

EVALUATING THE ACCURACY OF 1993 SIPP EARNINGS THROUGH THE USE OF MATCHED SOCIAL SECURITY ADMINISTRATIVE DATA

Benjamin Bridges, Linda Del Bene, and Michael V. Leonesio
Social Security Administration, Washington, D.C. 20254

KEY WORDS: Annual earnings, matched data, reporting error

This paper examines the accuracy of calendar year (CY) 1993 Survey of Income and Program Participation (SIPP) data on wages and self-employment income through the use of exact-matched data from the Social Security Administration's (SSA) Detailed Earnings Records (DER).¹ Various SIPP-DER comparisons are presented.

The study's sample is drawn from the 1992 and 1993 panels of the SIPP. It includes only those persons who have survey information for the reference month of December 1993 and who (according to the survey) were aged 15 or more in the reference month. The analysis sample includes 72,515 persons.

1. Construction of earnings variables

At the person level we construct for CY 1993 two earnings amount variables—wage and salary income (WS) and self-employment income (SEI)—from each of the two data sources. This section discusses the construction of those variables.

1.1. DER earnings amounts

We construct the two DER variables for each person with a match to SSA administrative records. The information used to create the DER variables comes from forms and schedules used by the federal government in the administration of tax laws. For each wage earner there is a DER for each of their 1993 employers that contains wage, salary, and tip income information reported on Form W-2. For each person with 1993 SEI taxable under the Hospital Insurance (HI) program there is a single DER that contains the taxable Old-Age, Survivors, and Disability Insurance (OASDI) and HI SEI information reported on Schedule SE. There is also a DER that contains information reported on Form 4137 for tip income not reported to employers.²

Our measure of DER wage and salary income is the sum of (1) wage, salary, and tip income taxable under the federal individual income tax (excludes deferred WS)³ and (2) deferred wages (contributions to 401(k) plans, etc.). It is the sum over all the person's 1993 Detailed Earnings Records. DER WS includes some types of pay in-kind; such in-kind pay is not very common.

Because the DER does not give total SEI, we constructed total 1993 SEI from HI taxable SEI. For the great majority of persons, HI taxable SEI is net profit income after deducting one half of OASDHI self-employment tax; this net profit income is our measure of total SEI. The purpose of this tax deduction is to make the tax treatment of SEI comparable to that of WS.⁴ For those persons, federal law specifies that HI taxable SEI be computed as follows: (1) If total SEI is less than \$400, then HI taxable SEI = 0.⁵ (2) If total SEI is \$400 or more, then HI taxable SEI = minimum of (\$135,000 minus HI taxable WS) and total SEI. \$135,000 is the maximum amount of HI taxable earnings. Farmers and other persons with very low SEI can compute HI taxable SEI by alternative methods.

For persons with HI taxable SEI whose HI taxable earnings are less than \$135,000 (4,121 persons in our sample), we set total SEI equal to HI taxable SEI.⁶ For persons with HI taxable SEIs

of \$135,000 (56 persons in our sample), we set their total SEIs at \$135,000; in effect, their total SEIs are topcoded.

Total SEIs of (1) persons with total SEIs less than \$400 and of (2) persons with HI taxable WSs of \$135,000 do not produce any HI taxable SEI. We treat their total SEI as zero.⁷ Of course, if a person with a match to SSA records has no 1993 Detailed Earnings Records, both of our constructed DER earnings variables are set equal to zero.

This study and one by Marc Roemer are the first to use DER data to evaluate the accuracy of SIPP earnings data. Roemer examines SIPP wage data.⁸ Recent studies by Roberto Pedace and Nancy Bates use the SSA's Summary Earnings Record (SER) data to examine SIPP earnings (wage and self-employment income combined).⁹ The SER summarizes the taxable earnings information from the person's DERs. Earlier studies by John Coder and Charles Nelson use federal income tax return data to evaluate SIPP wage data.¹⁰

For such evaluations the DER data has a number of advantages over the other two administrative data sets. The SER contains only OASDI taxable earnings. Taxable earnings (1) exclude wages from noncovered jobs (many government jobs are not covered), (2) are top coded at the taxable maximum (\$57,600 for 1993), and (3) combine wage and self-employment income. In the Coder and Nelson studies for joint tax returns only earnings amounts for couples were available. Moreover, deferred wages such as contributions to 401(k) plans are not reported on tax returns.

1.2. SIPP earnings amounts

For CY 1993 for each person we constructed two earnings income variables from the 1992 and 1993 public use longitudinal file monthly values. For each 4-month reference period the file gives monthly earnings amounts for up to two jobs (ern-amt1 and ern-amt2) and for up to two businesses (se-amt1 and se-amt2). The file also gives monthly values of incidental or casual earnings and of national guard or reserve pay. The SIPP does not record amounts of in-kind pay.

Our measure of SIPP wage and salary income is the sum of (1) employee wage and salary income from the main jobs (ern-amt), (2) earnings of the owner from their incorporated businesses (se-amt), (3) incidental or casual earnings (a g-amt variable), and (4) national guard or reserve pay (another g-amt variable).¹¹

Our measure of SIPP self-employment income is earnings of the owner from their unincorporated businesses (se-amt). In our analysis we treat SIPP SEI amounts of less than \$400 as zero; this is done to make the SIPP concept more comparable to the DER concept (see Section 1.1). SIPP SEI is not net profit or loss; it is the "income received" by an owner from their business during the period.

In determining whether earnings from a business are included in wage income or in SEI we use the form of business codes (corporation, sole proprietorship, or partnership) and the business ID number.

Some large SIPP earnings amounts are topcoded. The complex topcoding procedures are applied separately to each job or business; those procedures use amounts for single months, 4-

month periods (waves), and 12-month periods to determine whether monthly amounts in excess of \$8,333 should be topcoded at \$8,333. For a given job or business the maximum annual earnings amount shown on the SIPP public use longitudinal file for CY 1993 is usually \$99,996; in addition, wave earnings in excess of \$33,332 are often shown as \$33,332 or less.

2. Interpretations of comparisons of earnings amounts

The DER and SIPP WS measures are conceptually quite similar, but the administrative WS data are much more accurate than the survey WS data. The DER WS data contain some measurement error, particularly nonreporting and underreporting of income (for example, the underground economy). We interpret measures of the size of SIPP-DER wage differences as indicators of the accuracy of SIPP WS data--the larger the differences the less accurate are the SIPP data.

The DER and SIPP SEI measures are conceptually quite different. Also, the DER SEI data are less accurate than the DER WS data. Our measures of the size of SIPP-DER SEI differences provide useful pictures of such differences, but should not be interpreted as indicators of the accuracy of SIPP SEI data.

3. Analysis samples

The sample of persons age 15 or more who have survey interview information for December 1993 totals 72,840. Three hundred twenty-five of those persons were excluded from all the paper's analyses for any of three reasons: (1) the Summary Earnings Record showed positive 1993 taxable earnings but no 1993 Detailed Earnings Records were available (254 persons),¹² (2) the person had HI taxable earnings of \$135,000 and had both HI taxable WS and HI taxable SEI (20 persons),¹³ or (3) the person's administrative data exhibited obvious inconsistencies (51 persons).

The remaining sample includes persons for whom no match with SSA administrative data was made (nonmatch persons) and persons who have no SIPP interview information for one or more

months of CY 1993 (part-year interview persons). Most of this paper focuses on persons with matches and 12-month SIPP information. The paper does include brief looks at nonmatch persons and part-year persons.

4. Findings: matched full-year interview persons

First, we make some overall DER-SIPP comparisons for the matched full-year population. Then we turn to analyses of selected parts of this population.

4.1. Overall comparisons

Matches were found for 61,467 of the 68,794 full-year interview persons for a match rate of 89.3 percent.¹⁴ The match and nonmatch cases were rather similar in terms of gender mix, average age, and median earnings. A bit more on the characteristics of nonmatch persons appears in section 5.1.

Of the match cases 26.1 percent have neither DER nor SIPP earnings and 65.4 percent have both DER and SIPP earnings (Table 1). However, 3.2 percent of match cases have DER earnings but no SIPP earnings and 5.4 percent have SIPP earnings but no DER earnings.

For SIPP earners median SIPP earnings are \$16,936; for DER earners median DER earnings are \$17,639 (Table 2). For persons with both SIPP and DER earnings median SIPP earnings (\$18,182) amount to 97.6 percent of median DER earnings (\$18,632). Median SIPP earnings for the persons with SIPP earnings and no DER earnings are \$2,000; median DER earnings for the persons with DER earnings and no SIPP earnings are \$1,126.

4.2. Persons with wages only from both DER and SIPP

This paper focuses on the 35,362 persons with wages only from both DER and SIPP (the WS-WS group). There are several reasons for this emphasis. First, the group is large: it accounts for 88 percent of all persons with earnings from both data systems and 78 percent of all persons with earnings from at least one of the two data systems. Second, we want to focus on

DER earnings	All	SIPP earnings				
		No earnings	With earnings			
			Some	Wages only	SE only	Both sources
Number of persons						
Total	61,467	18,004	43,463	39,416	2,429	1,618
No earnings	19,333	16,043	3,290	2,473	664	153
With earnings	42,134	1,961	40,173	36,943	1,765	1,465
Wages only	38,059	1,743	36,316	35,362	195	759
SE only	2,311	190	2,121	486	1,433	202
Both sources	1,764	28	1,736	1,095	137	504
Percent of persons						
Total	100.0	29.3	70.7	64.1	4.0	2.6
No earnings	31.5	26.1	5.4	4.0	1.1	0.2
With earnings	68.5	3.2	65.4	60.1	2.9	2.4
Wages only	61.9	2.8	59.1	57.5	0.3	1.2
SE only	3.8	0.3	3.5	0.8	2.3	0.3
Both sources	2.9	0.0	2.8	1.8	0.2	0.8

Table 2.--Median SIPP and DER earnings by SIPP and DER reciprocity status: match persons

DER earnings	All	SIPP earnings				
		No earnings	With earnings			
			Some	Wages only	SE only	Both sources
Median SIPP earnings						
Total	\$8,710	\$0	\$16,936	\$16,894	\$15,400	\$21,091
No earnings	0	0	2,000	1,025	7,450	9,350
With earnings	17,160	0	18,182	18,000	19,000	22,079
Wages only	17,020	0	18,019	18,000	12,000	21,246
SE only	15,800	0	17,900	10,418	20,000	19,988
Both sources	21,959	0	22,164	21,827	18,519	24,321
Median DER earnings						
Total	\$7,533	\$0	\$16,768	\$17,611	\$5,867	\$13,508
No earnings	0	0	0	0	0	0
With earnings	17,639	1,126	18,632	19,128	10,678	16,069
Wages only	18,042	936	19,064	19,154	6,475	16,778
SE only	8,851	2,912	9,704	6,748	11,025	9,009
Both sources	21,494	5,838	21,955	25,152	12,516	19,563

persons with earnings from both data systems so that we can compare SIPP earnings amounts to DER amounts. Third, of the persons with earnings from both data systems, SIPP and DER amounts are more comparable for the WS-WS group than for any of the other groups. SIPP and DER definitions of wages are much more similar than are their definitions of self-employment income. Moreover, for some of the other groups the SIPP sources of earnings differ from the DER sources.

4.2.1. Whole group

For the WS-WS group, median SIPP wages (\$18,000) are 6 percent below median DER wages (\$19,154). When we compare the distribution of earners by SIPP wage classes with the distribution by DER wage classes, we find that the SIPP distribution is more concentrated in the \$5,000-29,999 classes (58.2 percent of persons versus 53.5 percent) and less concentrated in the \$30,000 and over classes (25.3 percent versus 29.6 percent).

We now present a number of comparisons for the \$1-99,995 DER wage classes. Comparisons for higher DER wage classes are not presented because for such classes SIPP topcoding is frequent enough to substantially distort SIPP-DER wage comparisons. For the \$1-99,995 class, mean SIPP wages fall 7.5 percent below mean DER wages; this is a measure of net SIPP reporting error. When we look at narrower wage classes in the \$1-99,995 range, the ratio of mean SIPP wages to mean DER wages decreases as DER wage income rises toward \$99,995 (Table 3, column 3). Mean SIPP wages exceed mean DER wages for the below-\$15,000 DER wage classes and fall short of them for the \$15,000-99,995 DER wage classes.

We use the mean of the absolute values of SIPP-DER wage differences as a percent of mean DER wages as a measure of gross SIPP reporting error. For the \$1-99,995 DER wage class this absolute value measure equals 18 percent. As DER WS levels increase within the \$1-99,995 interval, this measure of gross error first decreases and then increases (Table 3, column 5).

We now turn to more detailed person-level comparisons of SIPP wages with DER wages. Table 4 presents statistics on

relative earnings differences by DER wage classes. The relative earnings difference is (SIPP earnings – DER earnings) divided by DER earnings. A negative (positive) difference indicates that SIPP wages are understated (overstated). The percentage of persons with negative differences rises steadily from a low of 40.5 percent to a high of 85.9 percent for the \$75,000-99,995 class as DER wages increase. For the \$1-99,995 class, 62 percent have negative differences, 37 percent have positive differences, and half a percent have zero differences.¹⁵

For the \$1-99,995 DER wage class, 25 percent have relative differences of 5 percent or less, 42 percent have differences of 10 percent or less, and 71 percent have differences of 25 percent or less. As DER wages increase the percent of persons in a relative difference size class first increases and then decreases; these percentages peak in the \$20,000-24,999 wage class. This pattern of differences shows that the relative accuracy of SIPP wages is the best in the \$20,000-29,999 DER wage interval.

A number of factors contribute to these quite substantial earnings differences. SIPP’s use of subannual reporting of WS causes some increase in reporting errors. A sizable number of WS earners may report after-deduction or net WS instead of before-deduction or gross WS; subannual reporting likely increases the frequency of such errors. Topcoding slightly increases our measured earnings differences.

The following section shows that imputation errors cause a modest increase in earnings differences.

4.2.2. Imputed vs. not imputed¹⁶

Of the 35,362 persons in the WS-WS group, 17.7 percent (6,262) have imputed SIPP wage income; wages are partially imputed for 14.2 percent (5,015) and fully imputed for another 3.5 percent (1,247). Partial imputation results when only some of the positive monthly wage amounts (from jobs, businesses, etc.) included in CY 1993 WS are imputed.

For not-imputed earners median SIPP wages (\$17,619) are 6 percent below DER median wages (\$18,720); for imputed earners the shortfall of SIPP wages is 8 percent (\$19,241 vs. \$21,000). For partially-imputed earners the shortfall is 6 percent

Table 3.--Number of persons and various measures of earnings difference by DER earnings class: persons with wages only in both DER and SIPP

DER earnings	Number of Persons	SIPP-DER Earnings Difference ¹			
		SIPP-DER	<u>SIPP-DER</u> DER (%)	SIPP-DER	<u> SIPP-DER </u> DER (%)
\$1-2,499	3,424	\$1,188	100.8	\$1,539	130.6
\$2,500-4,999	2,554	780	21.0	1,788	48.1
\$5,000-7,499	2,193	698	11.2	2,101	33.8
\$7,500-9,999	2,092	476	5.5	2,285	26.2
\$10,000-14,999	4,079	193	1.6	2,486	20.0
\$15,000-19,999	3,996	-415	-2.4	2,851	16.3
\$20,000-24,999	3,610	-1,064	-4.7	3,403	15.2
\$25,000-29,999	2,947	-2,106	-7.7	4,048	14.8
\$30,000-39,999	4,309	-3,165	-9.2	5,198	15.0
\$40,000-49,999	2,697	-4,919	-11.1	6,870	15.5
\$50,000-74,999	2,331	-8,317	-14.0	10,626	17.9
\$75,000-99,995	583	-17,824	-21.0	18,885	22.3
\$1-99,995	34,815	-1,695	-7.5	4,069	18.1

¹(SIPP-DER), |SIPP-DER|, and DER denote earnings class means.

(\$18,664 vs. \$19,809), but for fully-imputed earners the shortfall is a substantially larger 16 percent (\$21,920 vs. \$26,190). When the fully-imputed group is compared with the not-imputed and partially-imputed groups, the fully-imputed group is older (mean age of 43 vs. 37 and 39) and more male (57 percent vs. 50 and 53 percent).

The percentage of persons with imputed wages is 12-13 percent for DER wages below \$5,000, 17-22 percent for DER wages of \$5,000-99,995, and 31 percent for DER wages in excess of \$99,995.

In the remainder of this section we focus on the \$1-99,995 DER wage classes. When we classify persons by DER wage class we find that the ratio of mean SIPP wages to mean DER

wages decreases as DER wage income rises toward \$99,995 for both imputed and not-imputed earners. For all but one of these wage classes this ratio is farther from 1 for imputed earners than for not-imputed earners.

For the \$1-99,995 DER wage class, the absolute value measure of gross reporting error is substantially higher for imputed earners than for not-imputed earners (28 percent vs. 16 percent). The comparable figures for fully-imputed earners and partially-imputed earners are 37 percent and 25 percent, respectively. For all 12 earnings classes in the \$1-99,995 interval this measure of gross reporting error is markedly larger for imputed earners than for not-imputed earners.

For imputed earners for the \$1-99,995 DER wage class 14.9

Table 4.--Percent of persons by relative earnings difference and DER earnings class: persons with wages only in both DER and SIPP

DER earnings	Relative earnings differences						
	Negative	Zero	Positive	-5 to +5 (%)	-10 to +10 (%)	-25 to +25 (%)	-50 to +50 (%)
\$1-2,499	40.5	1.6	57.9	8.6	15.6	32.9	53.0
\$2,500-4,999	49.3	0.2	50.5	11.4	23.7	50.7	73.9
\$5,000-7,499	51.1	0.3	48.6	18.1	33.0	61.4	82.3
\$7,500-9,999	53.5	0.2	46.3	20.0	35.3	67.3	87.3
\$10,000-14,999	56.1	0.4	43.6	25.5	45.6	77.3	91.6
\$15,000-19,999	61.8	0.5	37.7	29.8	50.6	81.1	94.8
\$20,000-24,999	66.4	0.4	33.2	31.7	51.9	83.0	96.1
\$25,000-29,999	71.7	0.2	28.0	30.4	50.6	82.8	95.9
\$30,000-39,999	72.7	0.3	27.0	29.5	50.0	82.1	96.0
\$40,000-49,999	76.2	0.2	23.6	29.3	48.8	79.9	95.7
\$50,000-74,999	78.6	0.3	21.2	27.6	45.5	74.4	92.9
\$75,000-99,995	85.9	0.0	14.1	23.8	40.7	64.8	87.8
\$1-99,995	62.2	0.5	37.3	24.5	42.0	71.3	87.8

percent have relative earnings differences of 5 percent or less, 28.4 percent have differences of 10 percent or less, and 58.1 percent have differences of 25 percent or less; the comparable figures for not-imputed earners are substantially larger—26.5 percent, 44.8 percent, and 74.0 percent, respectively. This pattern of differences is present for each of the 12 DER wage classes in the \$1-99,995 range.

The pattern of earnings differences shows that SIPP wages are, as expected, markedly less accurate for imputed earners than for not-imputed earners. As DER wages increase for both imputed and not-imputed earners the percent of persons in a relative difference size class first increases and then decreases; those percentages peak in the \$20,000-24,999 wage class.

4.3. Persons with self-employment income only from both DER and SIPP

For this group of 1,433 persons (the SEI-SEI group) median SIPP SEI (\$20,000) is 81 percent higher than median DER SEI (\$11,025). Recall that the SIPP and DER concepts of SEI are quite different. SIPP SEI is the amount a person draws from their business to support their family. DER SEI is the tax-law concept of net profit. That conceptual difference is presumably a major cause of the excess of SIPP SEI over DER SEI. For 1993 Roemer finds that aggregate SIPP SEI exceeds aggregate SEI from the Current Population Survey (CPS) by about 30 percent and that the CPS aggregate is greater than aggregate IRS SEI.¹⁷ The CPS uses a net profit SEI concept. Due to tax preferences and tax evasion, DER SEI is often considerably less than 'true' net profit from self-employment.

SIPP SEI is imputed for 29.3 percent of the SEI-SEI group. Median SIPP SEI as a percent of median DER SEI is the same percent for imputed earners and not-imputed earners.

When we compare the distribution of earners by SIPP SEI classes with the distribution by DER SEI classes, we find that the SIPP distribution is less concentrated in the \$400-9,999 classes and more concentrated in the \$20,000 and over classes. The pattern is markedly different from that for the WS-WS earners.

In the remainder of this section we focus on the \$1-39,999 DER SEI classes. Comparisons for higher SEI classes are not presented because for such classes SIPP topcoding is frequent enough to substantially distort SIPP-DER SEI comparisons. We find, as expected, that the relative differences for the SEI-SEI group are much larger than those for the WS-WS group. For the \$1-39,999 SEI class the mean of the absolute values of SIPP-DER SEI differences is very large, 118 percent of mean DER SEI. The comparable figure for the WS-WS group is 18 percent. For each of the 5 DER earnings classes in this range, this measure of gross differences is much larger for the SEI earners than for the WS earners.

4.4. Persons with DER earnings but no SIPP earnings

Among match cases 3.2 percent (1,961 persons) have DER earnings but no SIPP earnings. For this group, 55 percent of which is female, mean age is 42.25 years and median DER earnings are only \$1,126. For the much larger group with both DER and SIPP earnings (40,173), the percent female is lower (48 percent), mean age is lower (38.50), and median DER earnings are much higher (\$18,632).

For the group with DER earnings but no SIPP earnings, 67 percent have earnings below \$2,500; only 5 percent have earnings of \$20,000 or more. Eighty-nine percent have wages only and 10 percent have SEI only; for the group with both DER and SIPP earnings the comparable percentages are 90 and 5.

4.5. Persons with SIPP earnings but no DER earnings

Among match cases 5.4 percent (3,290) have SIPP earnings but no DER earnings. For this group median SIPP earnings are \$2,000, mean age is 39.11 years, and 52 percent are female.

Seventy-five percent (2,473 of the 3,290) have wages only. Their median SIPP wages are only \$1,025 and 61 percent have wages of less than \$2,500. Forty-seven percent of this wages-only group report that their only source of wage income is incidental or casual earnings; they do not receive any wage income from jobs (ern-amt) or from incorporated businesses. The comparable figure for the much larger group with SIPP wages only and some DER earnings (36,943 persons) is less than 2 percent. Many of those 2,473 persons may actually receive wage income which is not reported to the Internal Revenue Service.¹⁸

5. Additional findings

In this section we look at two groups that are not part of the matched full-year population.

5.1. Nonmatch full-year interview persons

DER matches were not found for 7,327 (10.7 percent) of the 68,794 full-year interview persons. The percent of nonmatch persons with SIPP earnings (68 percent) is similar to that for match persons (71 percent). Moreover, the percentage distribution of nonmatch earners by type of earnings (wages only, etc.) is similar to that for match earners. Median SIPP earnings for nonmatch earners (\$15,950) are slightly lower than that for match earners (\$16,936). For wage only earners the median earnings for the nonmatch and match groups are \$15,602 and \$16,894, respectively. The mean age of nonmatch earners (38.19 years) is very similar to that for match earners (38.55). In addition, the percent female for nonmatch earners (48) is the same as that for match earners.

We have seen that nonmatch persons are similar to match persons in a number of respects. This similarity does not hold for imputation status. For nonmatch and match earners the percents with at least some imputed earnings are 45 and 19, respectively.

5.2. Matched part-year interview persons

Only part-year interviews were obtained for 3.7 percent (2,363 persons) of the 63,830 match cases.¹⁹ For part-year persons, CY 1993 SIPP earnings are computed by annualizing their interview-month earnings. That is, for a person with x interview months during CY 1993 and total earnings for those months of \$y, CY 1993 SIPP earnings are set equal to \$y multiplied by 12/x. This simple method has been used by others.

The percents with earnings are a bit higher for part-year persons than for full-year persons. For part-year persons the percent with earnings in both data systems is 68 percent; the comparable figure for full-year match persons is 65 percent. The percent of persons whose DER and SIPP records agree as to the presence or absence of earnings is lower for part-year persons (83.1 percent) than for full-year persons (91.5 percent).

Compared to full-year persons with both SIPP and DER earnings, part-year persons with both SIPP and DER earnings have substantially lower earnings (SIPP median of \$15,008 vs. \$18,182 and DER median of \$13,499 vs. \$18,632) but have a substantially higher ratio of median SIPP earnings to median DER earnings (1.11 vs. .98). The percents with at least some imputed earnings for part-year earners are the same as those for full-year earners (19 percent).

Compared to full-year persons, part-year persons are on average much younger (33.53 years vs. 43.56 years) and slightly less female (50 percent vs. 53 percent).

Sixty-three percent (1,488) of these part-year persons have wages only from both DER and SIPP. We compare earnings differences for this group of WS-WS earners with those for full-year WS-WS persons with at least some imputed wages. For the \$1-99,995 DER wage class the absolute value measure of reporting error for this part-year group is 34 percent; the comparable figures for all imputed earners, partially-imputed earners, and fully-imputed earners are 28, 25, and 37 percent, respectively.

6. Concluding remarks

SIPP wage and salary data contain substantial measurement error. Imputed WS contain considerably more error than do not-imputed WS. For the WS-WS group with DER earnings under \$99,996, SIPP WS income is understated by 7.5 percent for the whole group, by 6 percent for not-imputed earners, and by 8 percent for imputed earners. The average relative error is 18 percent for the whole group, 16 percent for not-imputed earners, and 28 percent for imputed earners.

SEI earnings differences are very large. Much of the empirical differences result because the SIPP and DER definitions of SEI are conceptually quite dissimilar.

Three percent of persons have DER earnings but no SIPP earnings. On average their earnings are quite low (median of about \$1,100).

Five percent of persons have SIPP earnings but no DER earnings. On average their earnings are also low (median of \$2,000). For a sizable proportion of the group, incidental or casual earnings are their only source of earnings income.

Various steps could appreciably improve the quality of SIPP earnings data. The following very briefly mentions several of these steps: (1) It would be useful for the survey to ask for CY amounts of WS and of two SEI concepts (net profit and the current SIPP concept). (2) Analysts can use matched SIPP-DER data files to (a) evaluate and improve imputation procedures, (b) examine the extent to which earners report net rather gross WS, and (c) examine and compare measurement errors for population subgroups (for example, categorize by age or occupation).

Acknowledgements: We would like to thank Patrice Cole, Janice Olson, and especially Russell Hudson and Denton Vaughan.

¹ The matching and processing of these data were done in strict accordance with relevant laws and regulations regarding data confidentiality.

² Our DER data file is an edited abstract created in 2001 from the detailed earnings information contained in SSA's Master Earnings File (MEF). The MEF also contains a summary segment or Summary Earnings Record (SER) that summarizes some of the detailed information. Our data file also contains some SER information. SSA's primary use of the MEF information is in determining program benefits; SSA also uses it to compute the annual average wage index that is used to adjust benefit formulas and taxable limits under the program.

³ In some cases the wage income taxable under the income tax was not properly recorded in the MEF and had to be estimated by SSA staff in Baltimore based primarily on properly-recorded information on wages taxable under the OASDI and HI programs.

⁴ See the brief discussion in Social Security Administration, *Annual Statistical Supplement to the Social Security Bulletin* (Washington, D.C.: U.S. Government Printing Office, 2001), page 12.

⁵ For these cases total SEI is net profit income after deducting one half of potential OASDI self-employment taxes; self-employment taxes are not actually paid on such low SEIs.

⁶ The authors did not have access to information on HI taxable wages. We estimated HI taxable wages using DER information on total wages and OASDI taxable wages and SIPP information on age and class of worker of longest job. Except for the relatively small number of jobs covered under HI but not under OASDI our estimates of HI taxable wages are very reliable.

⁷ A very small number of persons (14 in our analysis sample) with HI taxable wages of \$135,000 have SIPP SEI larger than \$400; comparisons of the SIPP and DER earnings variables of these persons suggest that for most of these persons these SIPP SEI amounts appear as wages on their DERs.

⁸ Marc Roemer, "Using Administrative Earnings Records to Assess Wage Data Quality in the March Current Population Survey (CPS) and the Survey of Income and Program Participation (SIPP)," these proceedings.

⁹ Roberto Pedace and Nancy Bates, "Using Administrative Records to Assess Earnings Reporting Error in the Survey of Income and Program Participation," *Journal of Economic and Social Measurement*, 26 (Fall-Winter 2000): 173-192, and Nancy Bates and Robert Pedace, "Reported Earnings in the Survey of Income and Program Participation: Building an Instrument to Target Those Likely to Misreport," presented at the annual meeting of the American Association for Public Opinion Research, Portland, Oregon, May 2000.

¹⁰ John F. Coder, "Using Administrative Record Information to Evaluate the Quality of the Income Data Collected in the Survey of Income and Program Participation," *Proceedings of Statistics Canada Symposium 92: Design and Analysis of Longitudinal Surveys* (November 1992), and Charles T. Nelson, "The Quality of Census Bureau Survey Data among Respondents with High Incomes," *Proceedings of the American Statistical Association, Social Statistics Section* (1993).

¹¹ For a minority of persons for a given business ID number the incorporation status as indicated by the form of business code changes at least once during CY 1993; these changes are often likely to be the result of survey errors. For these status change cases we assign all CY 1993 earnings to the earnings category (wage or SEI) that according to the form of business code accounts for the majority of the year's earnings from the business.

¹² This reflects a problem with our matched file and not with SSA's Master Earnings File.

¹³ Their total SEI is only partially included in their HI taxable SEI.

¹⁴ SSA earnings data are exact-matched to SIPP persons using Social Security Numbers (SSNs). SSA validates SSNs collected in the SIPP for accuracy by using name, date of birth, race, and gender. SSA also searches for SSNs of persons who did not report an SSN in the survey (provided they did not refuse to report) and persons who reported incorrect SSNs in the survey by using name, date of birth, race and gender information from the SIPP. Of course this matching process is not error-free.

¹⁵ The term zero difference is used to indicate that SIPP earnings are within \$1.49 of DER earnings.

¹⁶ Many of the numbers used in this text section and in sections 4.3, 4.4, 4.5, and 5 do not appear in the paper's tables.

¹⁷ Marc Roemer, "Assessing the Quality of the March Current Population Survey and Survey of Income and Program Participation Income Estimates, 1990-1996," U. S. Census Bureau, June 16, 2000, <<http://www.census.gov/hhes/www/income.html>>.

¹⁸ Because of problems with employer wage reports, SSA is not able to match about 1.5 percent of the wage items it receives to worker records. These unmatched items account for about 0.5 percent of taxable wages and almost half of them are from three industries—agriculture, services, and bars and restaurants.

¹⁹ The nonmatch rate for part-year interview persons is considerably higher (36.5 percent) than for full-year interview persons (10.7 percent).