Using Linked Survey and Administrative Data to Better Measure Income: Implications for Poverty, Program Effectiveness and Holes in the Safety Net

USDA-Census Administrative Data Conference
November 1, 2016

Bruce D. Meyer, University of Chicago and NBER

Nikolas Mittag, CERGE-EI/Charles University

The views expressed are not necessarily those of the U.S. Census Bureau and the New York State Office of Temporary and Disability Assistance. We thank the Census Bureau and the NY OTDA for providing the data.

General Overview

- Household surveys are a critically important source of information to both researchers and policy makers.
- However, the accuracy of household surveys is declining
- Here, we show substantial bias due to underreporting of transfer income in analyses of low-income households
- More generally, we argue that linking administrative records is an important way to improve survey data

Summary

- In the CPS, the official source of income data, missing gov't transfers are a major problem when studying those with low income.
- Using linked administrative data, we find about half of food stamps are missing in the survey as well as 2/3 of cash welfare and housing assistance dollars for recipients
- Correcting the survey data sharply changes our under-standing of the circumstances of the poor and program effects
- Several widely cited types of analyses need to be changed
  - Distributional analyses: Incomes at bottom much higher
  - Program effects: Poverty reducing effects of programs much greater
  - Holes in safety net: Many fewer people missed by transfer programs

Table 1: Survey Errors in Transfer Receipt Reporting, CPS New York , 2008-2011

Error Type	Sample	SNAP	Public Assistance	Housing Assistance
False negatives	True recipients	42.8%	63.3%	35.6%
False positives	True non-recipients	1.9%	0.7%	2.8%
Abs. error in amount >$500	Recipients who report	53.22%	87.89%	97.50%
Mean of true amount	Recipients who report	$3,389	$5,213	$12,000
Mean of reported amount	Recipients who report	$3,170	$3,152	$3,081
SD of error in amount	Recipients who report	$2,392	$4,619	$8,776
Corr. true & reported amount	Recipients who report	0.55	0.22	0.07

Note: Estimation uses households with at least one PIKed member only, weights are adjusted for PIK rates. SNAP and public assistance amounts are average annual receipt per household, housing assistance amounts are annualized from monthly amounts per household. False positives for housing assistance may be recipients of non-HUD housing programs and therefore should not necessarily be interpreted as survey errors.

Outline

1. Data Quality Problems in Surveys
2. Linking Survey and Administrative Data
3. Re-visiting 3 Prototypical Analyses
   - Income of Poor Households
   - Poverty Reducing Effects of Programs
   - Holes in the Safety Net
4. Conclusions, Caveats and Extensions

Why is there growing misreporting?

- I am not going to talk about why misreporting occurs and why it is getting worse over time. That is in other papers.
- See Meyer, Goerge and Mittag (2016), Celhay, Meyer and Mittag (2016) for the reasons people misreport.
- See Meyer, Mok, Sullivan (2015) for why these problems are getting worse.
- This paper is about whether declining survey quality matters.

Why focus on summary stats for income?

- Summary statistics on the income distribution are used by academics, policy analysts, and govt officials.
- The predictions of bias are clear and robust. If you miss a lot of dollars, people look worse off, programs appear less effective.
- If we were to look at causal models with benefit receipt as the dependent variable or explanatory variable the results would be more dependent on specification.
- If errors uncorrelated with explanatory variables, large attenuation of coefficients in both cases.
- But errors are correlated with explanatory variables--Meyer and Mittag (2016)

Putting CPS data in perspective

- In this paper, we focus on the CPS, which is the source of official poverty statistics
- However, problems are similar in other household surveys.
- For comparison, consider net reporting rates calculated as in Meyer, Mok, Sullivan (JEP 2015):

$$\$\ Reporting\ Rate = \frac{Total\ weighted\ \$\ reported\ in\ survey}{Total\ \$\ paid\ out}$$

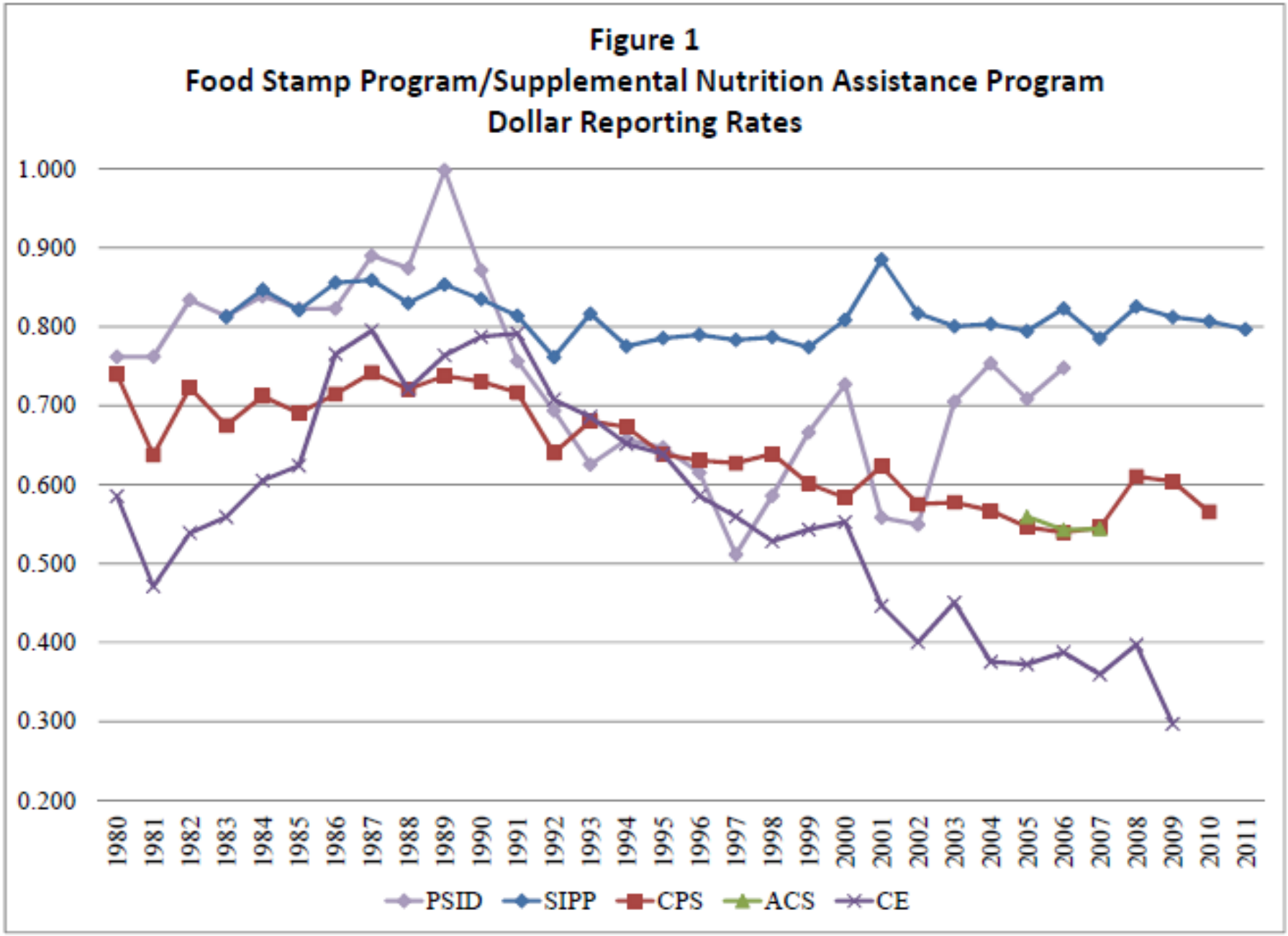


Figure 1
Food Stamp Program/Supplemental Nutrition Assistance Program
Dollar Reporting Rates
1.000
0.900
0.800
0.700
0.600
0.500
0.400
0.300
0.200
1980
1981
1982
1983
1984
1985
1986
1987
1988
1989
1990
1991
1992
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
PSID
SIPP
CPS
ACS
CE

Linked Admin and Survey Data

- Household Survey Data
  - 2008-2013 CPS ASEC (income data for 2007-2012)
- Cash Welfare and Food Stamp Data
  - 2007-2012 NY OTDA
  - SNAP, TANF, GA
  - Recipients, dates, amounts
- Public and Subsidized Housing Data
  - 2009-2012 HUD PIC and TRACS (gives housing in 2008-2011)
  - Recipients, dates, rent measures, location

Methods

- Link the data sources using the Protected Identification Key (PIK) an anonymized version of the SSN attached to each source
- The admin data have a PIK attached more than 99 percent of the time
- The CPS has a PIK attached over 90 percent of the time at household level (our unit of analysis)
- Use IPW to account for the probability that a household has a PIK (doesn't affect results much)
- Substitute admin receipt and amounts for reported ones

Advantages of Linked Data

- We argue that administrative record can be considered “truth” here:
  - Actual payments validated by agencies
  - Definitions comparable to survey definition
  - High match rate
- Rare opportunity to compare reports to “truth”
- Resulting dataset combines detail of survey and accuracy of administrative measure
- Caveat: in-kind transfers valued at face value

More Methods and Definitions

- We combine TANF and GA into public assistance (PA) because there is program substitution by the state and confused respondents
- Rental Assistance
  - Value of rental assistance = gross rent - tenant payment
  - If gross rent missing impute: zip code, household size
  - Data do not include non-HUD programs, so we rely on survey report when a household is not in admin files
  - CPS sharply under-imputes rental assistance
  - Experimental, supplemental files have better measures
- We use two alternative base income measures: official pre-tax cash income and SPM-type income

Table 1: Survey Errors in Transfer Receipt Reporting, CPS New York , 2008-2011

Error Type	Sample	SNAP	Public Assistance	Housing Assistance
False negatives	True recipients	42.8%	63.3%	35.6%
False positives	True non-recipients	1.9%	0.7%	2.8%
Abs. error in amount >$500	Recipients who report	53.22%	87.89%	97.50%
Mean of true amount	Recipients who report	$3,389	$5,213	$12,000
Mean of reported amount	Recipients who report	$3,170	$3,152	$3,081
SD of error in amount	Recipients who report	$2,392	$4,619	$8,776
Corr. true & reported amount	Recipients who report	0.55	0.22	0.07

Note: Estimation uses households with at least one PIKed member only, weights are adjusted for PIK rates. SNAP and public assistance amounts are average annual receipt per household, housing assistance amounts are annualized from monthly amounts per household. False positives for housing assistance may be recipients of non-HUD housing programs and therefore should not necessarily be interpreted as survey errors.

Characterizing misreporting

- 36, 43, and 63 percent of housing assistance, SNAP, and PA recipients, respectively, do not report
- Overall measurement error main source of error
  - Most is (0,1) receipt question
  - Amounts not strongly biased for SNAP
  - Bias and error in amounts for PA and housing assistance important, both biased sharply downward
    - PA payments typically include part that goes directly to landlord
    - Housing imputations – there are better imputations than those on ASEC file, use as robustness check

Prototypical Analysis I

- Studies of the income distribution, the distribution of transfers across the income distribution, and program targeting
  - Census (2015) finds poverty rates and poverty gaps high, and both have risen recently
  - Gottschalk and Smeeding (1997), Blank and Schoeni (2003), Gottschalk and Danziger (2005), Burkhauser, Feng and Jenkins (2009), Armour, Burkhauser and Larrimore (2013)

Prototypical Analyses I continued

- Income v. Consumption
  - Meyer and Sullivan (2012a, 2012b, 2014) find that consumption poverty and consumption percentiles show a different pattern than those using income, one that is much more favorable, especially over the last 15 years or so. Largest diffs for deep poor
  - We conjecture that much of the difference between income and consumption at the bottom is unreported government transfers

Figure 2. Official and Alternative Income Poverty Rates and Consumption Poverty Rate, 1960–2010[a]

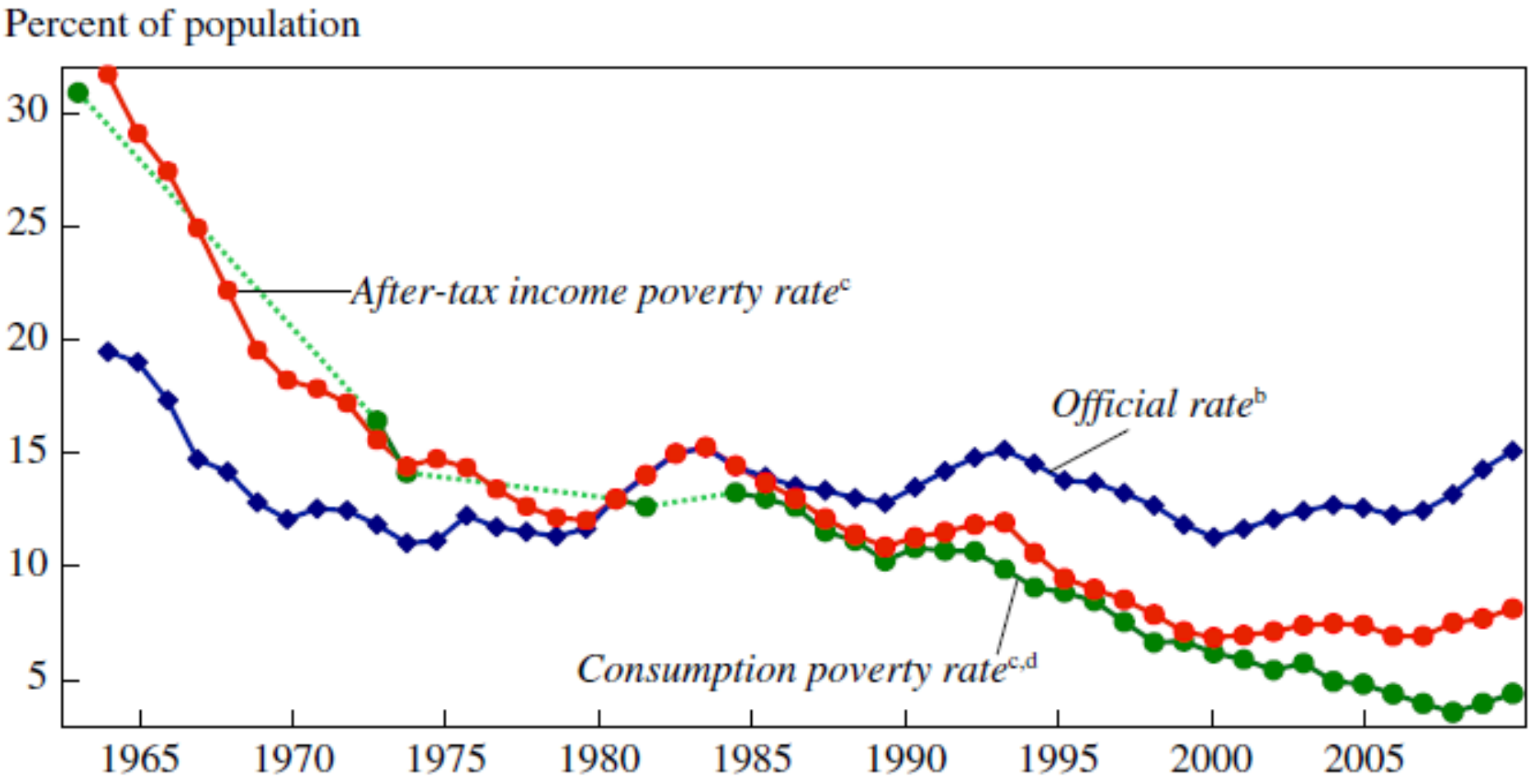


Figure 4. Change in the Average Poverty Gap Relative to 1980 for Poor Families for Income and Consumption Measures, 1960–2010[a]

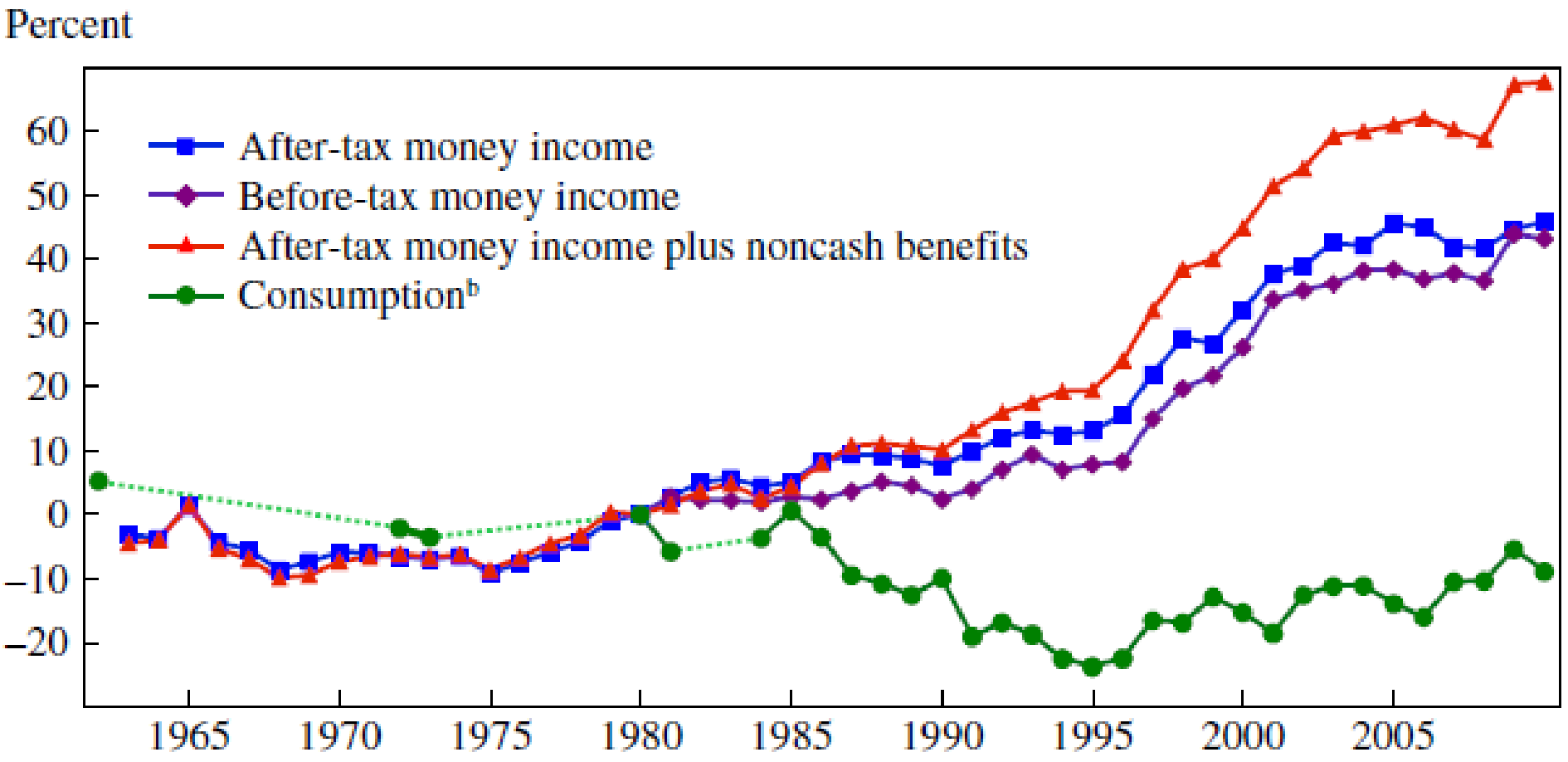


Table 2: Annual Unreported Per Capita Income by Source, CPS New York, 2008-2011

	< 50% FPL		50-100% FPL	
	amount	% of base	amount	% of base
PA	$373	28.6%	$187	3.4%
SNAP	$135	10.3%	$214	3.9%
Housing	$932	71.4%	$1,150	21.0%
All Programs	$1,438	110.2%	$1,548	28.3%

Base income is reported cash income. Income categories are defined based on pre-tax money income, poverty thresholds are the official poverty thresholds. Dollars received are 2012 dollars, but not adjusted for family size. Estimation uses households with at least one PIKed member only, weights are adjusted for PIK rates.

Effect of Missing Dollars on Income Distribution

- Unreported income equals more than reported cash income for those in deep poverty
  - 29 percent increase due to PA
  - 10 percent increase due to SNAP
  - 71 percent increase due to housing assistance
- Effect of including unreported dollars larger than effect of including reported non-cash benefits
- Effect on income fades out quickly as income rises
- For those between half the poverty line and the poverty line the increase is
  - Just over 7 percent for PA and SNAP combined
  - 21 percent for housing assistance

Missing Dollars Across Income Distribution

- Public Assistance:
  - Due to low net reporting rates at low income levels; often two-third or more of dollars per person missing for very low reported income cells.
  - Missing dollars fall off quickly as income (and reporting rates) rise
- SNAP:
  - Missing dollars peak at 100-150% of the poverty line
  - Remain substantial at higher income levels
- Housing Assistance:
  - Missing dollars spread out across lower parts of the income distribution

Deep poor subgroups

- For single mothers, unreported transfers even more of an issue and bias greater
- For disabled bias about average
- For 65+ bias smaller
  - But, we don’t have administrative data on the key programs for the disabled and aged
- I will talk about the actual numbers more in the next section

Comments

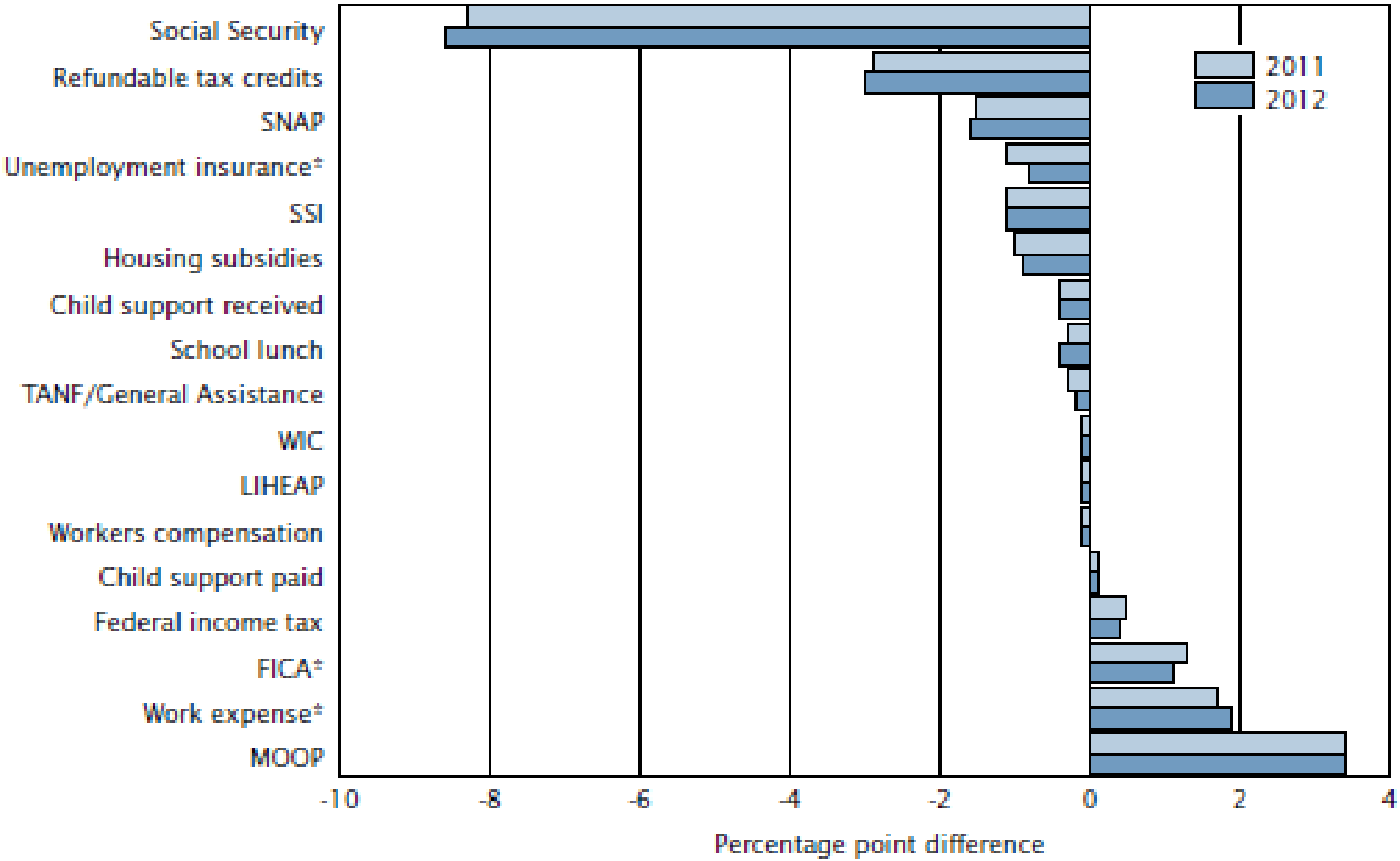
- We are reporting only the role of *missing* benefits; the role of all SNAP and housing benefits is greater.
- However, we are also not including SNAP and housing assistance in our base income
- We do include these benefits in our base in alternative estimates where base income is a version of SPM income (after-tax plus non-cash benefits as imputed in CPS)
- Lower percentage increases since base higher

Prototypical Analysis II

- Studies of the poverty reducing effects of programs
  - Census (annual) Supplemental Poverty Measure (SPM) report
  - Scholz, Moffitt and Cowan (2009), Ben-Shalom, Moffitt and Scholz (2012) use the SIPP

Figure 5.

Difference in SPM Rates After Including Each Element: 2011 and 2012



*Statistically significant change between 2011 and 2012.

Source: U.S. Census Bureau, Current Population Survey, 2012 and 2013 Annual Social and Economic Supplements.

Some try to account for misreporting

- Some papers try to account for under-reporting
  - Meyer and Sullivan (2006), Meyer (2010)
  - Scholz, Moffitt and Cowan (2009), Moffitt, Scholz (2010), Ben-Shalom, Moffitt and Scholz (2012)
  - Last set of papers most sophisticated, but
    - Uses observed reports to infer true reporting which is biased; see Meyer, Goerge and Mittag (2014), Celhay, Meyer and Mittag (2015)
    - Imputes receipt with certainty to those estimated to most likely be recipients rather than based on the probability they are recipients
    - Assumes no false positives

Measuring Poverty Reduction

- Compute the effect of a program on poverty as the fraction of recipients with income below the poverty line without, but above it with the transfer
- Assumes no behavioral response
  - Scholz, Moffitt and Cowan (2009): "...low-income individuals respond to these incentives, but that the magnitude of the response is small…"
  - Moffitt et al. (2012) bears out the above conclusion in simulations, with some caveats

Table 3: Poverty Reduction, CPS New York, 2008-2011

	Survey Change	Admin Change	Difference	% of Survey
PA	0.19%	0.47%	0.28%	149.7%
SNAP	1.59%	2.09%	0.50%	31.2%
Housing	0.86%	2.56%	1.71%	199.1%
PA, SNAP	1.89%	2.75%	0.86%	45.2%
All three	2.79%	5.29%	2.50%	89.7%

Note: Baseline Poverty Rate in the Survey is 13.65% (based on pre-tax cash income excluding public assistance).

Subgroup poverty reduction; Poverty gap

- Sharp differences by subgroup; for single mothers a 11 percentage point reduction from transfers missed, poverty falls from 37.5 percent to 19.2 percent. The reduction missed is 1.5 times the survey reduction
- You miss almost all of the effect of PA on the poverty rate for single mothers; about half or more of the reduction in the poverty gap
- For all groups combined almost half of poverty gap filled by PA, SNAP and housing benefits
  - 30 percent of this missed by the survey data

Program effects in detail, over time

- Misreporting varies by income and program, so survey distorts both overall and relative importance of programs
- The rise in the effect of the programs over these years is also 40 percent greater when one includes those transfers missed in the survey data

Prototypical Analysis III

- Holes in safety net; Often called "disconnectedness"
  - Defined as those without work income or welfare income (alternative looser definition includes those with up to 1K of benefits and 2K of earnings)
  - Blank and Kovak (2007, 2009) find high rates and find that rates that have risen over time;
  - Bitler and Hoynes (2010), Loprest (2011), Loprest and Nichols (2011) and others have looked at "disconnectedness"
  - Closely related to "2 dollar a day" literature

Table 4: Disconnectedness Rates, CPS New York, 2008-2011

	No earnings and program receipt			Low earnings and program receipt		
Included Programs	Survey	Admin	% Over-statement	Survey	Admin	% Over-statement
Public Assistance only	17.1%	12.8%	34%	22.8%	17.1%	33%
Public Assistance and SNAP	5.3%	3.2%	68%	7.4%	5.1%	46%
PA, SNAP and Housing	3.6%	1.7%	113%	5.1%	3.0%	71%
All Cash Programs	5.4%	3.6%	50%	9.8%	7.0%	40%
All Cash Programs and SNAP	not disclosed			3.5%	1.9%	82%

All definitions restrict the sample to households headed by an unmarried female with at least one own, grand, related or foster child present in the household. In column 2-4, we consider households that have no earnings and receive none of the programs in the first 3 columns as left behind, columns 5-7 also include those with yearly earnings less than $2000 and combined program receipt of less than $1000 (2005 dollars).

Estimates of safety net gaps: no work or welfare

- We consider a variety of definitions, varying what programs we include, initially require no earnings, no benefits, but then allow up to 2K in earnings, 1K in benefits
- Levels of disconnectedness overstated by 33-113 percent, combining all years.
- Numbers fall by 2/3 when include SNAP
- Almost no one disconnected when include all cash programs and SNAP
- Share of disconnected single mothers does rise over time still

Caveats

- Age, education, race, share Hispanic similar to all of U.S.
- We overstate the national problem because New York has a more generous welfare system
  - Poverty rate higher, PA, SNAP receipt rates higher and benefits per capita higher in NY
  - Housing assistance twice as common as in rest of U.S.
- We understate problem because
  - only account for admin data on a few programs (we miss errors in SSI, OASDI, UI, other pensions etc.)
  - New York CPS reporting better than other states
  - Problem getting worse over time
- Treat admin data as truth; less good with housing data which have been less scrutinized

Conclusions

- Accounting for unreported transfers sharply changes our understanding of the income distribution
  - Particularly at the bottom and for single mothers
  - Reduction in poverty as great for missing dollars as for reported non-cash transfers (including non-cash transfers key argument for new poverty measure--SPM)
- Unreported transfers lead to sharp understatement of program effects
- Unreported transfers lead to an overstatement of the number of those “falling through the cracks”
- Extensions: more states and programs
- Administrative data should be made more widely available and incorporated in surveys