Longitudinal Employer - Household Dynamics

Technical paper No. TP-2000-01

Longitudinal analysis of SSN response on SIPP 1990-1993 panels

Date	:	September 2000
Prepared by	:	Lars Vilhuber and Robert Pedace
Contact	:	U.S. Census Bureau, LEHD Program FB 2138-3 4700 Silver Hill Rd. Suitland, MD 20233 USA

This document reports the results of research and analysis undertaken by the U.S. Census Bureau staff. It has undergone a Census Bureau review more limited in scope than that given to official Census Bureau publications. [This document is released to inform interested parties of ongoing research and to encourage discussion of work in progress.]

This research is a part of the U.S. Census Bureau's Longitudinal Employer-Household Dynamics Program (LEHD), which is partially supported by the National Science Foundation Grant SES-9978093 to Cornell University (Cornell Institute for Social and Economic Research), the National Institute on Aging, and the Alfred P. Sloan Foundation. The views expressed herein are attributable only to the author(s) and do not represent the views of the U.S. Census Bureau, its program sponsors or data providers. Some or all of the data used in this paper are confidential data from the LEHD Program. The U.S. Census Bureau is preparing to support external researchers' use of these data; please contact U.S. Census Bureau, LEHD Program, Demographic Surveys Division, FOB 3, Room 2138, 4700 Silver Hill Rd., Suitland, MD 20233, USA.

Longitudinal analysis of SSN response on SIPP 1990-1993 panels

Lars Vilhuber and Robert Pedace
lars.vilhuber@census.gov

Original version: September 27,2000
This version: July 7, 2003

Abstract

This document describes the analysis of the SIPP-SSN match quality, and the file resulting for that analysis as distributable to the Census RDCs.

Contents

List of Tables

1 Analysis of coverage

The analysis presented here focuses on the SSN coverage in the SIPP 1990-1993 core wave files. The first time an individual is interviewed for the SIPP, his or her name and address information is requested, as well as the SSN. If the respondent did not refuse to provide the SSN, his or her information is sent to SSA to be validated, i.e. whether the provided SSN is an actual and valid SSN. If the respondent did not provide an SSN ("don't know" and "none"), the individual's information is clerically reviewed and an SSN assigned.

To analyze the coverage of the SIPP by this variable, which is crucial for files using data merged on by SSN, we construct two variables, derived from internal use variables on the U.S. Census internal SIPP files. The validity indicator (`first_ssnvald`) contains information on the validity of the first SSN provided to SIPP interviewers, which should be the one provided to the validation process. The SSN provision status (`first_ssnstat`) indicates whether a respondent provided an SSN, didn't know his or her SSN, refused an SSN, or provided none, as of the first time that respondent appeared in the SIPP. Starting with Table 1 on page 4, we cross-tabulate these two crucial variables.

It is the Census Bureau's policy to eliminate persons that refused to provide their SSN from the validation process, as evidenced by the tabulations in Tables 1 to 4.[1] However, our analysis revealed a limited number of cases where individuals' provision status (`ssnstat`) changed after the first occurrence. For instance, they may have provided an SSN at the first contact, and thus were validated, but show a refusal code on later waves. It is our understanding that the control card typically is not updated after first contact. It thus remains unclear whether changes in this variable represent data errors or changes in response. Changes have also been found in the respondent-provided SSN. We have documented these changes in Tables 8 on page 12 and 13 on page 17.

For analysis purposes, we also tabulate the value of the control card variable `SSNSTAT` the last time an individual was a respondent in the SIPP (Table 6 on page 10), and the value of the "best" (most favorable) value of `SSNSTAT` over the time an individual participated in the SIPP (Table 7 on page 11). The primary analysis tables 1 to 4 report the information for the first occurence of either variable, irrespective of subsequent changes.

Furthermore, it seems that not all individuals who did not refuse the SSN were sent to validation. In some preliminary analysis not reported here, we

[1] If an SSN was provided *and* the control card shows a "refusal" code for `SSNSTAT`, then the provided SSN takes precedence, and the refusal code is taken to be a coding error, and recoded to show "provided".

looked at the age distribution of this group, on the assumption that some age cutoff kept those individuals out of the validation process. However, no clear pattern emerged, and we have not pursued this avenue of research any further. In the variables in the analysis tables that follow, we have denoted these cases by "Not verified" (Row 4). Between 7.93 % and 10.55 % of all cases in the base SIPP files are affected.

Finally, we have noted a very limited number of cases where a respondent is listed on the validation crosswalk file, but the SIPP public-use identifier is NOT on the SIPP core data files. We have no explanation at this point, and this should be pursued further. These case are listed in Row 1 in Tables 1 to 4 (in 1990, 4 records. 1991, 31 records. 1992, 11 records. 1993, 26 records).

A simple coverage ratio analysis would concentrate on the ratio between `first_ssnvald = 1,2` (Rows 2 and 3 in Tables 1 through 4 and the total number of cases in the SIPP core files (the grand total denoted in the lower left. This ratio varies between 82 percent (1993) and 85 percent (1990). However, insofar as there may be differences between respondents who voluntarily provided an SSN, and those that did not without refusing, a more detailed breakdown may be needed. The files resulting from this analysis, described in a separate document, and available for research only at the U.S. Census, permit further analysis, after merging the variables created in this analysis to the core SIPP files.

Table 1: SSN Coverage SIPP 1990

Frequency
Percent
Row Pct
Col Pct

first_ssnvald	**first_ssnstat First provided SSN Status Indicator**					
First provided SSN Validity Indicator	Missing	Provided	Don't know	Refused to provide	None	Total
Verified: Invalid	2 0.00 0.04 50.00	383 0.55 7.81 0.83	3351 4.83 68.33 22.07	0 0.00 0.00 0.00	1168 1.68 23.82 30.91	4904 7.06
Verified: Valid as provided	0 0.00 0.00 0.00	43445 62.57 94.56 93.84	2204 3.17 4.80 14.52	0 0.00 0.00 0.00	293 0.42 0.64 7.75	45942 66.16
Verified: Valid as updated	2 0.00 0.02 50.00	2287 3.29 17.48 4.94	8799 12.67 67.25 57.95	0 0.00 0.00 0.00	1996 2.87 15.26 52.82	13084 18.84
Not verified	0 0.00 0.00 0.00	182 0.26 3.31 0.39	829 1.19 15.06 5.46	4173 6.01 75.79 100.00	322 0.46 5.85 8.52	5506 7.93
Total	4 0.01	46297 66.68	15183 21.87	4173 6.01	3779 5.44	69436 100.00

Table 2: SSN Coverage SIPP 1991

Frequency *Percent* *Row Pct* *Col Pct* **first_ssnvald**	**first_ssnstat First provided SSN Status Indicator**					
First provided SSN Validity Indicator	Missing	Provided	Don't know	Refused to provide	None	Total
Verified: Invalid	18 0.04 0.54 58.06	343 0.77 10.29 1.19	2351 5.29 70.52 22.44	0 0.00 0.00 0.00	622 1.40 18.66 32.75	3334 7.51
Verified: Valid as provided	0 0.00 0.00 0.00	27095 61.02 95.36 93.94	1203 2.71 4.23 11.48	0 0.00 0.00 0.00	115 0.26 0.40 6.06	28413 63.99
Verified: Valid as updated	13 0.03 0.16 41.94	1328 2.99 16.00 4.60	5912 13.31 71.21 56.42	0 0.00 0.00 0.00	1049 2.36 12.64 55.24	8302 18.70
Not verified	0 0.00 0.00 0.00	77 0.17 1.77 0.27	1013 2.28 23.26 9.67	3152 7.10 72.38 100.00	113 0.25 2.59 5.95	4355 9.81
Total	31 0.07	28843 64.96	10479 23.60	3152 7.10	1899 4.28	44404 100.00

Table 3: SSN Coverage SIPP 1992

Frequency
Percent
Row Pct
Col Pct

first_ssnvald	**first_ssnstat First provided SSN Status Indicator**					
First provided SSN Validity Indicator	Missing	Provided	Don't know	Refused to provide	None	Total
Verified: Invalid	6 0.01 0.12 54.55	430 0.69 8.57 1.08	3804 6.09 75.84 25.37	1 0.00 0.02 0.02	775 1.24 15.45 32.39	5016 8.04
Verified: Valid as provided	0 0.00 0.00 0.00	37358 59.85 95.67 93.76	1543 2.47 3.95 10.29	0 0.00 0.00 0.00	147 0.24 0.38 6.14	39048 62.55
Verified: Valid as updated	5 0.01 0.04 45.45	1962 3.14 15.08 4.92	9575 15.34 73.61 63.87	0 0.00 0.00 0.00	1466 2.35 11.27 61.26	13008 20.84
Not verified	0 0.00 0.00 0.00	93 0.15 1.74 0.23	70 0.11 1.31 0.47	5183 8.30 96.86 99.98	5 0.01 0.09 0.21	5351 8.57
Total	11 0.02	39843 63.83	14992 24.02	5184 8.30	2393 3.83	62423 100.00

Table 4: SSN Coverage SIPP 1993

Frequency
Percent
Row Pct
Col Pct

first_ssnvald	**first_ssnstat First provided SSN Status Indicator**					
First provided SSN Validity Indicator	Missing	Provided	Don't know	Refused to provide	None	Total
Verified: Invalid	8 0.01 0.17 30.77	420 0.67 8.89 1.08	3636 5.79 76.97 23.13	0 0.00 0.00 0.00	660 1.05 13.97 27.44	4724 7.53
Verified: Valid as provided	0 0.00 0.00 0.00	36327 57.89 96.18 93.74	1360 2.17 3.60 8.65	0 0.00 0.00 0.00	82 0.13 0.22 3.41	37769 60.19
Verified: Valid as updated	18 0.03 0.13 69.23	1890 3.01 13.86 4.88	10399 16.57 76.27 66.15	0 0.00 0.00 0.00	1328 2.12 9.74 55.22	13635 21.73
Not verified	0 0.00 0.00 0.00	116 0.18 1.75 0.30	326 0.52 4.93 2.07	5842 9.31 88.26 100.00	335 0.53 5.06 13.93	6619 10.55
Total	26 0.04	38753 61.76	15721 25.05	5842 9.31	2405 3.83	62747 100.00

2 Source data

At the source of this analysis lie the SIPP - SSN crosswalk files. These files came from the Income Surveys Programming Branch (Donna Riccini and Fuad Foty). For this analysis, SIPP files with internal use identifiers were used.

Furthermore, to establish the baseline for validation and coding success, SIPP internal use core wave files were used to gather information on the original provided SSN and status of such a provision.

All processing occurs on a cohort basis.

We use information on names, SSNs and response status for SSN, provided on the control cards for all SIPP waves. Thus, information can be available multiple times. The SSA validation process uses the first occurrence of SSN that is provided, but the SSN changes on the control card for some individuals, and we have compared the SSA-provided and verified SSN to all SSNs provided by a given individual.

Once merged with files from validation process, three new sets of variables are created. The vector *SSNSTAT* provides information on the response status for SSN (originally contained in variable *U_SSNDK*) from all waves on the internal SIPP files. *LAST_SSNSTAT* and *FIRST_SSNSTAT* provide the first and last response status for presence in the SIPP.

The vector *SSNVALD* provides information on the validity of the SSN *after* the validation process, and distinguishes whether any of the original information was valid, or if information was added by the validation process based on name, birthdate, etc. The variable *FIRST_SSNVALD* provides this information specifically for the first occurrence of an SSN, which should be the one sent to SSA for validation.

Finally, *SSNMULT* provides information on SSNs that show up on more than one SIPP ID (as opposed to multiple SSNs associated with a single SIPP ID). Research seems to indicate that the bulk of these individuals fell out of a couple waves and returned. All variables are described in more detail in a separate document.

Note that since the data quality of the 1990 internal use SIPP crosswalk file is suspect, so the analysis for 1990 relies solely on the public use crosswalk files.

Table 5: First provided SSN Status Indicator

Cohort	first_ssnstat	Frequency	Percent
	Missing	4	0.01
	Provided	46297	66.68
1990	Don't know	15183	21.87
	Refused to provide	4173	6.01
	None	3779	5.44
	Missing	31	0.07
	Provided	28843	64.96
1991	Don't know	10479	23.60
	Refused to provide	3152	7.10
	None	1899	4.28
	Missing	11	0.02
	Provided	39843	63.83
1992	Don't know	14992	24.02
	Refused to provide	5184	8.30
	None	2393	3.83
	Missing	26	0.04
	Provided	38753	61.76
1993	Don't know	15721	25.05
	Refused to provide	5842	9.31
	None	2405	3.83

Table 6: Last provided SSN Status Indicator

Cohort	last_ssnstat	Frequency	Percent
	Missing	4	0.01
	Provided	49530	71.33
1990	Don't know	12465	17.95
	Refused to provide	4216	6.07
	None	3221	4.64
	Missing	31	0.07
	Provided	30789	69.34
1991	Don't know	8753	19.71
	Refused to provide	3173	7.15
	None	1658	3.73
	Missing	11	0.02
	Provided	41937	67.18
1992	Don't know	13107	21.00
	Refused to provide	5274	8.45
	None	2094	3.35
	Missing	26	0.04
	Provided	40553	64.63
1993	Don't know	14023	22.35
	Refused to provide	5963	9.50
	None	2182	3.48

Table 7: Best provided SSN Status Indicator

Cohort	best_ssnstat	Frequency	Percent
	Missing	4	0.01
	Provided	50166	72.25
1990	Don't know	12744	18.35
	Refused to provide	3720	5.36
	None	2802	4.04
	Missing	31	0.07
	Provided	31122	70.09
1991	Don't know	8924	20.10
	Refused to provide	2868	6.46
	None	1459	3.29
	Missing	11	0.02
	Provided	42397	67.92
	Don't know	13359	21.40
1992	Refused to provide	4762	7.63
	None	1894	3.03
	Missing	26	0.04
	Provided	41066	65.45
	Don't know	14190	22.61
1993	Refused to provide	5451	8.69
	None	2014	3.21

Table 8: Change in SSN Status Indicator

Cohort	change_ssnstat	Frequency	Percent
	Missing	4	0.01
1990	No	65030	93.65
	Yes	4402	6.34
	Missing	31	0.07
1991	No	41856	94.26
	Yes	2517	5.67
	Missing	11	0.02
1992	No	59142	94.74
	Yes	3270	5.24
	Missing	26	0.04
1993	No	59804	95.31
	Yes	2917	4.65

Table 9: SSN Status Indicator

Cohort	ssnstat	Frequency	Percent
	Other	24933	35.91
	Provided	36248	52.20
1990	Don't know	5439	7.83
	Refused to provide	1809	2.61
	None	1007	1.45
	Other	15349	34.57
	Provided	22898	51.57
1991	Don't know	4127	9.29
	Refused to provide	1502	3.38
	None	528	1.19
	Other	33306	53.36
	Provided	23003	36.85
1992	Don't know	4071	6.52
	Refused to provide	1646	2.64
	None	397	0.64
	Other	22905	36.50
	Provided	30281	48.26
1993	Don't know	6258	9.97
	Refused to provide	2828	4.51
	None	475	0.76

NOTE: The table only denotes four types of such vectors: those for which the coding does not change across all waves. The remainder had codes that changed over the course of the SIPP follow-up period.

Table 10: First provided SSN Validity Indicator

Cohort	first_ssnvald	Frequency	Percent
1990	Verified: Invalid	4904	7.06
	Verified: Valid as provided	45942	66.16
	Verified: Valid as updated	13084	18.84
	Not verified	5506	7.93
1991	Verified: Invalid	3334	7.51
	Verified: Valid as provided	28413	63.99
	Verified: Valid as updated	8302	18.70
	Not verified	4355	9.81
1992	Verified: Invalid	5016	8.04
	Verified: Valid as provided	39048	62.55
	Verified: Valid as updated	13008	20.84
	Not verified	5351	8.57
1993	Verified: Invalid	4724	7.53
	Verified: Valid as provided	37769	60.19
	Verified: Valid as updated	13635	21.73
	Not verified	6619	10.55

Table 11: Last provided SSN Validity Indicator

Cohort	last_ssnvald	Frequency	Percent
	Verified: Invalid	4904	7.06
1990	Verified: Valid as provided	45252	65.17
	Verified: Valid as updated	13774	19.84
	Not verified	5506	7.93
	Verified: Invalid	3334	7.51
1991	Verified: Valid as provided	28095	63.27
	Verified: Valid as updated	8620	19.41
	Not verified	4355	9.81
	Verified: Invalid	5016	8.04
1992	Verified: Valid as provided	38615	61.86
	Verified: Valid as updated	13441	21.53
	Not verified	5351	8.57
	Verified: Invalid	4724	7.53
1993	Verified: Valid as provided	37401	59.61
	Verified: Valid as updated	14003	22.32
	Not verified	6619	10.55

Table 12: SSN Validity Indicator

Cohort	ssnvald	Frequency	Percent
	Other	14470	20.84
	Verified: Invalid	4904	7.06
1990	Verified: Valid as provided	31957	46.02
	Verified: Valid as updated	12599	18.14
	Not verified	5506	7.93
	Other	8422	18.97
	Verified: Invalid	3334	7.51
1991	Verified: Valid as provided	20270	45.65
	Verified: Valid as updated	8023	18.07
	Not verified	4355	9.81
	Other	19448	31.16
	Verified: Invalid	5016	8.04
1992	Verified: Valid as provided	20045	32.11
	Verified: Valid as updated	12563	20.13
	Not verified	5351	8.57
	Other	11514	18.35
	Verified: Invalid	4724	7.53
1993	Verified: Valid as provided	26698	42.55
	Verified: Valid as updated	13192	21.02
	Not verified	6619	10.55

Table 13: Change in SSN Indicator

Cohort	changessn	Frequency	Percent
	Missing	4	0.01
1990	No	66069	95.15
	Yes	3363	4.84
		31	0.07
1991	No	42299	95.26
	Yes	2074	4.67
1992		11	0.02
	No	59108	94.69
	Yes	3304	5.29
		26	0.04
1993	No	59649	95.06
	Yes	3072	4.90

Table 14: Multiple SSN Indicator

Cohort	ssnmult	Frequency	Percent
	N.a.	10410	14.99
1990	No	58566	84.35
	Yes	460	0.66
	N.a.	7689	17.32
1991	No	36264	81.67
	Yes	451	1.02
1992	N.a.	10367	16.61
	No	51320	82.21
	Yes	736	1.18
	N.a.	11343	18.08
1993	No	50767	80.91
	Yes	637	1.02