

SIPP User Notes

To: SIPP Users

From: Heather Boushey, economist, Center for Economic and Policy Research

RE: Set D: Labor Market and Employment Outcomes

Date: February 13, 2006

Set D contains data for analysis of labor force participation, employment, and earnings. In order to facilitate easier programming, Set D is organized by person-month-job because each SIPP respondent can provide information on up to two jobs per month. For person-month data, issue the code included in the “Clean Set D” program. Appendix 1 lists the variables included in Set D. Appendix 2 provides some information on the Census definitions of earnings and labor force status.

Earnings and Wages

Three different sources of monthly employment income are identified in the SIPP public use files: (1) wage and salary income, (2) self-employed earnings, and (3) other work arrangements. The variable *earnm* is total monthly earnings, not including earnings from own business. Self-employment income is included in the variable *selfinc* and income from other work arrangements (for the 1996 and 2001 panels only) is included in the variable *mooninc*.

Imputed earnings and wages

The U.S. Census Bureau imputes wages and earnings for some observations with missing data. The reasons for missing data for a single item is because respondents did not provide requested information, the interviewer failed to ask the question correctly, a respondent’s answer is inconsistent with related responses, or the interviewer makes an error in recording the data. Item nonresponse data are generally imputed for core items. (See *SIPP Users Guide*, chapter 4.)

Table 1 shows what proportion of those with earnings or hourly wages have these values imputed and the means and standard deviations for imputed and reported values. Although very few individuals have imputed earnings per month in the 1992 and 1993 panels, the percent difference in the means between those with imputed and without imputed earnings per month is quite large – upwards of 40 percent each year – whereas the difference is around 5 percent for all other years and for the wage data. Since so few individuals have these imputations, it makes sense to drop these values in estimations of earnings per month.

Calculated earnings and wages

CEPR calculates earnings per month and hourly wages for individuals who report one but not the other. The variables *earn_so* and *wage_so* indicate whether earnings and wages, respectively, are reported or calculated. The calculations include the hours worked per month, which is smoothed

over four- and five-week months (assuming that individuals are not varying their response to the question based on whether the month is four- or five-weeks long).

Table 2 shows that among those respondents who report both earnings per month and an hourly wage, there is a very close correlation between the actual and calculated values. The top panel of Table 2 compares the reported earnings per month and what it would be if CEPR had calculated earnings per month from the reported wages and hours worked per month, while the bottom panel does the same for hourly wages. In general, the differences are under 5 percent indicating that our method of calculating earnings per month and hourly wages is consistent with reported items.

Table 3 shows reported and calculated earnings per month and hourly wages among all respondents. We would expect that those who only report earnings per month are more likely than not to be salaried employees and should, therefore have higher average earnings than those who only report an hourly wage and who are more likely to be hourly employees. Table 3 confirms this hypothesis; those who report earnings per month have higher average earnings per month than those who have their earnings per month calculated from their hourly wage and hours worked per month. Similarly, those who report an hourly wage have lower average wages than those who have their hourly wage calculated from their earnings per month and hours worked per month.

Top-coding

To protect against the possibility that a user might recognize the identity of a SIPP respondent with very high income, income from every source is “topcoded” and each of the three earnings sources is topcoded separately. The topcodes are based on a limit of \$150,000 annual income. Thus, for each earnings source, monthly amounts over \$12,500 (one-twelfth of the \$150,000 annual benchmark) are topcoded if the total income from those sources from all 4 months in the wave is greater than \$50,000 (one-third of \$150,000). Individual monthly amounts above \$12,500 may occasionally be shown if the respondent's income varied considerably from month to month, as long as the average does not exceed \$12,500. For example, if a respondent's income from a single job were concentrated in only one of the four reference months, a figure as high as \$50,000 could be shown.

The U.S. Census Bureau cautions the user against trying to make much use of the occasional monthly figures above \$12,500, except in calculating aggregates or observing patterns across the 4-month period for a single individual, family, or household. Those units with higher monthly amounts shown are a biased sample of high-income units, more likely to include units with income from multiple sources than other units with equally high aggregate income, which comes from a single source. (See *SIPP User's Guide*, Ch. 10 and Appendix for information on topcoding.)

Topcoding has implications for analysis of wages in a number of ways. First, truncating the earnings and wage distributions at the topcode value will lower the mean and variance of the series relative to its true value. Further, the topcode values do not often change and thus will truncate increasing proportions of the distribution over time. This is likely to be especially the case with data from the late 1990s when incomes were rising relatively rapidly.

To adjust for these problems, “Clean Set D” includes programs to impute the mean above the topcode. The U.S. Census Bureau includes a formula for calculating the mean above the topcode for the 1996 panel. For the other panels, and for comparison to the Census-supplied means, we include programs that generate the mean above the topcode by estimating the distribution above the mean with a lognormal and a Pareto distribution.¹ Tables 4 and 5 shows the proportion of respondents with earnings and wages hitting the topcode and the estimated means above the topcode with these different techniques. In Table 4, we also include the mean above the topcode provided by the U.S. Census, which is much larger than the means estimated by the lognormal or Pareto distributions.

Trimming

Note that earnings and wages are untrimmed.

¹ John Schmitt, “Creating a consistent hourly wage series from the Current Population Survey’s Outgoing Rotation Group, 1979-2002,” unpublished.

Table 1. Census imputed earnings and wages (nominal, no trimming or top-code adjustments)

Earnings per month							
Panel	Year	Percent Imputed	Not Imputed		Imputed		Percent difference in means
			mean	Sd	mean	sd	
1992							
	1992	4.76	1863.65	1456.33	2785.89	1864.26	49.5%
	1993	5.20	1902.47	1457.71	2745.54	1895.44	44.3%
	1994	5.47	1937.39	1478.53	2744.10	1862.36	41.6%
1993							
	1993	5.31	1931.72	1490.40	2879.91	1911.24	49.1%
	1994	5.71	1966.63	1504.78	2848.59	1817.68	44.8%
	1995	5.63	2016.73	1517.70	2831.47	1876.87	40.4%
1996							
	1996	15.89	2209.03	2298.12	2185.62	2270.58	-1.1%
	1997	20.88	2303.55	2361.88	2183.58	2241.36	-5.2%
	1998	22.39	2426.82	2440.19	2286.34	2334.08	-5.8%
	1999	22.23	2537.81	2573.80	2388.83	2380.55	-5.9%
	2000	19.86	2496.15	2590.91	2439.58	2407.15	-2.3%
2001							
	2001	22.85	2710.13	2782.06	2595.21	2626.23	-4.2%
	2002	25.51	2806.32	2821.83	2745.50	2843.04	-2.2%
Hourly wage							
1992							
	1992	14.61	9.23	5.58	10.02	5.35	8.6%
	1993	16.83	9.39	5.59	9.96	5.28	6.1%
	1994	18.61	9.51	5.61	9.99	5.10	5.1%
1993							
	1993	15.84	9.40	5.55	10.20	5.03	8.5%
	1994	18.82	9.62	5.40	10.07	5.18	4.7%
	1995	19.77	9.91	5.45	10.03	4.93	1.2%
1996							
	1996	16.91	9.82	5.22	9.72	5.17	-1.0%
	1997	22.66	10.19	5.29	9.94	5.18	-2.4%
	1998	24.06	10.62	5.39	10.29	5.25	-3.1%
	1999	23.61	10.98	5.47	10.67	5.42	-2.8%
	2000	21.76	11.16	5.49	10.70	5.41	-4.2%
2001							
	2001	22.84	11.83	5.72	11.70	5.68	-1.2%
	2002	25.53	12.14	5.88	11.95	5.80	-1.6%

Note: Values are for those reporting earnings/wages and do not include those calculated by CEPR.

Table 2. CEPR calculated earnings and wages respondents reporting both earnings and wages

		Among those with both earnings and wages				
		Not calculated		Calculated		Percent difference in means
Panel	Year	mean	sd	mean	sd	
Earnings per month						
1992						
	1992	1446.70	1045.21	1482.34	1056.88	2.5%
	1993	1478.72	1043.46	1508.99	1075.94	2.0%
	1994	1508.55	1057.81	1536.59	1067.38	1.9%
1993						
	1993	1485.59	1052.54	1525.28	1066.71	2.7%
	1994	1533.22	1071.59	1564.61	1058.46	2.0%
	1995	1577.30	1081.38	1605.44	1057.90	1.8%
1996						
	1996	1618.06	1358.37	1600.76	1079.41	-1.1%
	1997	1683.86	1458.78	1646.74	1094.97	-2.2%
	1998	1765.24	1501.52	1712.26	1115.62	-3.0%
	1999	1839.34	1514.74	1768.58	1130.78	-3.8%
	2000	1761.64	1335.77	1777.37	1135.52	0.9%
2001						
	2001	1921.68	1594.75	1898.72	1196.50	-1.2%
	2002	2005.54	1670.35	1922.51	1216.60	-4.1%
Hourly wage						
1992						
	1992	9.39	5.56	9.16	5.92	-2.4%
	1993	9.52	5.53	9.34	5.77	-1.9%
	1994	9.64	5.53	9.47	5.86	-1.8%
1993						
	1993	9.56	5.46	9.32	5.73	-2.5%
	1994	9.75	5.34	9.56	5.83	-2.0%
	1995	9.98	5.35	9.82	6.00	-1.5%
1996						
	1996	10.05	5.25	10.58	13.33	5.2%
	1997	10.38	5.31	11.25	17.03	8.4%
	1998	10.78	5.40	11.76	17.70	9.1%
	1999	11.14	5.49	12.34	19.99	10.8%
	2000	11.27	5.50	11.64	14.58	3.3%
2001						
	2001	12.03	5.74	12.76	16.97	6.1%
	2002	12.32	5.88	13.59	18.34	10.3%

Note: Includes values imputed by the U.S. Census Bureau. Dollar values are nominal and do not include trimming or topcode adjustments.

Table 3. CEPR calculated earnings and wages among respondents reporting either wages or earnings

		All					
Panel	Year	Percent calculated	Not calculated		Calculated		Percent difference in means
			mean	sd	mean	sd	
Earnings per month							
1992							
	1992	1.53	1907.58	1491.30	956.69	917.04	-49.8%
	1993	1.37	1946.34	1495.44	992.97	991.88	-49.0%
	1994	1.38	1981.49	1513.20	989.45	991.03	-50.1%
1993							
	1993	1.39	1982.02	1530.49	990.44	918.26	-50.0%
	1994	1.34	2016.95	1538.02	1005.67	986.58	-50.1%
	1995	1.33	2062.63	1551.58	995.35	928.43	-51.7%
1996							
	1996	5.30	2205.31	2293.78	1140.47	960.03	-48.3%
	1997	5.84	2278.50	2337.73	1200.35	997.52	-47.3%
	1998	5.49	2395.37	2417.55	1251.35	1018.84	-47.8%
	1999	5.28	2504.68	2532.86	1324.87	1042.42	-47.1%
	2000	5.22	2484.92	2555.52	1388.31	1111.54	-44.1%
2001							
	2001	4.28	2683.87	2747.65	1357.38	1116.58	-49.4%
	2002	3.95	2790.80	2827.37	1326.72	1111.96	-52.5%
Hourly wage							
1992							
	1992	38.68	9.35	5.55	14.47	10.95	54.8%
	1993	38.19	9.49	5.54	14.82	12.70	56.2%
	1994	37.49	9.60	5.52	15.17	11.14	58.1%
1993							
	1993	38.68	9.53	5.48	15.17	14.03	59.2%
	1994	37.32	9.71	5.36	15.50	14.02	59.6%
	1995	36.77	9.93	5.35	15.89	12.29	60.0%
1996							
	1996	35.87	9.80	5.21	17.52	26.33	78.8%
	1997	36.07	10.13	5.27	18.21	24.72	79.8%
	1998	36.44	10.54	5.36	19.27	26.55	82.8%
	1999	36.81	10.91	5.46	20.23	28.74	85.5%
	2000	36.87	11.06	5.47	19.62	21.51	77.4%
2001							
	2001	36.75	11.80	5.71	21.71	28.91	83.9%
	2002	37.66	12.09	5.86	22.93	32.15	89.7%

Note: Includes values imputed by the U.S. Census Bureau. Dollar values are nominal and do not include trimming or topcode adjustments.

Table 4. Means above topcodes for earnings per month

Panel	Year	SIPP Topcode	Share of all monthly earners topcoded	SIPP imputed topcode*	Imputed mean above the SIPP topcode			
					lognormal	Pareto from:		
					70th	80th	90th	
1992								
	1992	8,333	0.7		10,438	12,494	12,067	11,675
	1993	8,333	0.7		10,333	12,365	11,958	11,548
	1994	8,333	0.8		10,358	12,345	11,916	11,470
1993								
	1993	8,333	0.9		10,628	12,777	12,355	11,966
	1994	8,333	0.9		10,538	12,656	12,175	11,773
	1995	8,333	1.0		10,558	12,736	12,364	11,947
1996								
	1996	12,500	0.4	29,475	15,824	19,639	19,099	18,760
	1997	12,500	0.4	31,059	15,763	19,462	18,945	18,505
	1998	12,500	0.5	32,805	15,791	19,499	18,920	18,393
	1999	12,500	0.5	34,384	15,916	19,698	19,133	18,589
	2000	12,500	0.5	35,686	15,896	19,725	19,076	18,531
2001								
	2001	12,500	0.7	29,565	15,943	19,939	19,324	18,520
	2002	12,500	0.8	31,145	15,908	19,905	19,273	18,889

Note: Dollar values are in 2002 dollars and do not include trimming. Values include those imputed by the U.S. Census or calculated by CEPR.

* The U.S. Census Bureau began supplying a formula for calculating the imputed mean above the topcode for earnings per month as of the 1996 panel. See *SIPP User's Guide*, Appendix B.

Table 5. Means above topcodes for wages

Panel	Year	SIPP Topcode	Share of all wage earners topcoded	Imputed mean above the SIPP topcode			
				lognormal	Pareto from:		
				70th	80th	90th	
1992							
	1992	30	2.7	36.77	44.48	42.71	41.70
	1993	30	2.8	36.75	44.40	42.29	41.24
	1994	30	2.9	36.97	44.69	42.76	41.32
1993							
	1993	30	2.8	36.91	44.15	42.51	41.34
	1994	30	3.0	36.98	44.14	42.35	41.12
	1995	30	3.2	37.06	43.98	42.38	41.09
1996							
	1996	30	4.0	37.33	47.70	45.68	44.54
	1997	30	4.4	37.35	47.48	45.42	44.39
	1998	30	5.1	37.25	47.85	45.54	45.27
	1999	30	5.9	37.55	49.45	47.02	45.25
	2000	30	5.6	37.59	48.07	45.55	42.01
2001							
	2001	30	6.4	37.83	48.29	44.53	39.21
	2002	30	7.2	37.90	48.15	45.42	37.42

Note: Dollar values are in 2002 dollars and do not include trimming. Values include those imputed by the U.S. Census or calculated by CEPR.

* The U.S. Census Bureau began supplying a formula for calculating the imputed mean above the topcode for earnings per month as of the 1996 panel. See *SIPP User's Guide*, Appendix B.

Appendix 1: Variable List**Variable name Label****Variables that appear in every set**

id	Unique ID
srefmon	Reference month
wave	Wave of data collection
age	Age in this month

Variables in this set

Abmsum	Allocation flags for business income (96 and 01)
class	Class of worker
earn	Original SIPP monthly earnings
earn_calc	Earnings variable calculated
earn_imp	Earnings imputed (if have earn)
earn_tc	Earnings topcoded
earnm	CEPR Earnings this month
earntc	Top coded earnings with mean above tc
earntc_7	Earnings per month w/ Pareto 70th topcode
earntc_8	Earnings per month w/ Pareto 80th topcode
earntc_9	Earnings per month w/ Pareto 90th topcode
earntc_1	Earnings per month w/ log normal topcode
emonth	Ending month for job
emp	With job all month
ersend	Reason stopped working
esr	LF: Employment status recode
estlemp	With same employer as previous month
eyear	Ending year for job
female	Female
flag_jobid_chan	Job ID changed: Stinson, pgm (90, 91, 92 and 93)
hasjob	Has a job this month
hours	hours per week
hours_m	Hours per month at that job
hrly	Paid by the hour this month
ind	Detailed industry
ind23	Major industries (23)
job	Job ID
jobid_revised	Revised job ID, Stinson, pgm (90, 91, 92 and 93)
jobno	Job number 1 or 2
mjobid	Merged with revised job id (90, 91, 92 and 93)
mooninc	Moonlighting income (96 and 01)
multjob	Multiple job holder
njobs	Number of jobs this month
occ	Detailed occupation
occ14	Major occupations (14)
prim	Primary job number
primjob	Primary job indicator
rate	Regularly hourly pay rate
rate_tc	Wages topcoded
ratetc_7	Wages per month w/ Pareto 70th topcode
ratetc_8	Wages per month w/ Pareto 80th topcode
ratetc_9	Wages per month w/ Pareto 90th topcode
ratetc_1	Wages per month w/ log normal topcode
rmwkjwb	LF: Number of weeks with a job in mont
rot	SU: Rotation of data collection
rwkesr2	Employment Status Recode for Week
selfinc	Self-employment income

SIPP User Notes: Set D: Labor Market and Employment Outcomes

selftc	Top-coded self-employment income with mean above topcode
smonth	Starting month for job
syear	Starting year for job
tempall	# of employees at worker's firm all locations (96 and 01)
tempsiz	# of employees at worker's location firm (96 and 01)
tpearn	PE: Total person's earned income for t
union	Union member
wage	CEPR Hourly rate this month
wage_calc	Wage variable calculated?
wage_imp	Wages imputed (if have rate)
wagetc	Top-coded wages w/ mean above wages \geq \$30 in CPS
wksper	LF: Number of weeks in this month
year	SU: Calendar year for this reference month
wage_calc	Wage variable calculated
wage_imp	Wages imputed (if have rate)

Note: variables are available in all panels except where noted

Appendix 2: SIPP Concepts

From the *SIPP Users Guide*:

Earnings. The concept of earnings used in the SIPP survey differs from that of the Bureau of Labor Statistics (BLS). The SIPP survey asks about actual earnings on up to two jobs per month. The self-employed are also asked about their business earnings, although these are not included in the wage and earnings variables. The BLS publishes quarterly averages for an earnings concept called “usual weekly earnings” for employed wage and salary workers, which differs from the SIPP concept since it is based on usual, not actual earnings, excludes the self-employed, and excludes earnings from secondary jobs.

Worked each week. Persons “worked each week” in a month if, for the entire month, they were “with a job” and not “absent without pay” from the job. In other words, a person worked each week in any month when they were (a) on the job the entire month, or (b) they received wages or a salary for all weeks in the month, whether they were on the job or not. Persons also worked each week if they were self-employed and spent time during each week of the month at or on behalf of the business or farm they owned, as long as they received or expected to receive profit or fees for their work. In the CPS, the concept at “work” includes those persons who spent at least 1 hour during the reference week at their job or business. In the CPS, however, “at work” does not include persons who were temporarily absent from their jobs during the entire reference week on paid vacation, sick leave, etc. In SIPP, “worked each week” does include persons on paid absences.

With a Job. Persons are classified “with a job” in a given month if they were 16 years old or over and, during the month, either (a) worked as paid employees or worked in their own business or profession or on their own farm or worked without pay in a family business or farm or (b) were temporarily absent from work either with or without pay. In general, the word “job” implies an arrangement for regular work for pay where payment is in cash wages or salaries, at piece rates, in tips, by commission, or in kind (meals, living quarters, supplies received). “Job” also includes self-employment at a business, professional practice, or farm. A business is defined as an activity that involves the use of machinery or equipment in which money has been invested or an activity requiring an office or “place of business” or an activity which requires advertising; payment may be in the form of profits or fees.