Fixing Errors in a SNAP: Addressing SNAP Under-reporting to Evaluate Poverty

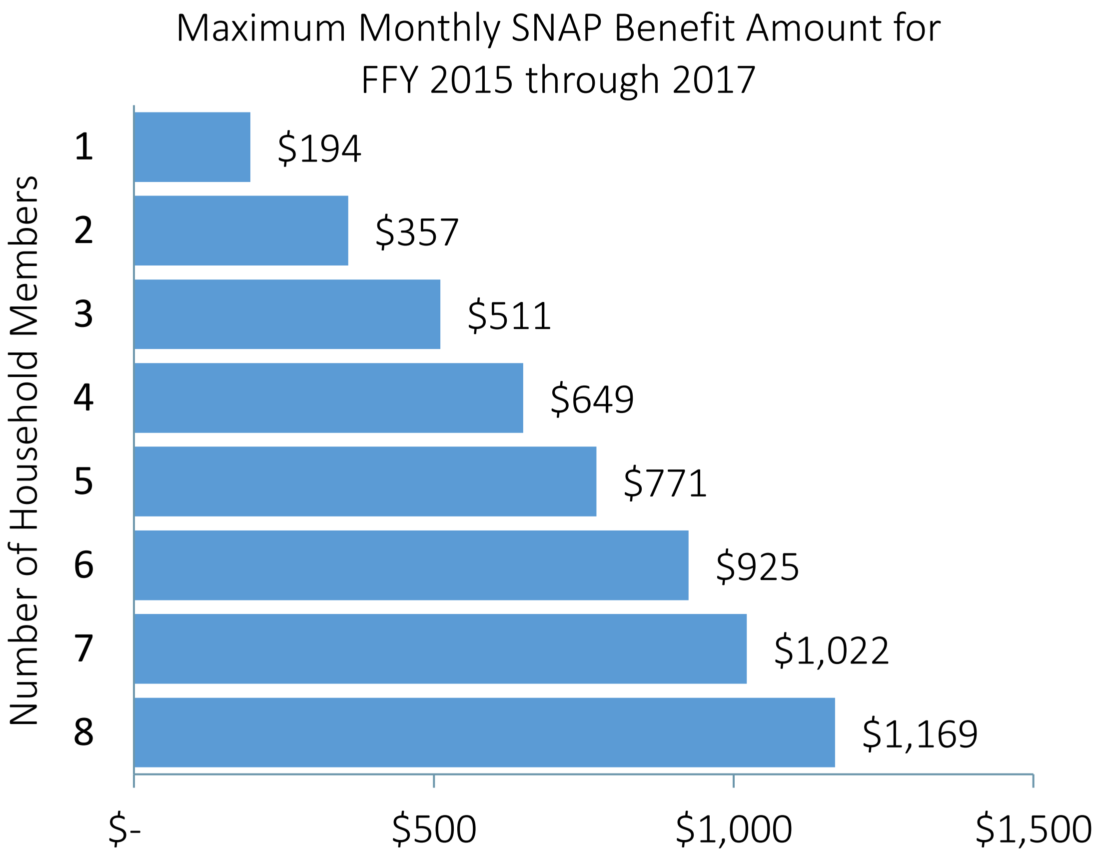
Jonathan Rothbaum, U.S. Census Bureau
Liana E. Fox, U.S. Census Bureau
Kathryn Shantz, Urban Institute

BLS Consumption Symposium

United States Census Bureau
U.S. Department of Commerce
Economics and Statistics Administration
U.S. CENSUS BUREAU
census.gov

Supplemental Nutrition Assistance Program (SNAP)

- In-kind benefit
- Eligibility requirements
  - Gross income test: 130% of FPG
  - Net income test: 100% of FPG
  - Asset limits
  - Work requirements
- Benefit amount calculation
- Average of 47.6 million recipients each month in 2013
- 79.9 Billion dollars in 2013

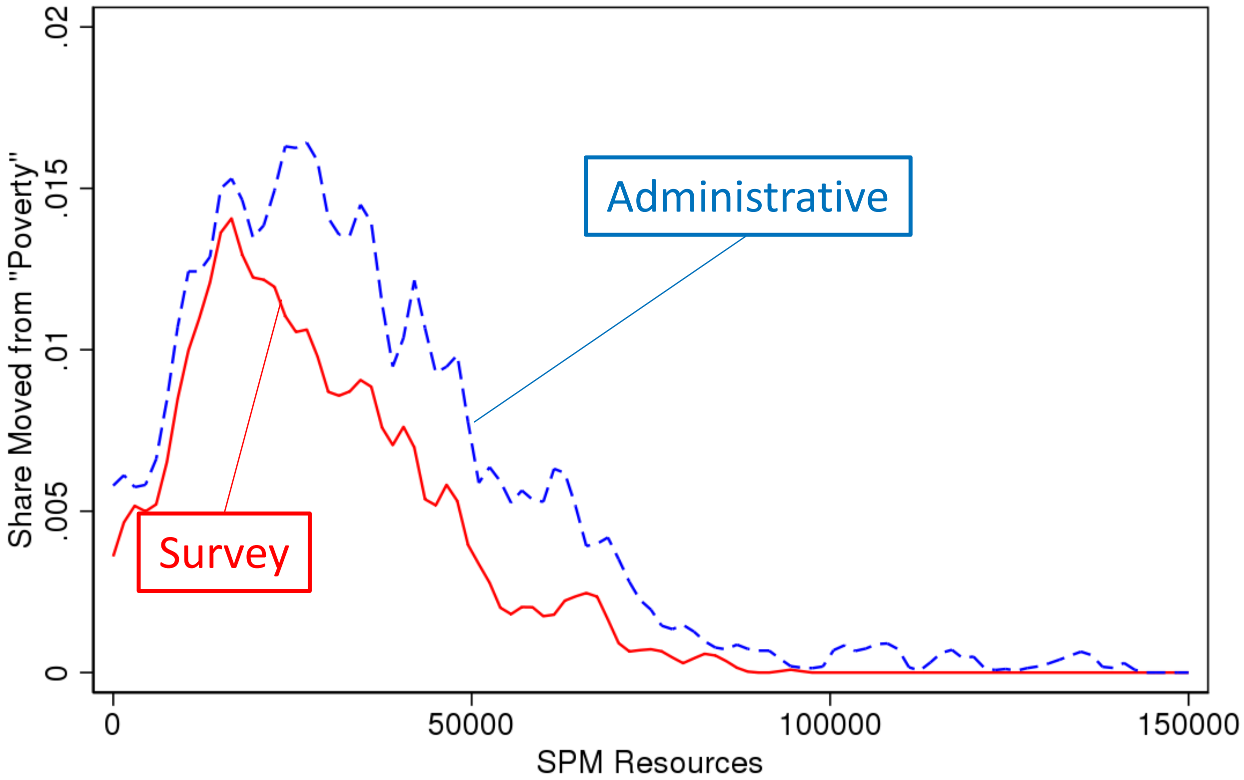


Source: United States Department of Agriculture Food and Nutrition Services. 2017. “Supplemental Nutrition Assistance Program (SNAP) Cost of Living Adjustment (COLA) Information.” Retrieved September 19, 2017 (https://www.fns.usda.gov/snap/cost-living-adjustment-cola-information).

Under-reporting Receipt

- A few examples
  - About 40% of SNAP recipients in NY did not report receipt in the CPS (Meyer and Mittag 2015)
  - About 46% of SNAP recipients in AZ, ID, IL, MD, OR, TN, and VA do not report receipt in the CPS ASEC (Stevens et al. 2018)
  - About 16% of SNAP recipients in IL, MD, and VA did not report receipt in the Survey of Income and Program Participation (SIPP) (Colby et al. 2017)

Under-reporting and Who's Affected by SNAP Benefits

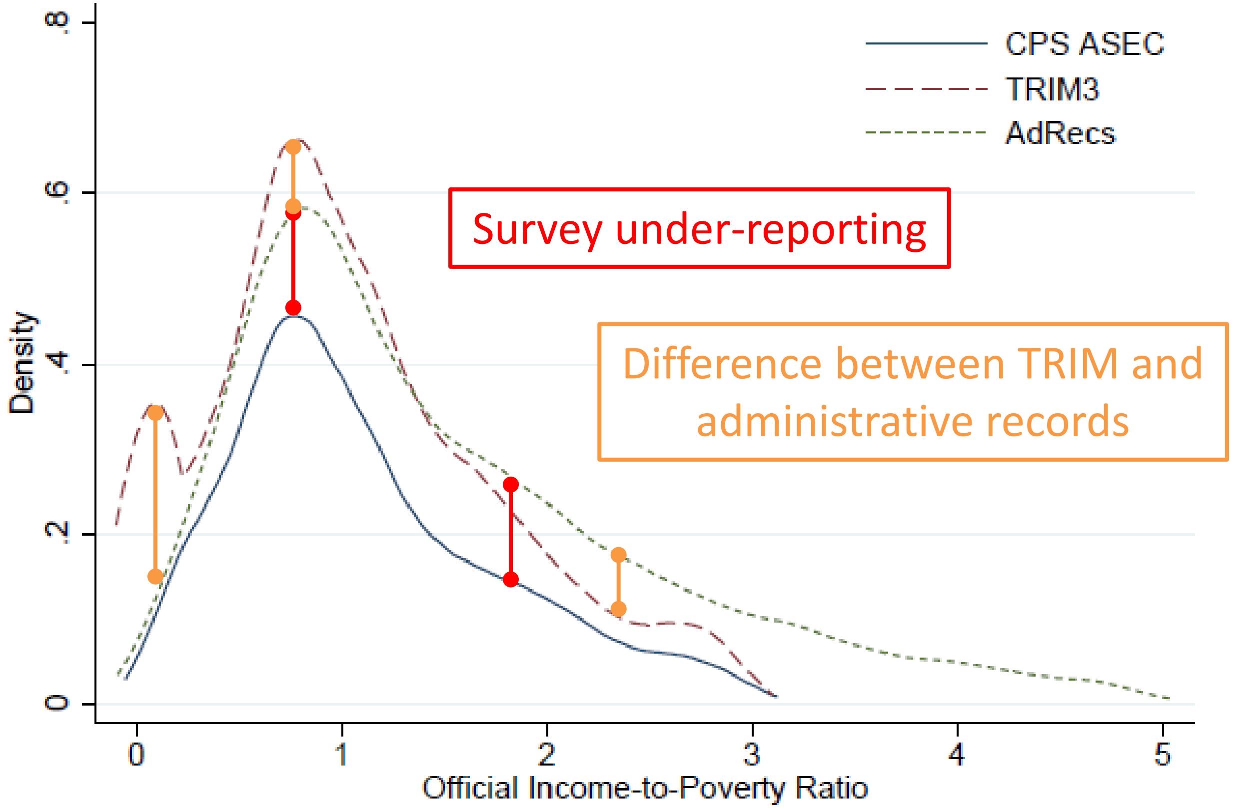


Source: U.S. Census Bureau, Current Population Survey, 2014 Annual Social and Economic Supplements (CPS ASEC) and state Supplemental Nutrition Assistance Program (SNAP) administrative records.

Solutions to Under-reporting

- Direct Replacement
  - Most accurate
  - Limited data availability
- Adjustments Based on External (Unlinked) Data
  - Examples
    - HHS/Urban Institute's TRIM3 microsimulation model – use program rules, aggregate data, quality control data (unlinked microdata on income and program participation) to allocate missing benefits
    - CBO's regression-based adjustments – use predicted receipt from regressions using survey responses to allocate missing benefits
    - Census imputation of the value of school lunch during the pandemic – allocates amounts based on state-level differences in in-person school attendance and program administration
  - Potential for mis-allocation based on reported characteristics
- Release Public-use conditional relationships from linked data (Mittag, 2019)
- Model-Based Imputation
  - Ideally, match conditional distributions, require little understanding of error-correction models (just use the data) and match public aggregates
  - Along with potential benefit of disclosure protection of microdata (essentially synthetic data)

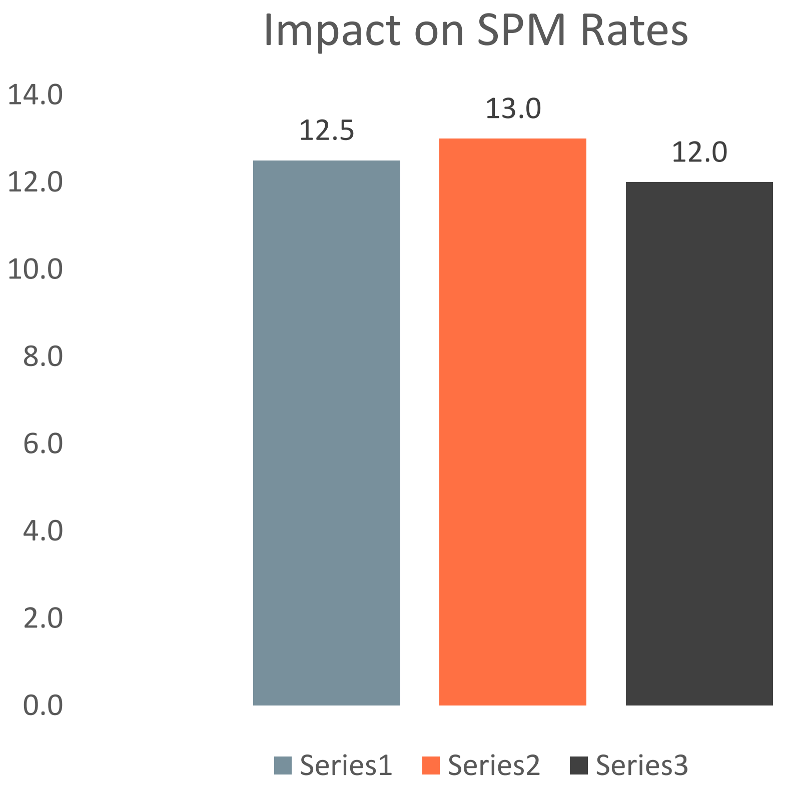
Solutions to Under-reporting
Who Gets Benefits with Survey, TRIM, and Adrecs?



Source: Shantz and Fox, 2019 using the 2010–2016 Annual Social and Economic Supplements (CPS ASEC), Transfer Income Model version 3 (TRIM3), and state Supplemental Nutrition Assistance Program (SNAP) administrative records (AdRecs).

Note: Adjusted using IPW, excluding full line imputes, excluding imputed SNAP receipt and amount, and excluding the top and bottom five percent of observations. The densities have been scaled based on the rates of SNAP receipt. The density for the administrative records curve is one. The unit of analysis is the SPM unit. Values are conditional on positive SNAP benefits in each data source. For information on confidentiality protection, sampling error, non-sampling error, and definitions, see https://www2.census.gov/programs-surveys/cps/techdocs/cpsmar16.pdf.

Solutions to Under-reporting SPM for Survey, TRIM, and Adrecs



Source: Shantz and Fox, 2018

Goal for This Paper

- SNAP data available for some, but not all (or most) states
- Can we use available SNAP data to correct for under-reporting of SNAP benefits nationally?
  - Focus on bias in several sets of estimates
    - Mobility Curves
    - Poverty (SPM)
    - Regression coefficients (SNAP as dependent and independent variable)

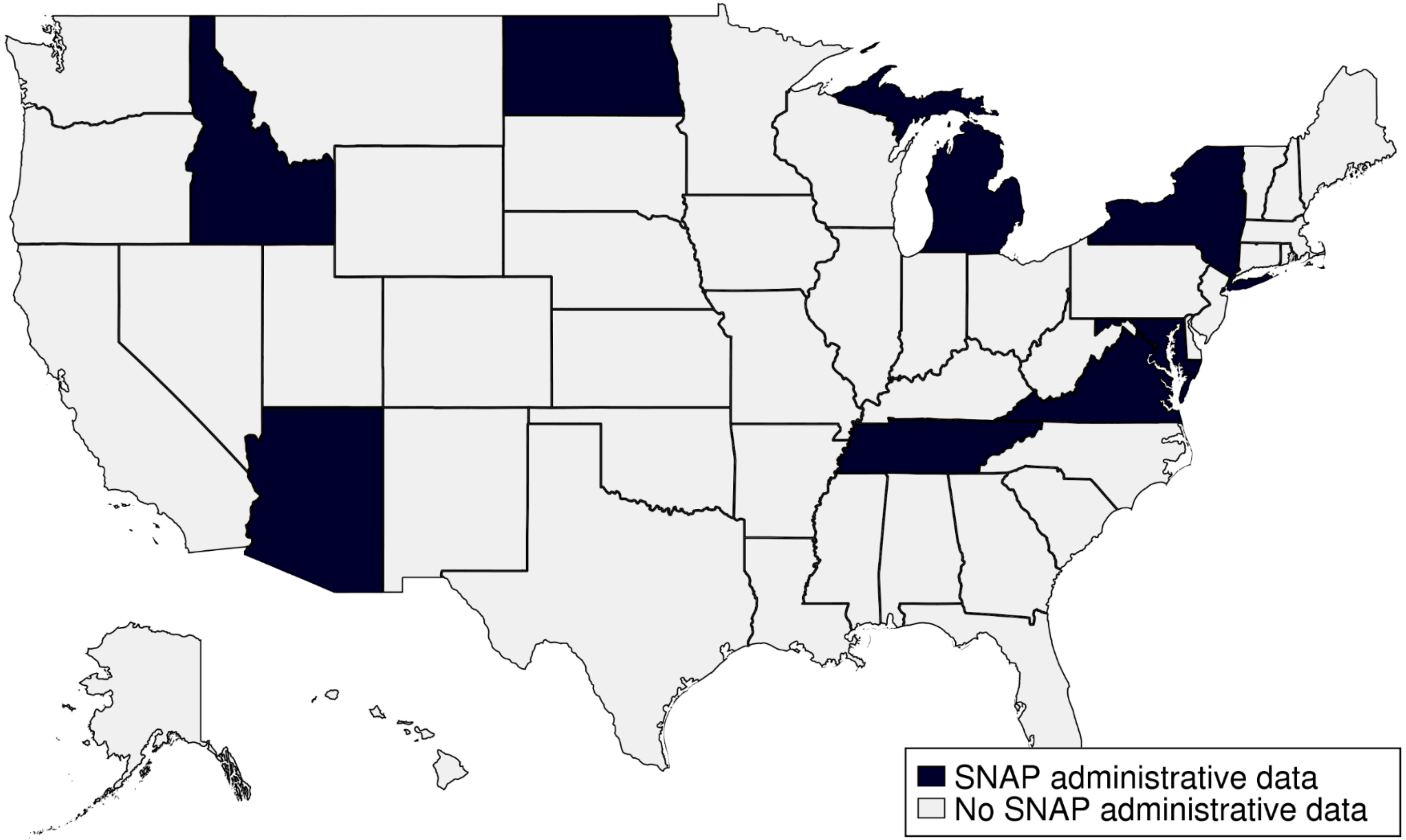
Broader Goals/Research Agenda

- Can we use model-based imputation to address under-reporting in surveys when administrative data is not available for all individuals or in all time periods?
  - Correct under-reporting in time *t* with data available from time *t+/-s*
    - Timely estimates before administrative data is available
    - Historic estimates when administrative data is available in some but not all years
  - Compare to other approaches to imputing missing benefits
- General proof of concept
  - Take administrative data with limited geographic or temporal coverage and impute to places/times without available data
  - Validate the imputation/data synthesis against specific targets of interest
  - Also, protect against disclosure of underlying linked, unreleasable data

Data

- Current Population Survey Annual Social and Economic Supplement (CPS ASEC) for 2014
  - Fielded in February through April
  - Asks respondent about SNAP receipt in the previous calendar year
  - Exclude households with non-PIKed head (<10%)
- IRS and SSA income data (1040s, W2s, 1099-R's and DER files)
- Matched to state administrative SNAP records for 8 states
  - Arizona, Idaho, Maryland, Michigan, New York, North Dakota, Tennessee and Virginia
- USDA state-level data on SNAP receipt rates and benefit amounts (monthly averages)

SNAP Administrative Data Used



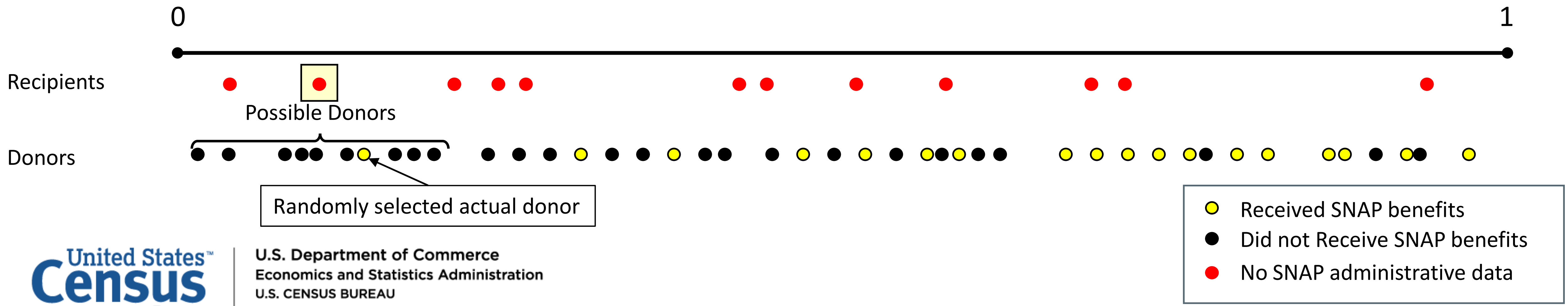
Treat as a Problem of Missing Information

- Want administrative SNAP data from all states
- Only observed in some
- Solved with Imputation
  - Similar to non-response
  - Use observable information to impute plausible administrative SNAP data to households in states where it is not available

Imputation using Predictive Mean Matching

1. Regress administrative SNAP receipt on predictors
   - Use some form of regularization to select from a large set of potential model predictors
   - Use resampling (Bayes Bootstrap) to properly account for uncertainty in model variables and coefficients
2. Assign expected probability of receipt to all households in sample (in states with and without administrative records)
3. For each household in state ***without*** SNAP administrative data (recipients), randomly draw actual SNAP receipt from nearest households in an administrative record state (donor)

Expected Probability of SNAP Recipiency (from regression)



Validation – How do we know if the model is any good?

- Cannot test imputation assumptions directly
- Indirect test
  - Leave-one-out imputations (LOO) – test imputations against observed administrative in each state s with adrecs, if imputation uses only data from the other adrec states

Leave-One-Out Imputation

- 8 separate imputation models
  - One for each state s with administrative data
  - Ignore administrative data state s and use other 7 states in imputation
- Validation
  - SNAP summary stats
  - Mobility curves
  - SPM rates
  - Regression coefficients

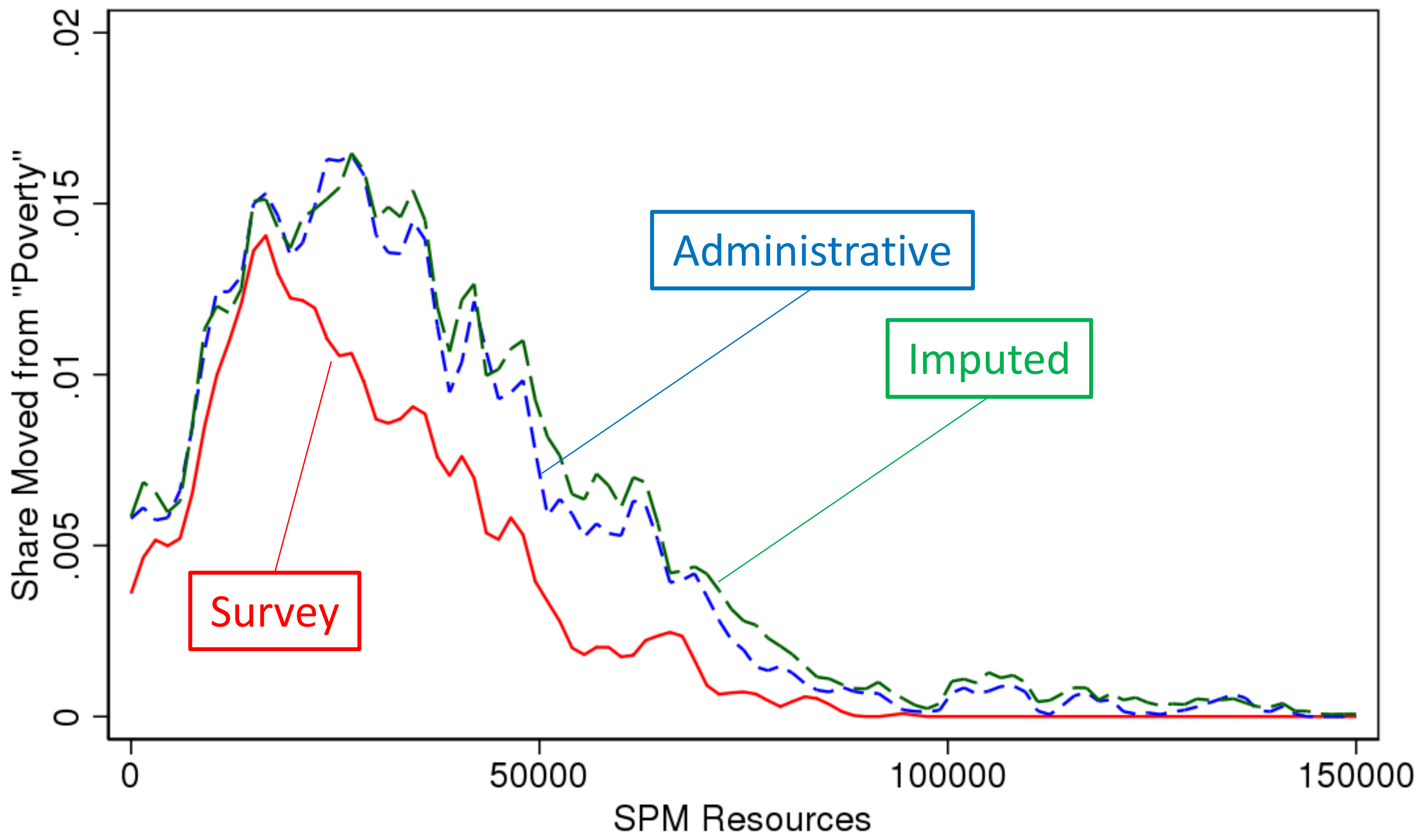
SNAP Validation

Pooled Leave-one-out Imputation in States with Administrative Data

Group	Administrative Data	Leave-One-Out Imputation	Difference
SNAP Receipt	20.0	20.1	0.2
SNAP Benefit Amount			
Average	2,613	2,388	-225**
10th Percentile	396	315	-82*
25th Percentile	992	833	-160*
Median	2,060	1,870	-189**
75th Percentile	3,693	3,200	-493*
90th Percentile	5,821	5,479	-342

Note: ***, **, * indicate difference is significant at the 99, 95, and 90 percent confidence levels, respectively.
Source: U.S. Census Bureau, Current Population Survey, 2014 Annual Social and Economic Supplements (CPS ASEC) and state Supplemental Nutrition Assistance Program (SNAP) administrative records.

Mobility Curve Validation
Pooled Leave-one-out Imputation vs. Administrative and Survey Data



Source: U.S. Census Bureau, Current Population Survey, 2014 Annual Social and Economic Supplements (CPS ASEC) and state Supplemental Nutrition Assistance Program (SNAP) administrative records.

Validation Summary

Pooled Leave-one-out Imputation in States with Administrative Data

- SNAP Results
  - Very good match overall (and by state) of SNAP receipt
  - Understate benefits conditional on receipt, especially for New York and Virginia
- Mobility Curve
  - Very good match for effect at all SPM resource levels
- SPM
  - Agreement between adrecs and leave-one-out relative to survey
  - Leave-one-out point estimates slightly overstate impact of SNAP, but only one difference from adrec is statistically significant
- Regressions
  - Regressions with SNAP as dependent and independent variable
    - No significant difference between SNAP administrative data and leave-one-out imputes
    - Both differ from survey estimates in earnings regression

Next Steps

- Add more states and years
- Continue to improve imputation model
  - Include information on rules and requirements in different states and counties
  - County-level SNAP aggregates from USDA
  - Better regularization/variable selection
  - Better handling of heterogeneity
- Expand to other topics where data is available for some individuals/households, but not others (by geography, time, etc.)
- Release the data as a research extract
- Incorporate into National Experimental Wellbeing Statistics (NEWS) project
  - NEWS – using survey and administrative data to improve income estimates
  - Experimental

Contact Information

Jonathan Rothbaum

Research Economist

Social, Economic, and Housing Statistics Division

jonathan.l.rothbaum@census.gov

(301) 763-9681

Extra Slides

Descriptive Statistics

	States with Admin Records	Other States	Difference
SNAP State-Level Aggregates (USDA)			
Average monthly share of HHs with SNAP	29.4	26.3	-3.10***
Average Monthly Amount *12	3,192	3,305	113***
Survey SNAP Receipt Rate	14.2	12.5	-1.7**
SNAP Benefits	3,508	3,565	58
Poverty Rate (SPM)	15.5	15.5	0.01
Age			
Under 18 years	22.9	23.8	0.9***
18 to 64 years	62.6	62.1	-0.5*
65 years and older	14.5	14.1	-0.4

Note: ***, **, * indicate difference is significant at the 99, 95, and 90 percent confidence levels, respectively.
Source: U.S. Census Bureau, Current Population Survey, 2014 Annual Social and Economic Supplement (CPS ASEC).

Descriptive Statistics

	States with Admin Records	Other States	Difference
Race/Hispanic Origin			
White	73.5	78.6	5.1***
Black	16.2	12.2	-3.9***
Asian	6.3	5.3	-1.0**
Hispanic (any race)	13.9	18.1	4.3***
Education			
No HS Diploma	7.9	7.8	-0.1
HS, No College	20.0	19.8	-0.1
Some College, No Degree	17.3	17.9	0.6
Bachelor's Degree or Higher	22.5	21.1	-1.5***

Note: ***, **, * indicate difference is significant at the 99, 95, and 90 percent confidence levels, respectively.
Source: U.S. Census Bureau, Current Population Survey, 2014 Annual Social and Economic Supplement (CPS ASEC).

Imputation

- Notation
  - O – Characteristics observed for all individuals
    - Includes survey responses, administrative data on W2's, 1040s, 1099Rs
  - Y – SNAP administrative data, $Y = (Y_1, \ldots, Y_p)$
    - Observed if $A = 1$ (adrec states), but not if $A = 0$ (non-adrec states)
- Imputation model is based on the conditional joint density

$$f(Y|O, \theta)$$

where $\theta = (\theta_1, \ldots, \theta_p)$, and each θ_j is a vector of parameters for each Y_j such as regression coefficients and dispersion parameters

Imputation

- For any imputation model, we must impose assumptions on f and θ to assign plausible values to Y where data are missing

- Underlying assumptions for any imputation model
  - Missing at Random (MAR) – don't need unobservable information to account for missingness
$$f(Y|O, A = 1) = f(Y|O, A = 0)$$
  - Proper/Congenial – for a statistic Q (estimate, regression coefficient, etc.), the imputation model is congenial if:*
$$E\left(\hat{Q}\middle|O, A = 1\right) = E\left(\hat{Q}\middle|O, A = 0\right)$$

* Congeniality also requires the distributions of $\hat{Q}$ to converge so that SE estimates and confidence intervals are valid

Validation – How do we know if the model is any good?

- Cannot test MAR or congeniality assumptions directly
- Indirect test
  - Leave-one-out imputations (LOO) – test imputations against observed administrative in each state s with adrecs, if imputation uses only data from the other adrec states
  - MAR

$$f(Y|O, A = 1, State \neq s) = f(Y|O, A = 1, State = s)$$

  - Congeniality

$$E(\hat{Q}|O, A = 1, State \neq s) = E(\hat{Q}|O, A = 1, State = s)$$

Imputation Process

- Impute (Y)
  1. Household SNAP receipt
  2. Months of receipt
  3. Annual SNAP amount
- Model (f) – predictive mean matching
- SNAP predictors (O) – what goes in the models?
  - The kitchen sink!
  - Survey responses for household head, spouse, and summarized at the household level
  - Tax information
    - W-2 earnings histories
    - 1099-R retirement income (defined-benefit pensions and defined-contribution withdrawals)
    - 1040 income information (AGI, interest, dividends, gross rent, …)
  - State-level SNAP aggregates

Imputation Process – Too Much Detail

- Practical considerations – O has too many variables
  - Reduce dimenstionality of O using variable selection model
  - In practice, this imposes assumptions on θ_j
- θ_j is unknown and must be estimated
  - Uncertainty in $\hat{\theta}_j$ must be accounted for
- Use Bayes' Bootstrap before selection model and predictive mean matching regressions to approximate distribution of $\hat{\theta}_j$ and account for uncertainty
- Create multiple implicates (independent imputations) to calculate uncertainty for any $\hat{Q}$ of interest

SPM Validation

Pooled Leave-one-out Imputation in States with Administrative Data

Group	Survey Only (Official)	Administrative Data	Leave-One-Out Imputation	Difference		
				Adrecs - Survey	LOO - Survey	LOO - Adrecs
All People	14.49	14.09	13.95	-0.41***	-0.54***	-0.13
Age						
Under 18	15.35	14.73	14.70	-0.62**	-0.65**	-0.04
18 to 64	14.48	14.10	14.00	-0.37***	-0.48***	-0.11
65 and older	13.18	12.96	12.56	-0.22	-0.62	-0.40
Race/Hispanic Origin						
White	12.17	11.72	11.60	-0.45***	-0.57**	-0.13
Black	23.66	23.35	23.17	-0.32	-0.49	-0.18
Hispanic	24.89	24.09	24.29	-0.80*	-0.59	0.21

Note: ***, **, * indicate difference is significant at the 99, 95, and 90 percent confidence levels, respectively.
Source: U.S. Census Bureau, Current Population Survey, 2014 Annual Social and Economic Supplements (CPS ASEC) and state Supplemental Nutrition Assistance Program (SNAP) administrative records.

SPM Validation

Pooled Leave-one-out Imputation in States with Administrative Data

Group	Survey Only (Official)	Adrec States	Leave-One-Out Imputation	Difference		
				Adrecs - Survey	LOO - Survey	LOO - Adrecs
Education						
No HS Diploma	33.01	31.21	30.89	-1.81***	-2.13**	-0.32
HS, No College	16.00	15.79	15.54	-0.21	-0.46*	-0.25
Some College, No Degree	11.31	11.27	11.05	-0.04	-0.26	-0.22
Bachelor's Degree or Higher	6.30	6.19	6.09	-0.11**	-0.21**	-0.10

United States Census Bureau
U.S. Department of Commerce
Economics and Statistics Administration
U.S. CENSUS BUREAU
census.gov

Note: ***, **, * indicate difference is significant at the 99, 95, and 90 percent confidence levels, respectively.
Source: U.S. Census Bureau, Current Population Survey, 2014 Annual Social and Economic Supplements (CPS ASEC) and state Supplemental Nutrition Assistance Program (SNAP) administrative records.

Regression Validation - Predicting "True" SNAP Receipt from Survey Responses

Pooled Leave-one-out Imputation in States with Administrative Data

	Adrecs (1)	Leave-One-Out (2)	Difference (3)
Survey SNAP Recipient	0.729***	0.752***	0.023
	(0.032)	(0.041)	(0.053)
Non-Response to SNAP Receipt	0.195***	0.243***	0.048
	(0.022)	(0.026)	(0.031)
Female	-0.019***	-0.014	0.005
	(0.007)	(0.009)	(0.011)
Race			
Black	0.079***	0.055***	-0.025
	(0.014)	(0.013)	(0.018)
Native American	0.013	0.016	0.003
	(0.032)	(0.039)	(0.041)
Asian	0.013	0.022	0.009
	(0.016)	(0.019)	(0.024)
Pacific Islander	-0.128***	-0.008	0.120
	(0.038)	(0.188)	(0.204)
Hispanic	0.074***	0.064***	-0.010
	(0.013)	(0.020)	(0.023)

	Adrecs (1)	Leave-One-Out (2)	Difference (3)
Survey SNAP Amount	0.007	0.006	-0.001
	(0.014)	(0.017)	(0.022)
Survey SNAP Amount^2	-0.001	-0.001	0.000
	(0.002)	(0.002)	(0.002)
Age	-0.001	0.001	0.002
	(0.002)	(0.002)	(0.003)
Age^2	-0.000017	-0.000029	-0.000012
	(0.000015)	(0.000022)	(0.000025)
Household Income	0.019*	0.022	0.003
	(0.010)	(0.022)	(0.026)
Household Income^2	-0.0031***	-0.0030***	0.0000
	(0.0005)	(0.0010)	(0.0012)
Household Income $\neq 0$	0.090	-0.002	-0.092
	(0.083)	(0.142)	(0.167)
Constant	0.280***	0.275***	-0.005
	(0.062)	(0.089)	(0.104)
R2	0.48	0.48	
N	7,400	7,400	

Note: ***, **, * indicate difference is significant at the 99, 95, and 90 percent confidence levels, respectively.
Source: U.S. Census Bureau, Current Population Survey, 2014 Annual Social and Economic Supplements (CPS ASEC) and state Supplemental Nutrition Assistance Program (SNAP) administrative records.

Regression Validation – Association between Earnings ($\neq 0$) and SNAP Receipt

Pooled Leave-one-out Imputation in States with Administrative Data

	Administrative Data	Leave-One-Out	Survey
SNAP Receipt	-0.175***	-0.178***	-0.227***
	(0.0185)	(0.0184)	(0.0225)
Female	0.0284***	0.0269***	0.0256***
	(0.00951)	(0.00971)	(0.00972)
Urban	0.00465	0.00169	0.00373
	(0.0110)	(0.0117)	(0.0115)
Race/Ethnicity			
Black	-0.00384	-0.00541	-0.00519
	(0.0149)	(0.0146)	(0.0147)
Native American	-0.0280	-0.0275	-0.0174
	(0.0336)	(0.0352)	(0.0348)
Asian	-0.0153	-0.00975	-0.0200
	(0.0192)	(0.0189)	(0.0192)
Pacific Islander	0.0682*	0.0967	0.0789**
	(0.0410)	(0.0601)	(0.0375)
Hispanic	0.00884	0.00787	0.00304
	(0.0199)	(0.0191)	(0.0199)

	Administrative Data	Leave-One-Out	Survey
Education			
High School	0.118***	0.119***	0.109***
	(0.0257)	(0.0257)	(0.0255)
Some College	-0.0206	-0.0215	-0.0215
	(0.0182)	(0.0184)	(0.0183)
Associates	0.0630***	0.0645***	0.0609***
	(0.0177)	(0.0184)	(0.0186)
Bachelors	-0.00478	-0.00358	-0.000165
	(0.0138)	(0.0143)	(0.0138)
Masters	0.00919	0.0109	0.0122
	(0.0124)	(0.0127)	(0.0125)
Age	0.0306***	0.0312***	0.0306***
	(0.00393)	(0.00398)	(0.00395)
Age^2	-0.000399***	-0.000405***	-0.000400***
	(0.0000449)	(0.0000454)	(0.0000451)
Constant	0.258***	0.248***	0.263***
	(0.0914)	(0.0910)	(0.0905)
R^2	0.14	0.14	0.15
N	5,200	5,200	5,200

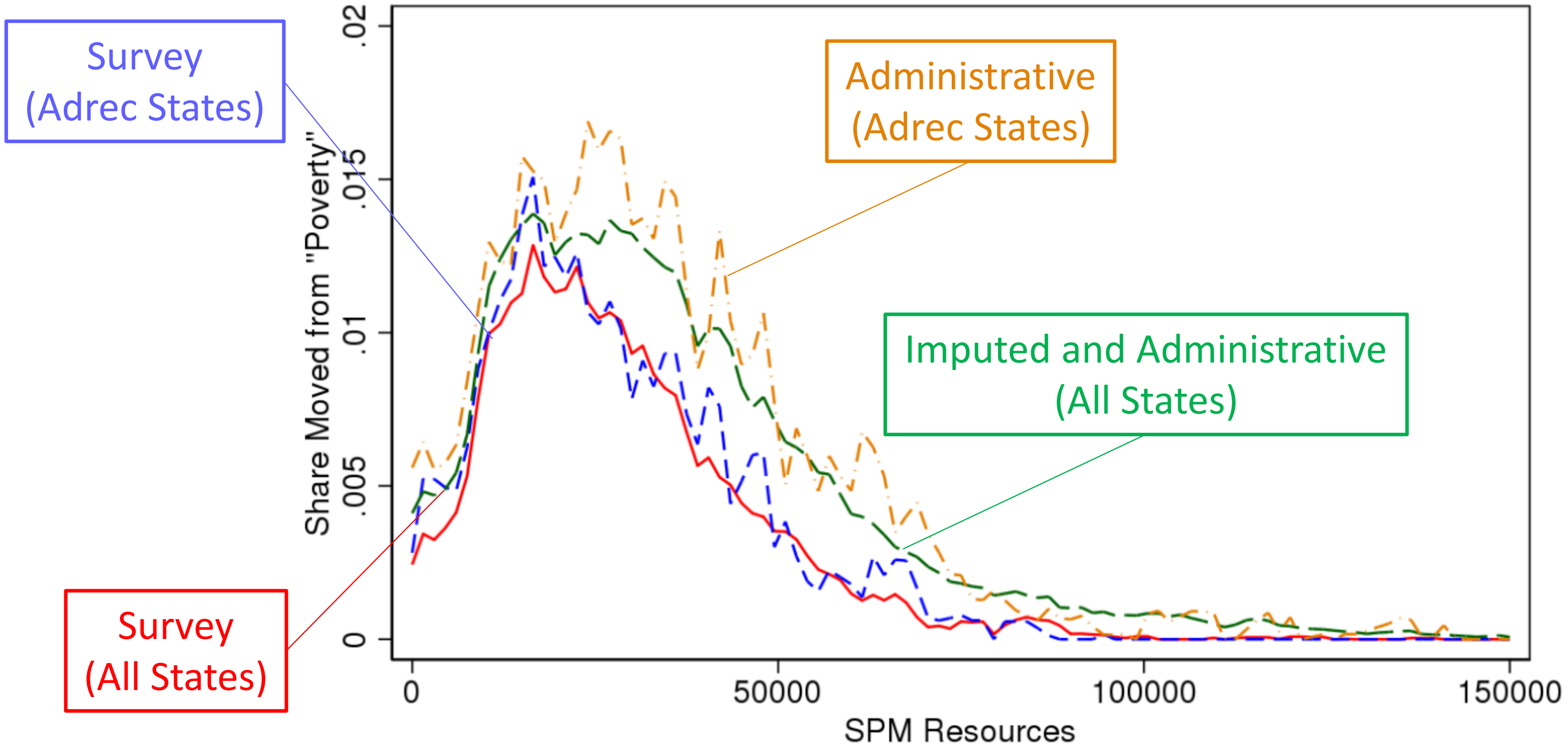
Note: ***, **, * indicate difference is significant at the 99, 95, and 90 percent confidence levels, respectively.
Source: U.S. Census Bureau, Current Population Survey, 2014 Annual Social and Economic Supplements (CPS ASEC) and state Supplemental Nutrition Assistance Program (SNAP) administrative records.

National Estimates

- Impute SNAP administrative data for the other 42 states and DC
- Use administrative data for the 8 states where it is available
- Estimate
  - Mobility curve
  - SPM
  - Earnings regression

National Estimate – Mobility Curve

Imputed in Non-adrec States and Administrative Data in Adrec States



Source: U.S. Census Bureau, Current Population Survey, 2014 Annual Social and Economic Supplements (CPS ASEC) and state Supplemental Nutrition Assistance Program (SNAP) administrative records.

National Estimate - SPM

Imputed in Non-adrec States and Administrative Data in Adrec States

Group	Survey Only (Official)	Imputed + Adrecs	Difference
All People	14.58	14.36	-0.22***
Age			
Under 18	15.40	15.38	-0.02
18 to 64	14.35	14.08	-0.28***
65 and older	14.24	13.89	-0.34***
Race/Hispanic Origin			
White	12.75	12.55	-0.20**
Black	24.14	23.56	-0.58
Hispanic	23.84	23.43	-0.41

Note: ***, **, * indicate difference is significant at the 99, 95, and 90 percent confidence levels, respectively.
Source: U.S. Census Bureau, Current Population Survey, 2014 Annual Social and Economic Supplements (CPS ASEC) and state Supplemental Nutrition Assistance Program (SNAP) administrative records.

National Estimate – Earnings Regressions

Imputed in Non-adrec States and Administrative Data in Adrec States

- A handful of coefficients differ between survey-only regressions for samples in Adrec ($A = 1$) and non-adrec ($A = 0$) states
  - Such as for Pacific Islanders, associate's degree holders
- Find similar differences in regressions uses administrative data estimates
- No significant differences in difference-in-difference estimates
  - Difference 1: administrative – survey estimates
  - Difference 2: adrec – non-adrec states