- Student-level administrative data
 - Roughly 14,000 8th graders
 - Indicators for enrollment in free or reduced-price lunch
 - Demographics (race/ethnicity, gender, language status)
 - Annual end-of-grade achievement test scores (standardized)
- IRS 1040 tax data
 - IRS-reported household income: Sum of total money income reported on all 1040s in the household for the calendar year in which student began focal grade
- Indicators based upon IRS-reported household income:
 - Indicators for free or reduced-price lunch eligibility
 - Household income to poverty ratio (8th grade calendar year)
 - From K to 8th grade percentage of years household income is 1) below poverty threshold and 2) reduced-price lunch threshold

Distribution of IRS-reported Household Income-to-Poverty Ratio in 8th Grade by NSLP Enrollment in 8th Grade



Note: Sample sizes have been rounded and distributions have been modified to comply with U.S. Census Bureau disclosure requirements.

Source: Linked California school district records, for academic years 2008-09 – 2013-14, and Internal Revenue Service 1040 Tax data, from 2008 – 2013 (N ≈ 13,000).

OLS Regression on Standardized 8th Grade English Language Arts Scores with California School District and IRS Tax Record Predictors

	Model 1 Coef.	Model 2 Coef.
NSLP Free Lunch	-.392 ***	
NSLP Reduced Price Lunch	-.200 ***	
IRS Free Lunch		-.153 ***
IRS Reduced Price Lunch		-.160 ***
HH Income to Poverty Ratio		
HH Inc:Poverty Ratio Squared		
Demographic Controls		
% Years in Poverty		
% Years NSLP Eligible		
School Fixed Effect	+	+
Academic Year Fixed Effect	+	+
Constant	.282 ***	.122 ***
R-Squared	0.099	0.078

Demographic Controls include race/ethnicity and nativity, gender, and English language status.

Source: Linked California school district records, for academic years 2008-09 – 2013-14, and Internal Revenue Service 1040 Tax data, from 2000 – 2013.

Note: Sample sizes have been rounded to comply with U.S. Census Bureau disclosure requirements (N ≈ 14,000).

Data Limitations

- NSLP FRL eligibility may not match IRS predicted FRL eligibility
- Calendar year covered by IRS data captures only a portion of the August to May school year
- IRS income measure is not sensitive to month-to-month income volatility

NSLP FRL Findings

- Analyses with Oregon data yield similar results
- NSLP data are at best imperfect proxies for students' household income in a given year
- NSLP free or reduced-price lunch enrollment information is valuable
 - It appears to capture additional aspects of disadvantage not captured by IRS income measures

Case Study 2: Career Paths and Earnings of Former Teachers

- Teachers have a pronounced impact on student achievement and long run economic success
- Because of this impact, teacher turnover may be harmful for student achievement
- Large literature on the predictors of teacher attrition and mobility
- Little evidence on what happens to teachers after they leave teaching

Research Questions

1. Are most teachers that leave the school district ('leavers') still employed?
2. Do leavers move somewhere else?
3. Do leavers earn more elsewhere?

Data

- Administrative data from a California school district
 - Large urban school district
 - In a given year, roughly 5 to 7% of teachers leave
- IRS data
 - W-2 records
 - Data on businesses from the Business Register
- Linked using Protected Identification Keys (PIKs)
 - PIKs are assigned using personal identifying information
 - Teacher-level PIK rate: ~99.7%

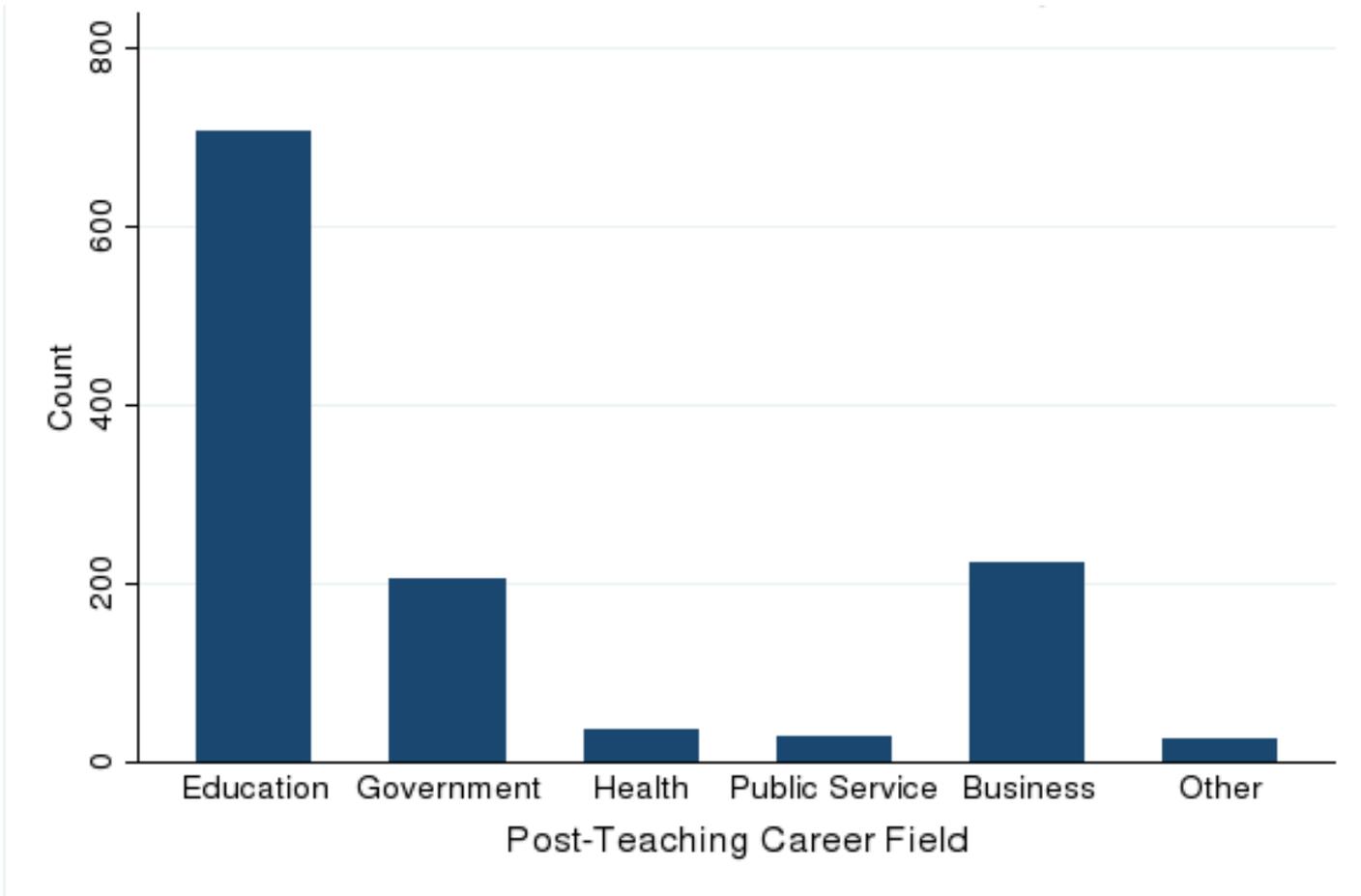
Employment of Teachers after Leaving

District exit code	N	Percentage Employed:		
		1 Year after Teaching	3 Years after Teaching	5 Years after Teaching
All leavers	≈ 3,000	49.7	47.6	48.4
Retirements	≈ 700	6.7	5.1	4.2
Involuntary leavers	≈ 600	61.9	55.6	55.7
Voluntary leavers, no retirements	≈ 1,700	64.1	61.5	61.7

Source: Linked school district administrative records and W-2 income data.

Note: The sample sizes have been rounded to comply with U.S. Census Bureau Disclosure Requirements.

Teacher Industry 1 Year after Leaving School District, Voluntary Leavers

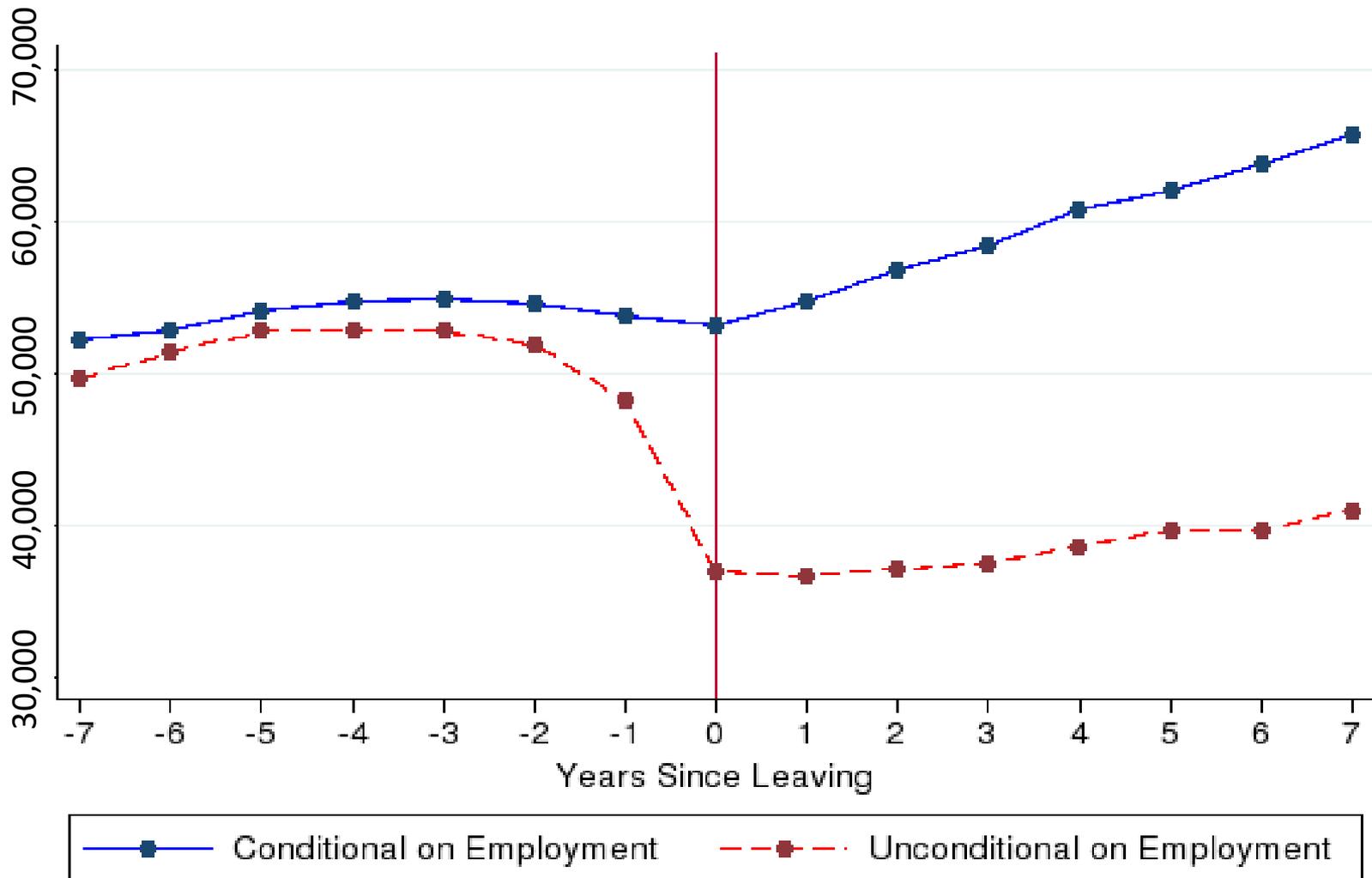


Source: Linked school district administrative records, W-2 income data, and industry from Business Register (N≈ 1,000).

Geographic Location of Former Teachers

- Among individuals who continue teaching, only about a quarter leave the state
- Moreover, individuals who continue teaching are also most likely to continue working in the same metropolitan area

Earnings for Voluntary Leavers



Source: Linked school district administrative records and W-2 income data (N ≈ 15,000).

Note: Sample sizes have been rounded to comply with U.S. Census Bureau Disclosure Requirements.

Teacher Mobility Findings

- Many teachers leave the workforce after they leave teaching
- If working, the majority of former teachers continue to work in a nearby school district
- Conditional on working, earnings increase for voluntary leavers
 - This pattern holds across all industries but is strongest for individuals who likely are leaving teaching for business occupations

School Records Project Next Steps

- Expand usage of school records to improve Census Bureau operations and answer important research questions
- Explore ways to efficiently expand our school records catalogue
 - Explore methods to harmonize records across localities and time

Project Team

Joint effort

- CARRA, Census Bureau
 - Quentin Brummet, PhD, Nikolas Pharris-Ciurej, PhD and Sonya Porter, PhD
- University of California, Irvine
 - Andrew Penner, PhD and Emily Penner, PhD
- University of North Carolina, Chapel Hill
 - Thurston Domina, PhD

Thank you!

Nikolas Pharris-Ciurej

Nikolas.D.Pharris-Ciurej@census.gov

Description of School Records

California School District #1

- Mid-sized urban school district
- Universe: 2008-09 to 2014-15 8th graders
 - Records for roughly 26,000 students
 - Student records for grades prior to 8th were obtained from 2008-09 to 2013-14 for 8th grade cohorts
 - Placed a PIK on 94% of student records

Description of School Records

California School District #2

- Large urban school district
- Universe: District applicants and employees from 2003 through 2015
 - Roughly 1.3 million records
 - Placed a PIK on 99.8% of records
- Also have school and classroom level data

Description of School Records

State of Oregon

- Universe: Student and Staff data from 2004 to 2015

Student Data

- Records for roughly 1.25 million students, including select pre-Kindergarten and post-secondary enrollment records
- Placed a PIK on 93% of student records

Staff Data

- Records for roughly 50,000 staff (teachers)
- Placed a PIK on 97% of staff records

Description of School Records

- Common student record data elements include:
 - Demographic information
 - School attended and grade level
 - Course enrollment
 - GPA and credits earned
 - End of year standardized test scores
 - Enrollment in free or reduced-price lunch program
 - Special education placement
 - English language learner status

Description of School Records

- Common staff data elements include:
 - Demographic information
 - Credentials, education
 - Experience
 - Salary
 - Position(s) held in school district
 - School assigned
 - Courses taught
 - Course roster information
 - Average class standardized test scores