Using Administrative Records to Assess Accuracy of Reporting in the Survey of Income and Program Participation: The Case of SNAP in Texas

Daniel Perez-Lopez (US Census Bureau)
Jeongsoo Kim (US Census Bureau)

ABSTRACT

Previous research has found that benefit receipt tends to be under-reported in population surveys. This study examines the accuracy of reporting of SNAP benefits and amounts in the Census Bureau's Survey of Income and Program Participation (SIPP). We linked Texas SNAP administrative records to SIPP 2004 panel data covering 2004 through 2007. Using the linked data, we utilized descriptive statistics and logistic regression to analyze the factors associated with under-reporting. The preliminary results revealed that about 15% of person-months of the linked data under-reported SNAP receipt in SIPP. Logistic regression showed that Hispanic, never-married, less than high school graduate, and female householder populations were more likely to under-report than their counterparts. For example, Hispanics are more likely to under-report their SNAP receipt than non-Hispanics by 20 percentage point, holding other factors constant.

Extended Abstract

The number of participants in the Supplemental Nutrition Assistance Program (SNAP), the largest anti-hunger program administered by the U.S. Department of Agriculture (USDA), has increased over the past decade by 130%, from 17 million in 2000 to 40 million in 2010 (USDA, 2011). The U.S. Census Bureau's Survey of Income and Program Participation (SIPP) is a nationally representative household survey that collects data on SNAP participation, and the amount of SNAP benefits, in addition to demographic and work history characteristics, and participation in other social welfare and insurance programs.

Several studies have shown that estimates for government program benefits that relied on self-report are inaccurate. Meyer and Goerge (2010) examined SNAP receipt and Temporary Assistance for Needy Families (TANF) receipt in the Current Population Survey (CPS) in 2004 and indicated that about forty percent of months of program receipt were not reported.

In an effort to improve SNAP coverage in surveys, the Census Bureau has requested SNAP administrative records from state program agencies, and is currently matching them to existing

survey data. As part of this effort, the Census Bureau matched SNAP records from Texas to the 2004 SIPP panel, which covers the years 2004 through 2007.

Using the file of matched administrative data and SIPP data, this paper examines the under-reporting of SNAP receipt and benefit amounts in the SIPP. We examine what share of SIPP data (at the person-month level) under-reports the SNAP receipt, how large are the differences in the SNAP benefit amount reported in the SIPP and in the administrative records, how under-reporting behavior differs by household characteristics, and what factors are associated with under-reporting (at the person level).

Data and Methods

We used two different data sets for this study. First, the administrative record file from Texas, which has a different entry for every person in every month they received SNAP benefits. In months in which a respondent did not receive SNAP benefits from Texas, there is no entry in the administrative record.

Our second data set, the SIPP, collects monthly data in 4-month "waves". While every effort is made to contact respondents once they enter the SIPP and to follow them if they move between waves, there is some level of wavely non-response. The matching of the administrative record to the SIPP was done at the person-month level for every SIPP person-month where a respondent was interviewed or an interview was given by proxy. Children under the age of 15 always respond by proxy. Person-months without a response in the SIPP were disregarded, even if they were matched to a person-month of SNAP receipt in the administrative record. This is because the purpose of this research is to explore the accuracy of reporting of benefits in the survey. Meanwhile, since there is no negative "response" possible on the administrative record, months of SIPP response that did not have a corresponding person-month entry in the administrative record were considered matched to the administrative record, and a month of non-receipt in the administrative record was tabulated.

The 4 years of data were matched under the conditions described above, yielding a total of 170,176 matched person-months. Based on the person-month level data, means and frequencies for all demographic characteristics were calculated.

We also conducted logistic regression analysis at the person level to examine the factors associated with under-reporting, where the number of observation was 5,331. The dependent variable was coded as 1 if the person-month under-reported and coded as 0 otherwise. The independent variables were selected to capture important demographic and household characteristics such as race, education, age, age squared, household type, and citizenship

status. All estimates in the regression were weighted and clustered robust standard errors were used to address SIPP's complex sample design.

Preliminary Findings

Descriptive statistics for our person-month dataset show that about one quarter of the person-months in our analysis came from children, and just over 53% were from women. (Table1)

Table 2 shows a weighted average of responses by person-month over the four years, 2004-2007 from the SIPP and Texas administrative record, which includes the incidence of under-reporting. Responses in the lower left quadrant in table 2 are incidences of under-reporting , or false negatives as respondents reported not receiving SNAP, but the Texas administrative record showed that the person received SNAP in that month. On the other hand, responses in the upper right quadrant in the table are over-reporting, or false positives, as respondents report receiving SNAP in a given month but the administrative record shows that Texas did not give them benefits in that month. The average annual under-reporting (false negative) rate for these person-months was 14.5%, while the average annual over-reporting (false positive) rate was 5%.

In table 3 we examine the reporting of SNAP benefit amounts received. To this end, we separate out "true positive" person-months of reporting. These are months in which the Texas administrative report shows receipt of SNAP benefits and the respondent reports receipt of benefits in the SIPP. There are 13, 832 person-months available for analysis that met these criteria. The average monthly benefit reported in the SIPP for these person-months was $279, while the average monthly benefit recorded in the administrative record for these matched months was $289. About one fifth of these person-months showed accurate reporting of SNAP benefit amounts, and an additional 52.5% of person-months had an absolute difference in amounts that was within 25% of the amount in the administrative record.

Table 4 presents the logit regression coefficients estimates, which examines factors associated with under-reporting. For easier interpretation of logit coefficients, we calculate the predicted probability of under-reporting for each covariate. The logistic regression result shows that Hispanic, never-married, less than high school graduate, and female householder populations were more likely to under-report than their counterparts. For example, Hispanics are more likely to under-report their SNAP receipt than non-Hispanics by 20 percentage point, holding other factors constant.

Implications and Extension of research

This research is meaningful in that it uses data improved by record linkage. This leads to more accurate data measurement and analysis, and research findings for SNAP participation, using socio-economic information missing in administrative data.

The Census Bureau is in the process of getting state-level program administrative data from more states. One extension of current research is to increase the number of states' data matched to SIPP and to analyze the nationwide under-reporting rate. Another extension of this research is to match the administrative data to other survey data to look at under-reporting in other surveys such as CPS and ACS. This would further the goal of improving instrument design in a way that lessens under-reporting.

Table 1: Characteristics of sample by person-month (2004-2007)

	All Years		2004		2005		2006		2007	
	Total	Share of all person-months	Total	Share of all person-months	Total	Share of all person-months	Total	Share of all person-months	Total	Share of all person-months
Number of person-months=	170,176	100.0%	52,848	100.0%	50,352	100.0%	39,690	100.0%	27,286	100.0%
Age distribution										
15 or younger	42,958	25.2%	13,730	26.0%	12,973	25.8%	9,916	25.0%	6,339	23.2%
16-19 yrs. Old	9,400	5.5%	2,948	5.6%	2,667	5.3%	2,140	5.4%	1,645	6.0%
20-24 yrs. Old	10,490	6.2%	3,168	6.0%	3,007	6.0%	2,529	6.4%	1,786	6.5%
25-34 yrs. Old	21,879	12.9%	7,012	13.3%	6,687	13.3%	4,967	12.5%	3,213	11.8%
35-44 yrs. Old	23,564	13.8%	7,648	14.5%	6,851	13.6%	5,379	13.6%	3,686	13.5%
44-54 yrs. Old	24,761	14.6%	7,431	14.1%	7,292	14.5%	5,855	14.8%	4,183	15.3%
55-64 yrs. Old	17,705	10.4%	5,392	10.2%	5,241	10.4%	4,130	10.4%	2,942	10.8%
65 yrs. and up	19,419	11.4%	5,519	10.4%	5,634	11.2%	4,774	12.0%	3,492	12.8%
Sex										
Men	79,769	46.9%	24,658	46.7%	23,666	47.0%	18,700	47.1%	12,745	46.7%
Women	90,407	53.1%	28,190	53.3%	26,686	53.0%	20,990	52.9%	14,541	53.3%
Race										
White alone	143,477	84.3%	44,320	83.9%	42,321	84.1%	33,520	84.5%	23,316	85.5%
Black alone	19,298	11.3%	6,128	11.6%	5,775	11.5%	4,471	11.3%	2,924	10.7%
Asian alone	3,446	2.0%	1,189	2.2%	1,052	2.1%	744	1.9%	461	1.7%
All other races	3,955	2.3%	1,211	2.3%	1,204	2.4%	955	2.4%	585	2.1%
Ethnicity										
Hispanic	63,003	37.0%	18,535	35.1%	18,026	35.8%	15,140	38.1%	11,302	41.4%
Not Hispanic	107,173	63.0%	34,313	64.9%	32,326	64.2%	24,550	61.9%	15,984	58.6%
Marital status										
Married	74,790	43.9%	23,499	44.5%	22,146	44.0%	17,191	43.3%	11,954	43.8%
Divorced/Seperated/Widowed	23,865	14.0%	7,194	13.6%	6,964	13.8%	5,758	14.5%	3,949	14.5%
Never married	71,521	42.0%	22,155	41.9%	21,242	42.2%	16,741	42.2%	11,383	41.7%
Educational attainment*										
LT high school graduate	25,058	19.3%	8,059	20.2%	6,964	18.2%	5,767	18.8%	4,268	19.9%

High school grad, some college or AA	78,687	60.5%	23,509	59.1%	23,380	61.2%	18,762	61.3%	13,036	60.8%
BA degree or higher	26,283	20.2%	8,240	20.7%	7,855	20.6%	6,068	19.8%	4,120	19.2%
Work*										
had paid job during reference period	85,497	65.8%	26,494	66.6%	25,261	66.1%	19,946	65.2%	13,796	64.4%
did not have a paid job	44,531	34.2%	13,314	33.4%	12,938	33.9%	10,651	34.8%	7,628	35.6%
Disability**										
Had a work-limiting mental or physical disability	14,114	12.1%	4,370	12.1%	4,220	12.3%	3,378	12.3%	2,146	11.2%
Did not have a disability	102,720	87.9%	31,619	87.9%	30,103	87.7%	24,049	87.7%	16,949	88.8%

**for 15 and older, ** for 15 to 69 years old inclusive*

Source: U.S. Census Bureau, Survey of Income and Program Participation (SIPP), 2004 Panel Waves 1-12. For information on confidentiality protection, sampling and nonsampling error see <http:/www.census.gov/sip/source.html>

Table 2. Person-months of SNAP receipt in Texas Administrative records and SIPP, 2004-2007 (Weighted average across 2004 to 2007)

Ad Rec		SIPP		
		Did not report SNAP receipt in month X	Reported SNAP receipt in month X	Total
	Not received SNAP in month X	39,594	2,083	41,677
		85.62	4.50	90.13
row		95.00	5.00	100.00
		98.36	34.79	
	Received SNAP in month X	662	3,903	4,565
		1.43	8.44	0.10
row		14.50	85.50	100.00
		1.64	65.21	
	Total	40,256	5,986	46,242
		87.06	12.94	100.00

Source: U.S. Census Bureau, Survey of Income and Program Participation (SIPP), 2004 Panel Waves 1-12. For information on confidentiality protection, sampling and nonsampling error see <http:/www.census.gov/sip/source.html>

Table 3. Amounts received for true positive person-months (2004-2007)

	All Years Combined, 2004-2007	
	Total	Share of all "true positive" person-months
Number of person-months=	13,832	100.0%
Mean Monthly Benefit Amount reported in SIPP	$279	n/a
Mean Monthly Benefit Amount reported in Texas Administrative Record	$289	n/a
Ratio: |SIPP-Ad.Rec| / Ad. Rec		
no difference	2,793	20.4%
up to and including 25%	7,200	52.5%
25%, up to and including 50%	1,805	13.2%
50%, up to and including 100%	1,302	9.5%
greater than 100%	608	4.4%

Source: U.S. Census Bureau, Survey of Income and Program Participation (SIPP), 2004 Panel Waves 1-12. For information on confidentiality protection, sampling and nonsampling error see <http:/www.census.gov/sip/source.html>

Table 4. Probability of Under-reporting: Logit Results

Dependent Variable: Under-report or not (1/0)

Explanatory Variables	Coefficient Estimates
Black	-0.348***
	(0.008)
White	-0.298***
	(0.007)
Age	-0.014***
	(0.0004)
Age^2	-0.0005***
	(0.000)
High School graduate	-0.118***
	(0.005)
College and beyond	-0.713***
	(0.005)
Never married	-0.054***
	(0.006)
Divorced/separated/ widowed	0.410***
	(0.005)
Female	0.333***
	(0.003)
Female householder	0.777***
	(0.003)
Hispanic origin	0.696***
	(0.003)
Disability	0.864***
	(0.005)
Work history	-0.822***
	(0.004)
Citizenship	-0.372***
	(0.006)
Metropolitan Status	0.691***
	(0.005)

*** p<0.001

Note1: Base group includes other race, Hispanic, 35 years old, female, married, no disability, citizen, in work force, and in metropolitan area.

Note2: Heteroskedasticity-robust standard errors are in parentheses.

Source: U.S. Census Bureau, Survey of Income and Program Participation (SIPP), 2004 Panel Waves 1-12. For information on confidentiality protection, sampling and nonsampling error see <http:/www.census.gov/sip/source.html>