LONGITUDINAL EMPLOYER - HOUSEHOLD DYNAMICS

TECHNICAL PAPER No. TP-2002-22

Using Administrative Earnings Records to Assess Wage Data Quality in the March Current Population Survey and the Survey of Income and Program Participation

Date   : November 2002
Prepared by   : Marc Roemer
Contact   : Ronald Prevost (Ronald.C.Prevost@census.gov)

U.S. Census Bureau, LEHD Program
FB 2138-3
4700 Silver Hill Rd.
Suitland, MD 20233 USA

**Using Administrative Earnings Records
to Assess
Wage Data Quality
in the
March Current Population Survey
and the
Survey of Income and Program Participation**

November 19, 2002

Marc Roemer
Longitudinal Employer-Household Dynamics Program
Demographic Surveys Division
U.S. Bureau of the Census
marc.i.roemer@census.gov

Abstract: The March Current Population Survey (CPS) and the Survey of Income and Program
Participation (SIPP) produce different aggregates and distributions of annual wages. An excess of
high wages and shortage of low wages occurs in the March CPS. SIPP shows the opposite, an
excess of low wages and shortage of high wages. Exactly-matched Detailed Earnings Records
(DER) from the Social Security Administration allow comparing March CPS and SIPP people's
wages using data independent of the surveys. Findings include the following. March CPS and
SIPP people differ little in their true wage characteristics. March CPS and SIPP represent a
worker's percentile rank better than the dollar amount of wages. Workers with one job and low
work effort have underestimated March CPS wages. March CPS has a higher level of
"underground" wages than SIPP, and increasingly so in the 1990s. March CPS has a higher level
of self-employment income "misclassified" as wages than SIPP, and increasingly so in the 1990s.
These trends may explain one-third of March CPS's 6-percentage-point increase in aggregate
wages relative to independent estimates from 1993 to 1995. Finally, the paper delineates March
CPS occupations disproportionately likely to be absent from the administrative data entirely or to
"misclassify" self-employment income as wages.

Summary of Findings

The March Current Population Survey (CPS) and the Survey of Income and Program Participation (SIPP) produce different aggregates and size distributions of annual wages. March CPS shows higher aggregate wages than SIPP. Compared to each other, an excess of high wages and shortage of low wages occurs in the March CPS. SIPP has an excess of low wages and shortage of high wages. SIPP's differing aggregates and size distributions may result from collecting employer-specific sub-annual amounts. Sample bias, data quality, worker characteristics, or other phenomena could also contribute to these differences.

Independent data show nearly identical wage distributions of CPS and SIPP workers. Exactly-matched Detailed Earnings Records (DER) from the Social Security Administration contain data on March CPS and SIPP workers' wages. These data on March CPS and SIPP workers are for the most part indistinguishable by wage size. March CPS and SIPP workers don't seem to differ in their true wage characteristics. Sample bias doesn't appear to be operating.

The accuracy of the wage data must account for some of the difference between distributions. March CPS produces a greater proportion of small and large discrepancies from administrative data than SIPP, while SIPP produces more medium-sized discrepancies, and fewer discrepancies in receipt of wages. SIPP's discrepancies are errors of understatement more often than March CPS's, shifting SIPP's wage distribution toward smaller wage amounts. Both surveys identify a wage earner's relative rank better than the dollar amount.

Worker characteristics affect data accuracy in March CPS. Low March CPS work effort relates to under-estimated wages, except among workers with more than 1 job. Year-round part-time workers, unlikely to receive wages as salary, show a ratio of March CPS to DER wages similar to full-time workers. Possibly, this finding constitutes evidence against the theory that respondents overstate annual wages by reporting current salary as last year's wages.

March CPS overestimates wages relative to administrative data. Part of the overestimation comes from wages accruing to workers in the March CPS but not in the DER. Wages missing from the DER increase during the 1990 to 1996 period, during which the computerized CPS questionnaire began. Some of the excess may result from employers paying wages underground, or contractors' reporting their pay as self-employment income in the DER but wages in the March CPS.

Background

Aggregate wages from the March CPS are consistently higher than those from the SIPP (Roemer, 2000) and increase 6 percentage points from 1990 to 1996 relative to independent estimates, from 96% to 102%. About 2 percentage points of this increase come from increasing the "topcode" on wages, the Census Bureau's limit on the size of wages. During the same period, aggregate SIPP wages remain at nearly the same level relative to independent estimates, from 90% to 91%.

After describing the data sources and methodology, this paper demonstrates how the March CPS and the SIPP produce different size distributions of wages. Second, it explains why sample bias is a less important factor driving the difference than data accuracy. Next, the analysis shows how characteristics of workers and changes in data collection in the March CPS relate to data accuracy. Focusing on March CPS's higher and increasing level of wages missing from the DER, the investigation offers several explanations such as changes in the CPS measurement environment, the surveys' differing levels of scrutiny, and changes in the world. Finally, the paper shows how this "underground" and "misclassified" pay appears in all size classes, and delineates occupations disproportionately lacking DER wages.

The analysis focuses more attention on the March CPS for 3 reasons. First, the March CPS is the source of the Census Bureau's official income and poverty estimates. Second, it increases 6 percentage points in aggregate wages relative to independent estimates while SIPP remains at nearly the same level. Third, the most compelling explanation of SIPP's underestimation is respondents reporting take-home pay.[1] Data on deductions, necessary to investigate the take-home effect properly, are entirely lacking.

March CPS, SIPP, and DER data collection. The Basic Current Population Survey (CPS) takes place every month. It focuses on current employment status. Every March, a supplementary questionnaire gathers information about income received during the previous calendar year. The March CPS interviewed people in approximately 60,000 households from 1991 until 1996, when the sample size decreased to 50,000 households. A new sample design based on the 1990 decennial census began in March 1994 and a computerized instrument replaced the paper questionnaire completely in March 1995. The survey asks for wages earned in the prior calendar year from the longest job and from other jobs. A few workers also receive "other income" which the Census Bureau may code as wages. The data in this investigation are from the March 1991, 1994, and 1997 CPS, representing wages earned in calendar years 1990, 1993, and 1996.

---

[1] Coder (1988) compares monthly wage data from the 1984 SIPP Panel to annual wage data collected the following year in a roundup interview. The Census Bureau has since abandoned this topical module. Fully-interviewed respondents with only 1 employer during the year report monthly wages summing to 6.8 percent less than the annual wages they report the following year. This suggests the 4-month reference period is the basis of SIPP's underestimation.

The Survey of Income and Program Participation (SIPP) is a longitudinal panel survey that focuses specifically on income rather than labor force participation. It uses a four-month reference period rather than a calendar year, and covers more detailed income sources than March CPS. In the 1990s, the SIPP sample size varied from 14,000 to 37,000 households. The Census Bureau automated the survey instrument beginning with the 1996 Panel, and added several new income sources. SIPP asks for wages earned from 2 employers. Wages in this paper also include severance pay, national guard or reserve pay, incidental or casual earnings, and moonlighting. The investigation includes SIPP people[2] in the 1990, 1993, and 1996 Panels who are present in December of the reference year.

The Detailed Earnings Records are an extract from the Social Security Administration's (SSA) Master Earnings File of nearly all workers in the United States. The data are annual records of the full amounts of wages (untopcoded) and self-employment paid to workers by each employer. An employer reports earnings to the SSA by Employer Identification Number (EIN)and the Social Security Number (SSN) of the worker. The SSA verifies the name and birth date of the record before entering the earnings data into the database. About 1 percent of the reported earnings records remains unverified and absent from the database supplying the DER. This analysis focuses primarily on DER wages, the sum of wages from all employers for each worker. Secondarily, the analysis uses the presence of DER self-employment income to detect "misclassification" of CPS earnings. Known reporting problems exist in the DER data, causing undercoverage of workers in private households, agriculture, and construction in particular. A special advantage of the DER data is that it's worker-specific.

The DER suits data quality research well. It contains full wage amounts, without any topcode applied. Unlike tax data used in many previous exact-match studies, the DER data include deferred wages, such as contributions to 401(k) plans, making the wage amounts more comparable to wages as defined in the Census Bureau surveys. Moreover, tax data is difficult to use and interpret because of joint tax returns, which contain the combined income of spouses.

For all SIPP Panels, and selected March CPS years, the Census Bureau collects SSNs of sampled people, and verifies these SSNs with the Social Security Administration. If a respondent doesn't know an SSN, or if it fails verification, the Census Bureau searches the SSA's database for it. If a respondent refuses to provide an SSN, the Census Bureau makes no attempt to find it. Validated SSNs link a sampled person to the DER.[3]

The universe for all of the analysis in this paper is survey people with an SSN. To make the March CPS and SIPP statistics comparable, and to account for missing SSNs, an adjusted survey

---

[2] The terms "SIPP people" and "March CPS people" refer to survey respondents, non-respondents, and sampled persons about whom proxy respondents provided information.

[3] SSNs are available for the following percentages of people 15 years or older: March CPS 1991, 87.0; 1994, 80.5; 1997, 77.3. SIPP 1990, 91.1; 1993, 87.9; 1996, 82.1.

weight applies. The inverse of the availability rate for SSNs by size of wages multiplied by the base weight produces an adjusted weight accounting for people missing SSNs.

Analysis

March CPS and SIPP wage distributions differ. See Figures A, B, and C, which show the sum of wages on the vertical in each wage range, with the size of wages increasing along the horizontal. The SIPP distribution sits well to the left of CPS (toward smaller wage size) every year. The ratio of aggregate March CPS to SIPP wages is 106.0% in 1990, 111.2% in 1993, and 112.8% in 1996. This could be evidence of sample bias.

Including noncontacts, refusals, and other nonparticipating eligible households, the rate of household noninterview varies.[4] More importantly, the characteristics of noninterviewed households may vary by survey. People are more likely to participate in a survey whose purpose is salient (Geuzinge et al., 1999). Workers may be more likely to participate in a labor force survey such as the CPS. Program participants, who are probably low wage earners, might be more likely to participate in a survey like SIPP, while high wage earners may refuse due to SIPP's high time costs.

Does sample bias cause the distributions to differ? The exact-matched Detailed Earnings Records (DER) form evidence to the contrary. Across the distribution of matched records, the CPS's and SIPP's DER data estimate wages similarly. See Figures D, E, and F. The effect of people without an SSN remains unknown, but judging by the matched DER data, one must conclude that the March CPS and SIPP people with an SSN are quite similar in their true wage characteristics. The ratio of aggregate CPS DER wages to SIPP DER wages is 99.0% in 1990, 98.8% in 1993, and 100.2% in 1996.

The distributional differences between the SSN-matched March CPS and SIPP must result from causes other than biased samples. At least, the DER data point us in another direction to explain the differences between the surveys. In particular, the reference period and worker characteristics may affect the nature of the data people report in a survey. For example, in the March CPS, salaried workers may report current annual salary (as of March the following reference year), inflating high wages. Asked for a sub-annual amount in SIPP, respondents might report take-home pay instead of gross pay despite prompts in the questionnaire for gross pay.

Data accuracy is a more important factor. Compared to the DER, the CPS has an excess of extremely high wages in March 1997 and a shortage of low wages every year. See Figures G, H, and I. In March 1997, low wages become somewhat depressed and extremely high wages become overstated. SIPP shows the opposite, an excess of low wages and a shortage of high wages every year. See Figures J, K, and L for SIPP. If the DER is the "truth," the difference between each

---

[4] Noninterview rates as percentages are as follows. March CPS: 1991, 5.0; 1994, 7.4; 1996, 9.2. SIPP: 1990 Wave 1, 7.3; 1993 Wave 1, 8.9; 1996 Wave 1, 8.0. See Tupek, 2001; U.S. Bureau of the Census, 1998; and U.S. Bureau of Labor Statistics, 2002.

survey and "truth" seems to account for a much larger share of the CPS-SIPP difference than than the discrepancy between CPS "truth" and SIPP "truth."

Factors besides response error surely contribute to discrepancies between the surveys and the DER. Errors could occur in the Census Bureau's editing, weighting procedures, or in matching the survey to the DER. Imputing for nonresponse may cause especially large discrepancies. Furthermore, the DER itself may contain unknown errors. Nevertheless, let's define "error" as the discrepancy between survey data and DER data.

Table 1 lists the frequency of people by size of "error." It shows the percent of people in each survey with annual wages within a specific percent of their matched DER amounts. In CPS, 18.6% of people have wages within 5% of their matched DER amount, while only 14.4% of SIPP people have similarly accurate annual wages. However, SIPP has fewer people (9.9%) with errors greater than 50%; in March CPS, 12.4% have such large errors. SIPP has a greater proportion of people (22.8%) with medium-sized errors (between 5% and 20% of their DER wages) than March CPS (17.7%).

Table 1. Distribution of Discrepancies Between Census Bureau Surveys and the 1996 DER by Size

Distance from DER wages to Census Bureau wages	March 1997 CPS (n=77,228)	SIPP 1996 Panel (n=55,127)
	Percent distribution	
within 5%	18.6	14.4
6 to 10%	8.6	10.2
11 to 20%	9.1	12.6
21 to 50%	11.7	14.8
more than 50%	12.4	9.9
survey wages = 0	5.4	4.0
both = 0	29.1	29.9
der wages = 0	5.2	4.4
	100.0	100.0

It's no surprise that SIPP appears less accurate, because it never asks respondents to report annual wages. The analysis constructs annual wages from monthly data. Regarding recipiency, however, SIPP is more consistent with the DER than CPS is. Almost eleven percent (10.8%) of March CPS people have wages only in the survey or only in the DER. In contrast, only 8.4% of SIPP people have such discrepancies.

Focusing on "errors" in the dollar amount of wages, Table 2 shows the direction of the discrepancies among the subset of workers with wages in both the survey and the DER. As one would expect from the its lower aggregate, SIPP has a greater proportion of workers showing a shortfall from DER wages (less than 100 percent ratio) than CPS. This would be consistent with the theory that SIPP captures net rather than gross wages for many people. (The Census Bureau is currently redesigning the SIPP to include a more explicit check for take-home pay).

Table 2. Percent of People with Wages with Specified Ratio to Matched DER Data,
People with Wages in the Survey and the DER, March 1997 CPS, SIPP Panel 1996, Income Year 1996

Survey/DER Ratio (%)	CPS (n=47,253)	SIPP (n=34,068)
Less than 75	15.1	17.4
75 to 79	2.8	4.4
80 to 84	3.7	5.6
85 to 89	5.0	7.6
90 to 94	8.0	9.9
95 to 99	14.4	12.1
Less than 100	**48.9**	**57.0**
100	**7.8**	**2.9**
More than 100	**43.4**	**40.2**
101 to 105	10.2	9.9
106 to 110	5.5	6.0
111 to 115	3.4	3.6
116 to 120	2.6	2.7
121 to 125	2.0	2.0
More than 125	19.8	16.0
	100.0	**100.0**

Besides showing the relative merits of each survey, this tabulation also shows that comparing a survey's aggregate wages to independent estimates may be an overly simplistic method of evaluating data quality. Despite its aggregate's shortfall from benchmark wages, some SIPP people's wages are overstated. Despite exceeding benchmark wages, March CPS has many workers with understated wages. Discrepancies occur in both directions, and later this paper will show that the direction in the March CPS depends partly on worker characteristics.

The ratios above are quite sensitive to the level of wages - a $1,000 absolute error, though small, is a 50 percent error if the DER amount is $2,000. Comparing a wage earner's relative position in the survey and DER distributions may be more informative. The $1,000/$2,000 case would have a small percentile rank difference, because both amounts sit very low in the size distribution.

Table 3 lists the frequency of people by difference in percentile rank. Eleven and one-half percent (11.5%) of CPS people have exactly the same rank in the CPS and DER. Fewer SIPP people have exactly the same rank in the DER (9.2%), but SIPP also misplaces fewer people than CPS by 25 percentile points or more.

Table 3. Frequency of Specified Distance Between Survey and DER Percentile Rank,
People with Wages in the Survey and the DER, March 1997 CPS, SIPP Panel 1996, Income Year 1996

Percentile Rank Distance	CPS (n=47,253)	SIPP (n=34,068)
0	11.5	9.2
1 to 5	51.3	53.8
6 to 10	11.6	12.5
11 to 15	7.4	7.8
16 to 20	4.4	4.3
21 to 25	3.0	2.5
more than 25	8.5	5.4

Both surveys measure percentile rank better than dollar amount. This suggests that analyses not requiring the explicit dollar amount may benefit from using the percentile rank to compare workers' wages.

CPS within 5% of DER:	32.4%
SIPP within 5% of DER:	24.9%
CPS rank within 5 of DER:	62.8%
SIPP rank within 5 of DER:	63.0%

Worker characteristics relate to data accuracy. One of the puzzles presented by the exactly-matched DER data is March CPS's apparent over-estimation of wages. Among workers with CPS or DER wages under $300,000 (the topcode applied to the Census Bureau's internal files in March 1991), the CPS/DER ratio is 102.6% in March 1991, 101.4% in March 1994, and 103.3% in March 1997. Some analysts (for example, Coder and Scoon-Rogers, 1996) have suggested this excess comes from respondents reporting the current year's salary. Someone's salary is likely to be higher in the current year than last year.

The current-salary hypothesis and work experience. Part-time workers are less likely to work for a salary than full-time workers. Comparing people by work experience (the distinctions between full-time and part-time, year-round and part-year) may provide some evidence relevant to the current-salary hypothesis. Work experience variables in CPS are only available for people who worked, so the universe for this comparison includes only workers with wages in the CPS and the DER.

Table 4 compares the ratio of CPS wages to DER wages for part-time and full-time workers. Among full-time workers, CPS wages are near 100 percent of DER wages each year. Year-round part-time workers also have a ratio near 100 percent. If part-timers seldom receive salaries, their ratios resembling full-time workers' constitutes evidence against the current-salary theory.

Table 4. CPS/DER Ratio and Distribution of Aggregate Dollars by Work Experience

Work Experience	CPS/DER Ratio (%)			Distribution of Aggregate Dollars (Mean)
	March 1991 (n=73,443)	March 1994 (n=63,868)	March 1997 (n=55,500)	
All with CPS or DER wages	97.7	102.0	104.7	X
CPS or DER wages, under $300K	102.6	101.4	103.3	X
CPS and DER wages, under $300K	100.7	99.5	100.2	100.0
Part-time	95.5	95.9	97.9	5.8
part-year	89.9	91.3	90.3	2.1
year-round	99.2	98.4	102.4	3.8
Full-time	100.8	99.6	100.2	94.2
part-year	99.8	100.8	99.2	10.2
year-round	101.0	99.5	100.3	83.9

This table suggests four other things. First, limiting the universe to people with wages in both the CPS and DER eliminates most of CPS's over-estimation. The excess wages must come in part from people having wages in the CPS but not in the DER, perhaps working in the informal economy or in industries under-covered by the DER. Second, the ratio is quite sensitive to high amounts of wages. A substantial amount of the excess must be high wages. Third, CPS under-estimates the wages of part-year, part-time workers. Fourth, full-time workers receive 94.2% of aggregate wages on average and therefore deserve special attention.

The level of wages. Table 5 focuses on full-time, year-round workers, listing the cumulative CPS/DER ratio by upper limit of wage size. Each row of the table includes any CPS or DER wages in the specified range. This tells a more complex story. Despite the figures in Table 4, some excess wages do exist among year-round, full-time workers with wages between $30,000 and $150,000. The excess dollars are simply offset by CPS missing wages at the low end.

Table 5. Cumulative Ratio of CPS Wages to DER Wages, by Size of Annual Wages, Year-Round Full-Time Workers, March 1991, 1994, and 1997, Workers with CPS and DER Wages

Upper Limit of Wage Size	Cumulative Ratio of CPS to DER (%)		
	1991 (n=44,043)	1994 (n=38,703)	1997 (n=33,847)
4,999	39.6	35.5	24.4
9,999	76.6	69.7	57.0
14,999	93.6	89.5	85.3
19,999	96.0	93.8	91.2
24,999	99.5	99.0	96.6
29,999	99.8	98.8	98.3
49,999	100.9	100.5	101.2
74,999	101.5	101.8	103.3
99,999	101.3	101.8	102.4
149,999	101.5	101.2	102.6
174,999	100.8	100.2	101.0
199,999	100.7	99.7	100.5
299,999	101.0	99.5	100.3

The excess at the high end and the deficit at the low end both increase from March 1991 to March 1997. Why might this happen?

Work effort: weeks and hours worked, and the number of jobs. Changes in March CPS data collection interacting with worker characteristics may affect the wage distribution. In the computerized instrument (phased in during 1994), respondents choose "the easiest way" to report earnings from the longest job: weekly, biweekly, monthly, quarterly, or annually. Most respondents report annual wages directly.[5] But if they report a sub-annual amount, the instrument calculates the annual amount and the respondent verifies it. Exaggerating the number of periods receiving pay might shift the CPS distribution further toward higher wages.

---

[5] Respondents may choose weekly, every other week, twice a month, monthly, or yearly for reporting wages from the longest job. In March 1999, respondents chose "yearly" for 90.0% of household members earning wages from the longest job (n=47,440; DeNavas, 2002).

A rigorous test of this hypothesis would require the unedited CPS data, which contains the period the respondent chose for reporting earnings. However, the edited data do include the number of weeks and the usual number of hours worked at the longest job during the reference year, perhaps useful proxies. If people misreport the number of periods receiving pay, perhaps they similarly misreport weeks and usual hours. In this case, workers with low weeks or hours should have a low ratio of CPS wages to DER wages; those with high weeks or hours should have a high ratio.

Table 6 confirms that errors in work effort and amount of wages may occur in tandem at the low end. Among people with only one job, low work effort relates to under- estimated wages, but high work effort doesn't appear to relate to over-estimated wages. Having more than one job seems to be a bigger factor driving some workers' over-estimated wages, although for them the particular effect of work effort on the CPS/DER ratio is less clear.

Table 6. Ratio (%) of CPS Wages to DER Wages, by Number of Wage Sources, Weeks and Hours Worked, and Survey Year, People with CPS and DER Wages

	Any CPS wages			CPS wages only from the longest job			CPS wages from the longest job and other jobs		
	1991	1994	1997	1991	1994	1997	1991	1994	1997
				n=51,630	n=45,538	n=39,045	n=11,843	n=9,654	n=8,208
Weeks									
All	**96.2**	**99.9**	**100.8**	**94.5**	**98.7**	**98.9**	**107.5**	**107.8**	**113.7**
Part-year									
01 to 09	50.2	59.3	61.9	45.7	56.6	55.9	85.6	77.4	106.1
10 to 19	80.3	66.6	80.9	75.8	67.6	77.4	98.9	62.8	96.2
20 to 29	94.3	95.6	95.6	89.1	95.5	92.5	108.5	95.9	106.1
30 to 39	99.5	97.6	99.9	96.8	93.0	95.9	106.3	111.0	110.7
40 to 49	102.2	106.8	103.5	98.1	102.1	98.3	111.5	118.4	116.0
Year-round									
50,51,52	96.1	100.3	101.1	94.7	99.4	99.5	107.0	107.6	114.2
Hours									
All	**96.2**	**99.9**	**100.8**	**94.5**	**98.7**	**98.9**	**107.5**	**107.8**	**113.7**
Part-time									
01 to 08	82.5	79.7	87.9	81.8	75.5	73.7	85.1	101.1	138.9
09 to 16	85.7	82.4	89.9	82.3	83.9	88.8	95.9	78.4	94.2
17 to 24	94.1	89.3	99.3	91.0	87.7	96.3	106.4	96.9	112.3
25 to 34	98.8	101.8	101.9	95.8	98.6	99.9	110.0	114.4	109.2
Full-time									
35 to 40	97.6	100.0	101.0	96.3	99.0	99.2	106.8	107.8	115.1
41 to 50	94.6	98.7	101.5	92.6	97.8	100.5	108.0	104.3	109.0
51 & up	94.4	100.8	100.8	91.9	99.4	98.9	108.8	108.9	113.0

Note: The table excludes workers whose longest job was unincorporated self-employment. Weeks worked and usual hours worked refer only to the longest job. CPS and DER wages are totals for all jobs.

CPS tends to underestimate the wages of part-year workers who worked fewer than 20 weeks. The CPS/DER ratio approaches 100 percent as the number of weeks increases. The highest ratios are generally among people who worked 40 to 49 weeks. However, it's unclear whether the underestimated wages occur because of under-reported weeks, or directly from under-reported wages.

CPS also tends to underestimate the wages of people usually working 16 or fewer hours per week. Among all part-time workers, the CPS/DER ratio increases with the number of hours worked. Here also, it's unclear whether the underestimation comes from calculating wages from under-reported hours, or under-reporting wages directly. Among full-time workers (35 hours or more), there is no discernible pattern.

A low number of weeks worked or usual hours worked clearly affects people with only one job. Possibly, missing wages from other jobs depresses the ratios of one-job workers with low work effort. Generally higher ratios in March 1997 suggest that the computerized instrument improves data for everyone by work effort.

Excess wages appear among people with more than one job. These workers have much higher CPS/DER ratios (between 107.5% and 113.7%) than those with wages only from the longest job (between 94.5% and 98.9%). This suggests again that the CPS contains wages from the informal economy. Strangely, the ratio among workers with more than one job but minimal weeks or hours increases wildly from 85.6% to 106.1% by weeks, and from 85.1% to 138.9% by hours.

Some March CPS wages are missing from the DER. Table 4 showed excess CPS wages increasing from March 1991 to March 1997. This finding confirms a study comparing March CPS income to benchmarks derived from the national accounts (Roemer, 2000). The table also showed that workers without DER wages contribute substantially to March CPS overestimating wages. Table 6 also suggests that March CPS captures some activity missing from the DER by measuring wages from jobs besides the longest job. These findings motivate a closer look at the March CPS workers appearing to have "underground" wages.

An increasing amount of wages appears from a CPS "job" with no DER employer. Table 7 compares the percent distribution of March CPS wages in March 1991, 1994, and 1997, by number of DER employers (EINs) and CPS jobs. The percent of wages earned in the CPS but missing from the DER increases from 4.6% in March 1991 to 6.6% in March 1997. Mean CPS wages in this group increase by 59.5%, from $14,618 (March 1991) to $16,836 (March 1994) to $23,323 (March 1997). In contrast, people with wages missing from March CPS have mean DER wages between $7,401 and $8,806.

Table 7. Percent Distribution of March CPS Wages, by Number of DER Employers and CPS Jobs

Number of DER Employers	March 1991 CPS n=68,118			March 1994 CPS n=59,203			March 1997 CPS n=51,294		
	1 Job	2+ Jobs	Total	1 Job	2+ Jobs	Total	1 Job	2+ Jobs	Total
0	4.3	0.3	**4.6**	4.3	0.3	**4.7**	6.3	0.4	**6.6**
1	67.8	3.3	**71.1**	69.9	3.3	**73.2**	65.5	2.8	**68.3**
2	9.9	7.0	**16.9**	9.1	6.5	**15.5**	10.7	6.5	**17.2**
3+	3.1	4.2	**7.4**	2.9	3.7	**6.6**	3.6	4.2	**7.8**
Total	**85.1**	**14.9**	**100.0**	**86.3**	**13.7**	**100.0**	**86.1**	**13.9**	**100.0**

Between 20 percent and 25 percent of March CPS wage earners have a different number of DER employers than CPS jobs. This situation suggests a substantial challenge in collecting detailed information about "jobs." In the 90s, SIPP asked for wages paid by only 2 specific employers, surely missing wages from 3rd or further employers. About 7 percent of March CPS wages accrue to people with 3 or more DER employers. The Census Bureau is currently redesigning the SIPP to collect data on third and further jobs.

Wages of any size may be absent from the DER. Using a scale on the Y-axis 1/14 that of the earlier figures, Figure M shows that the March CPS increasingly contains wages missing from the DER, at all wage levels. It's especially apparent at the high end, where the income limits increased.

Wages may be absent from the DER for at least 2 reasons. First, an employer may fail altogether to report an employee's wages to the Social Security Administration, in a sense paying the worker "underground." Second, a worker may have self-employment income classified as wages in the Census Bureau survey. Presence of self-employment income instead of wages in the DER may be evidence of a classification difference.

Table 8 distinguishes "underground" wages from "misclassified" wages. The levels of underground activity and misclassification differ between CPS and SIPP every year, especially the percent of wages misclassified. In 1996, a sharp increase in both phenomena appears in March CPS while SIPP shows no substantial change. Increasing underground and misclassified wages only in March CPS calls to mind the interplay of 3 factors: a) real change in the level of underground activity; b) differences between the March CPS and SIPP measurement environments; and c) changes in the March CPS and SIPP measurement environments.

Table 8. Percent Distribution of March CPS and SIPP Wages and Workers, by Income Year and Presence of Wages or Self-employment Income in the DER

	Wages			Workers		
	1990	1993	1996	1990	1993	1996
March CPS Wage Earners	100.0	100.0	100.0	100.0	100.0	100.0
DER wages present; no self-employment	91.5	91.6	89.6	89.8	89.9	88.8
DER wages and self-employment present	3.9	3.8	3.7	3.6	3.5	3.5
No DER wages or self-employment (CPS underground)	2.5	2.4	**3.6**	5.0	5.0	**5.5**
No DER wages, self-employment present (CPS misclassified)	2.1	2.2	**3.0**	1.6	1.7	**2.2**
	1990	1993	1996	1990	1993	1996
SIPP Wage Earners	100.0	100.0	100.0	100.0	100.0	100.0
DER wages present; no self-employment	92.1	92.5	92.7	89.9	90.4	89.7
DER wages and self-employment present	4.3	4.1	4.1	3.9	3.9	3.8
No DER wages or self-employment (SIPP underground)	2.2	1.9	1.8	4.6	4.1	4.8
No DER wages, self-employment present (SIPP misclassified)	1.4	1.5	1.5	1.6	1.7	1.7

Note: SIPP sample counts (n) are 25,706 in 1990; 21,270 in 1993; and 36,469 in 1996. Table 7 contains sample counts for March CPS.

The CPS measurement environment changed during the period. The computerized survey instrument replaced the paper questionnaire beginning in 1994. The paper version was actually 2 questionnaires, the Basic CPS about last week's labor force participation, and the Supplement about last year's income. Perhaps the break after completing the first questionnaire gave households with "underground" wages an opportunity to end the interview, or to decide to hide income. The computerized questionnaire proceeds seamlessly from the Basic to the Supplement, allowing no such opportunity, and keeping the "foot in the door." Having just reported information about last week's work, respondents may be less apt to hide last year's earnings.

The SIPP instrument converted from paper to computer with the 1996 Panel, and no great increase in underground wages occurs. However, the Census Bureau introduces SIPP as an income survey collecting data for social scientists and policy-makers, while March CPS is a labor force survey measuring employment and unemployment. The context of the survey surely influences behavior, and respondents may be more forthcoming in CPS where the purpose is more salient, and when they've already complied with the more innocuous labor force portion.

The surveys' differing levels of scrutiny may be a factor. Collecting monthly and weekly data, SIPP scrutinizes people more closely than March CPS. One might speculate that asking for more detailed information could inhibit people from reporting underground wages.

On the other hand, the greater level of scrutiny may allow SIPP to classify workers more accurately. During 1990 to 1996, the ratio of March CPS to SIPP wage earners increases from 94.6% to 98.3%, and the March CPS/SIPP ratio of self-employed workers decreases from 89.6% to 70.5% (derived from Roemer, 2000). March CPS gains wage earners relative to SIPP and loses self-employed workers relative to SIPP. These changes occur while the SIPP misclassification rate in Table 8 remains nearly constant and the March CPS misclassification rate increases.

It's clear from the survey questions why March CPS would be more apt than SIPP to classify earnings differently from the DER. Answering "yes" to the March CPS question "Did ... work for pay?" and "no" to "Did ... own a business?" makes someone a wage earner. There are many situations in which a worker, self-employed for tax purposes, may say a "no" to owning a business.

Changes in the world could have caused the DER to miss more wages. Underground activity could have increased. If so, March CPS is more successful than SIPP at measuring it. More people working on contract may play a role as well. SIPP appears to classify their pay more consistently with the DER while March CPS counts it as wages.

Underground and misclassified pay appear in all size classes. Figures N and O distinguish underground and misclassified pay by size. March 1997 CPS is in blue, SIPP 1996 is in yellow. CPS has a much larger level of total wages missed by the DER. Underground wages tend to be

lower amounts (the mean in CPS is $17,765; in SIPP, $8,570). Misclassified self-employment income tends to be higher amounts (the mean in CPS is $37,262; SIPP $21,264).

Certain occupations in March CPS are more likely to be DER misses. Appendix A contains tabulations showing how March CPS occupations differ between workers lacking DER wages and those having DER wages. The tables compare the composition of "underground" and "misclassified" workers to legitimate workers by occupation. The difference in occupations underscores 2 key findings of this paper. First, the March CPS indeed measures economic activity invisible to administrative systems. Second, a non-trivial difference in the definition of "self-employment" exists between the March CPS and administrative systems.

"Underground." An underground worker is a CPS wage earner with no DER wages or self-employment income. Legitimate workers are those with wages in the CPS and the DER. Table 9 in Appendix A lists occupations appearing disproportionally underground relative to legitimate workers. The occupations in the table comprise 17.8% of all legitimate workers and 46.9% of all underground workers.

As expected, workers without DER data are more likely to work in private households, construction, and agriculture. Informal occupations such as street and door-to-door sales workers, dancers, and bartenders appear as well. Among the occupations most disproportionately underground are workers in private households: child care workers, cooks, cleaners and servants. Family child care providers, real estate salespeople, and news vendors also appear disproportionately. Some underground occupations such as timber cutting earn very high wages.

Apart from the relative likelihood of an occupation being in the underground group, the absolute number of workers may be important. Occupations with a high number of underground workers not mentioned already include farm workers, groundskeepers and garderners, painters, construction laborers, carpenters, janitors and cleaners, laborers, waiters and waitresses, and cooks. It's easy to imagine how such occupations might escape the Social Security Administration's record-keeping system. For example, some may work for short periods and their employers never report the wages paid.

"Misclassified." Table 10 in Appendix A shows occupations disproportionally having CPS earnings classified as wages and DER earnings classified as self-employment income. The occupations listed comprise 15.0% of legitimate workers and 56.7% of workers with misclassified pay. The occupations most disproportionately misclassified are real estate sales, barbers, fishers, social scientists (of unspecified type), dentists, and clergy. Other numerous occupations include truck drivers, sales supervisors and proprietors, hairdressers and cosmetologists, and carpenters.

There may be accounting idiosyncracies at work here. Commissions from real estate sales may follow peculiar accounting conventions, allowing workers to opt out of the Social Security system. Tax law can consider the clergy self-employed, although in the March CPS context they

clearly work for wages because they have an employer and don't own a business. In this case, "misclassification" is really "different classification."

Many occupations in Table 10 such as designers and construction laborers may actually be self-employed contractors. Some are high-paying occupations such as lawyers, insurance salespeople, financial services providers, management analysts, and physicians. Small business occupations such as barbers, roofers, and carpet installers appear as well. These workers, who have a "job" in March CPS, may actually be business owners. Occupations such as real estate sales appearing in Table 9 and 10 may be in Table 9 because workers, not their employers, fail to report their self-employment income to Social Security.

Conclusions

March CPS and SIPP appear to contain workers differing little in their true wage characteristics. The peculiarities of each survey, more than selection bias, seem to drive the differences in the level and size distribution of wages. Both March CPS and SIPP represent a worker's relative rank in the wage distribution better than the dollar amount of wages.

Data accuracy drives the differences between March CPS and SIPP wages more than differential household nonresponse, and varies by worker characteristics. In March CPS, working part-year and part-time relates to underestimated wages, especially with a very low number of weeks or usual hours worked. Having more than one job relates to overstated wages relative to administrative data.

March CPS may be superior to SIPP at capturing wages from the underground economy. Increased coverage of "off-the-books" workers by March CPS seems to have aggravated differences from the SIPP. March CPS wages with no corresponding wages in the administrative record-keeping system increase by 2 percentage points during the 1990 to 1996 period, possibly explaining one-third of the 6 percentage-point increase in aggregate wages relative to independent estimates. Many informal occupations appear disproportionally among CPS workers lacking any earnings record. Administrative data such as the DER may inadequately represent certain areas of the economy such as private household services, agriculture, and construction.

Classification peculiarities seem to contribute to the differences between March CPS and SIPP wages as well. Among workers with wages according to March CPS and self-employment according to administrative records, informal and small-business occupations appear disproportionally. March CPS seems to define some contract workers and small business owners as wage earners rather than self-employed. SIPP seems to define the class of worker more consistently with administrative data. The operating definition of self-employment clearly differs between administrative systems and Census Bureau surveys, especially March CPS. The rate of misclassifying self-employment as wages in March CPS increased in the 90s.

Figure A: March 1991 CPS and SIPP 1990 Wage Distributions
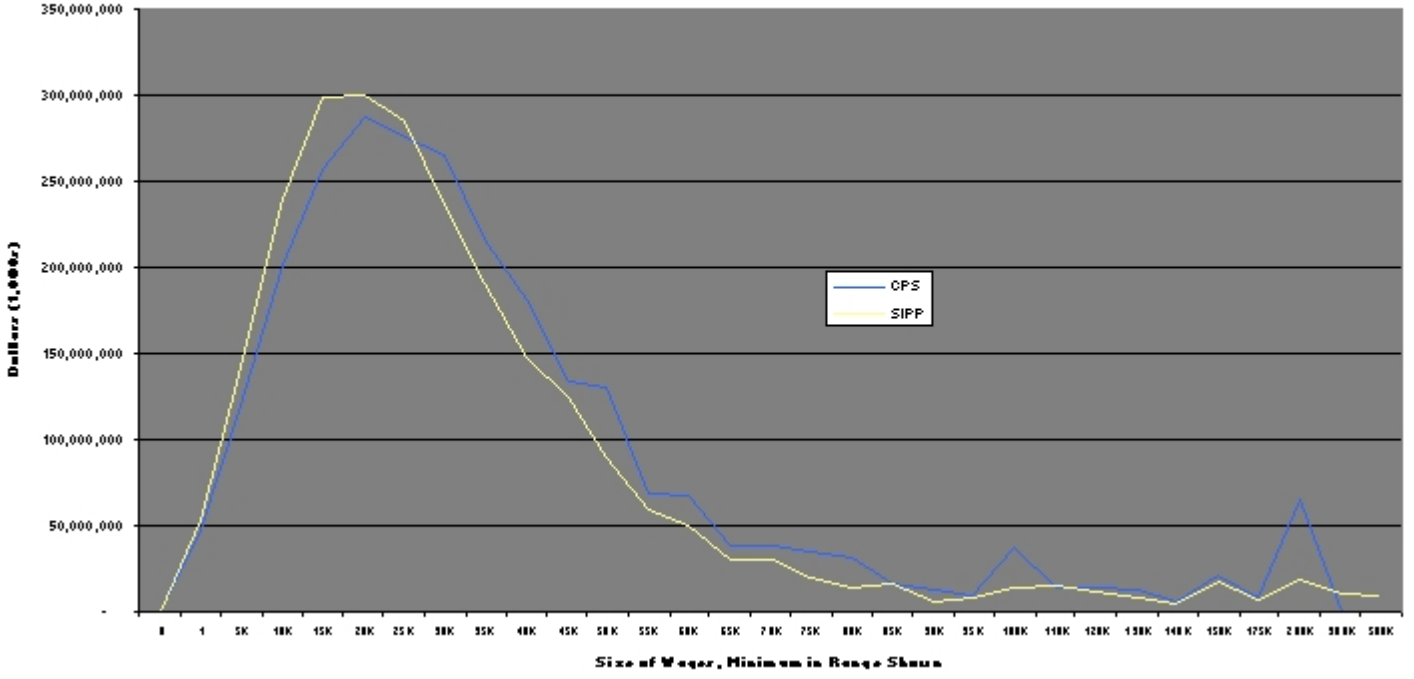
Figure B: March 1994 CPS and SIPP 1993 Wage Distributions
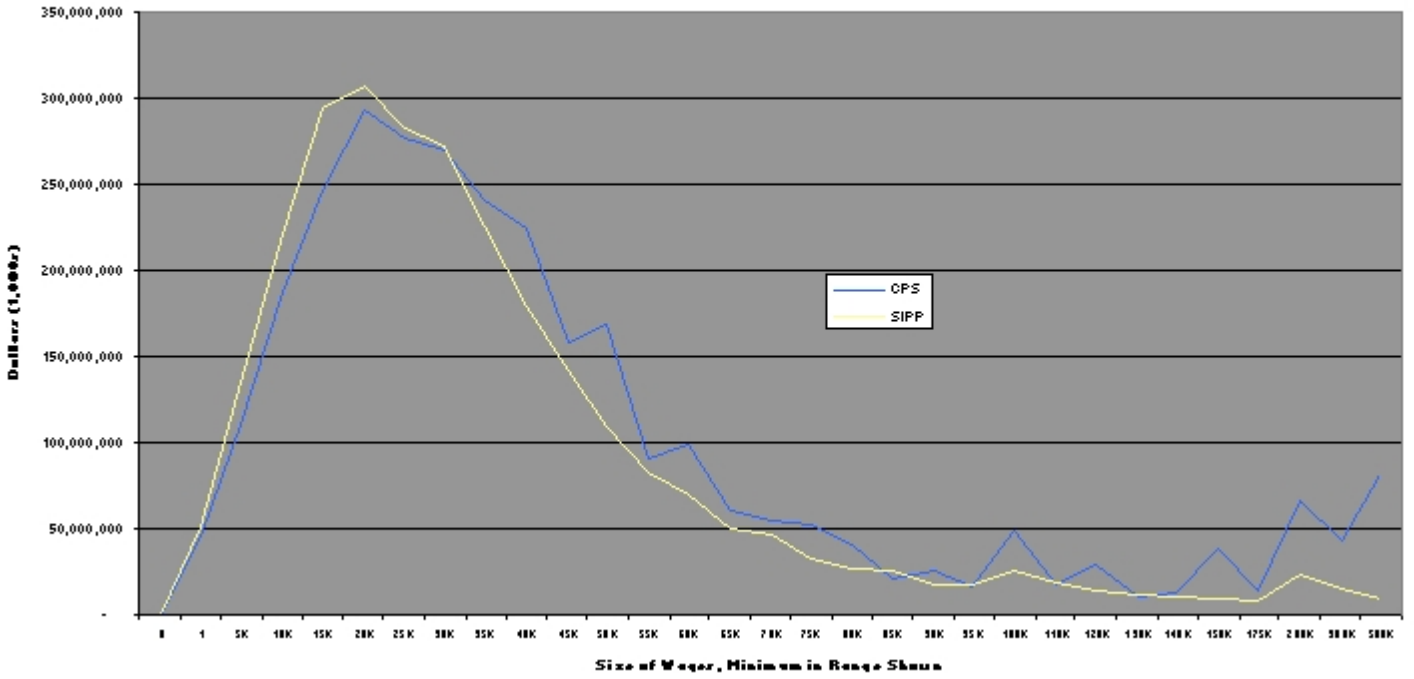
Figure C: March 1997 CPS and SIPP 1996 Wage Distributions
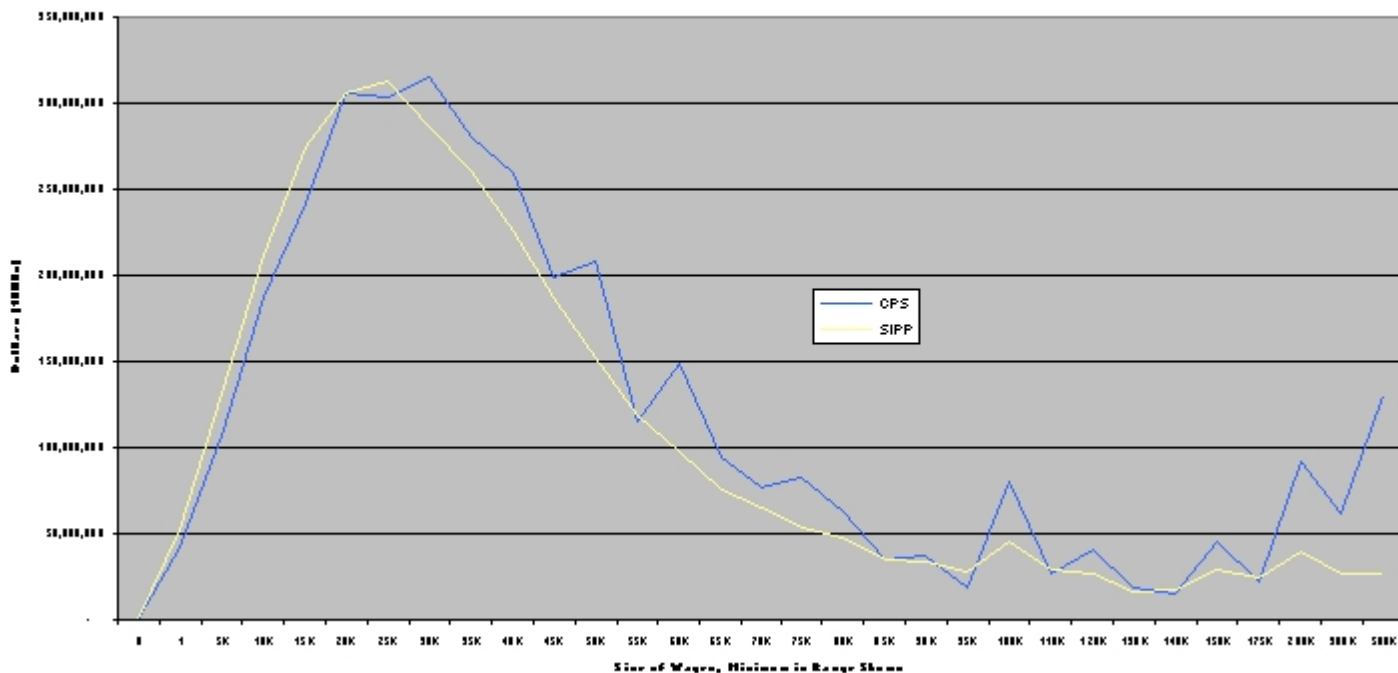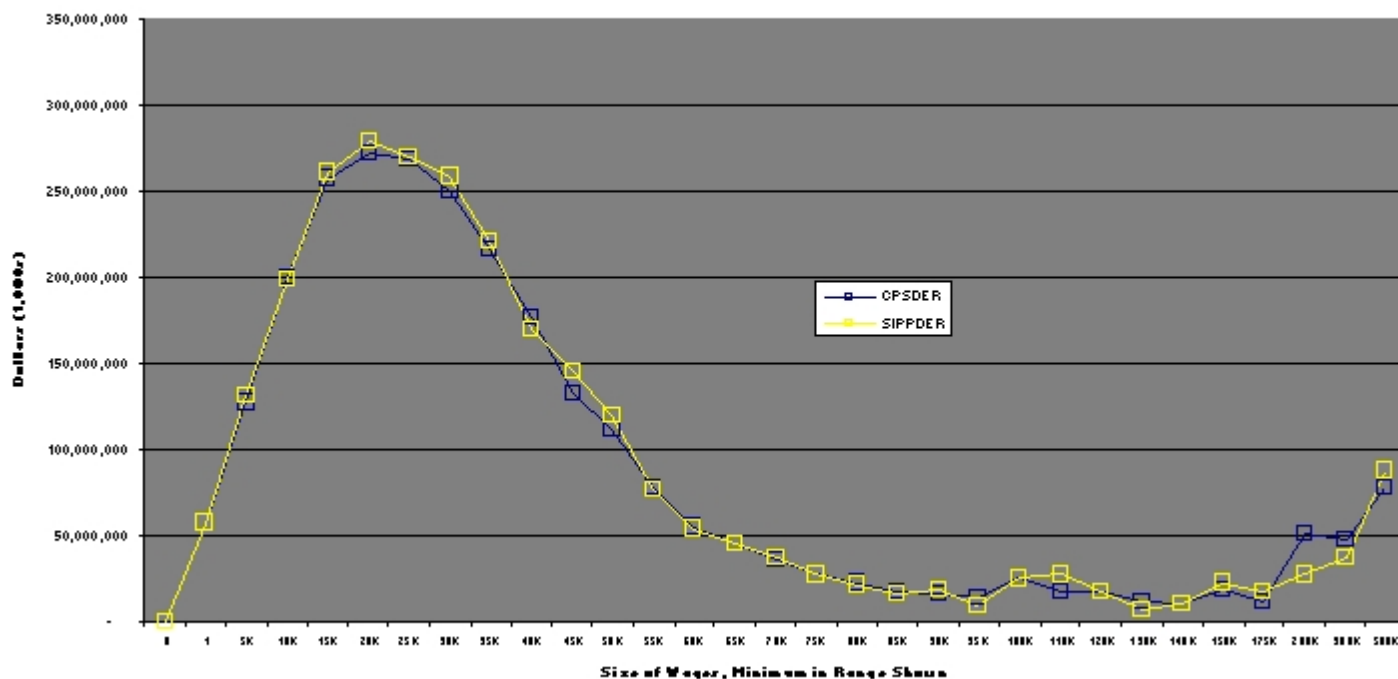

Figure D: DER 1996 Wage Distributions

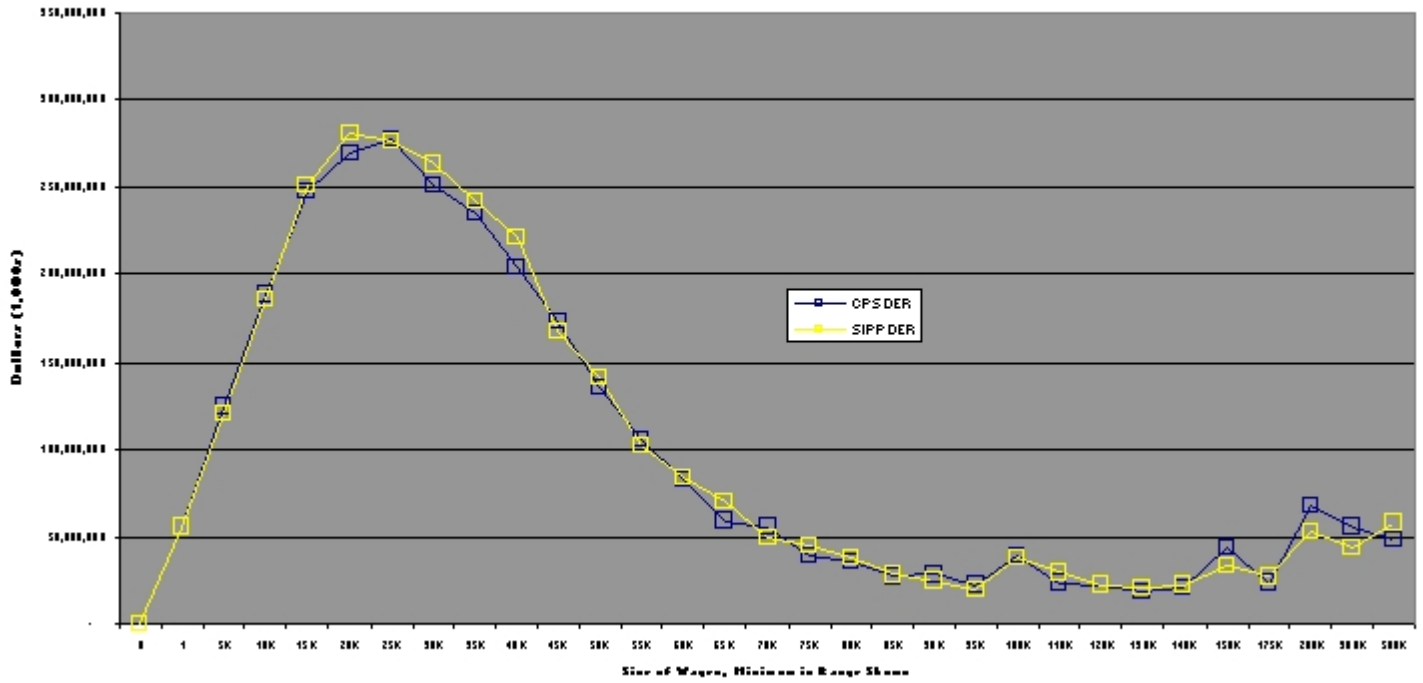Figure E: DER 1993 Wage Distributions
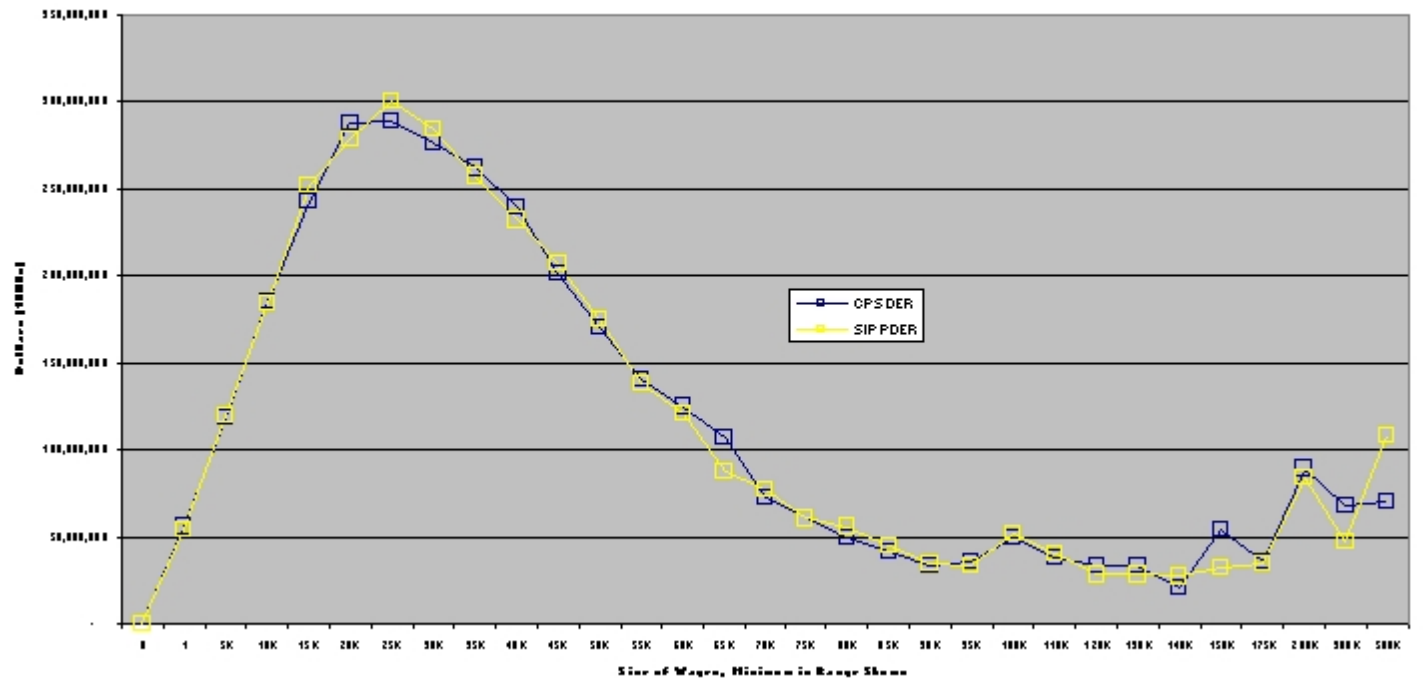

Figure F: DER 1996 Wage Distributions

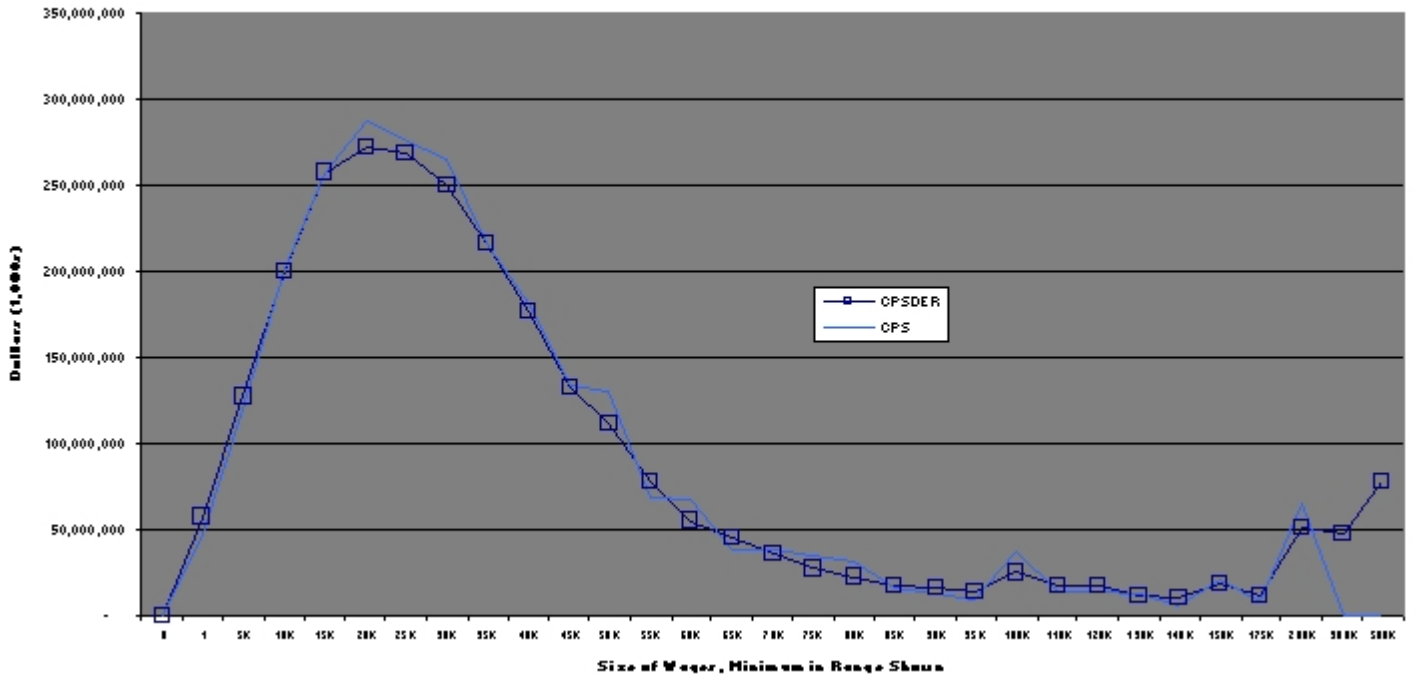Figure G: March 1991 CPS and DER 1990 Wage Distributions

Figure H: March 1994 CPS and DER 1993 Wage Distributions

Figure I: March 1997 CPS and DER 1996 Wage Distributions

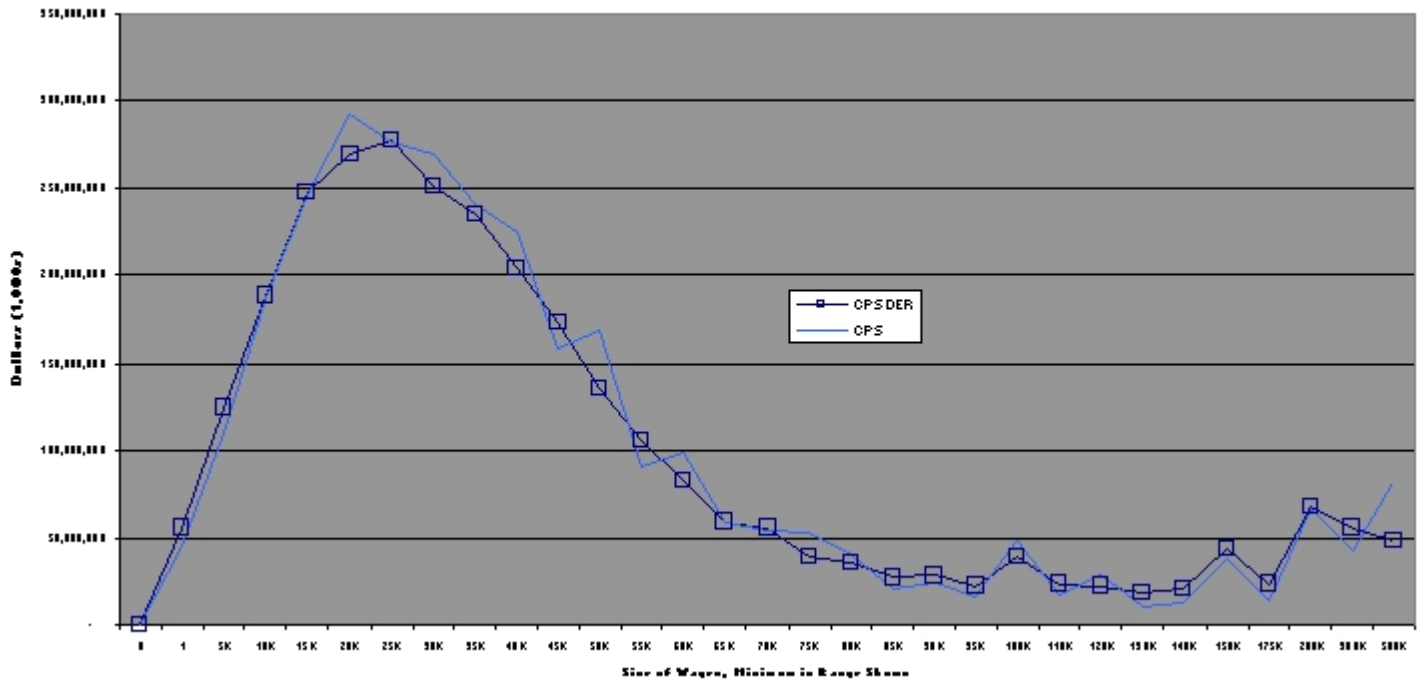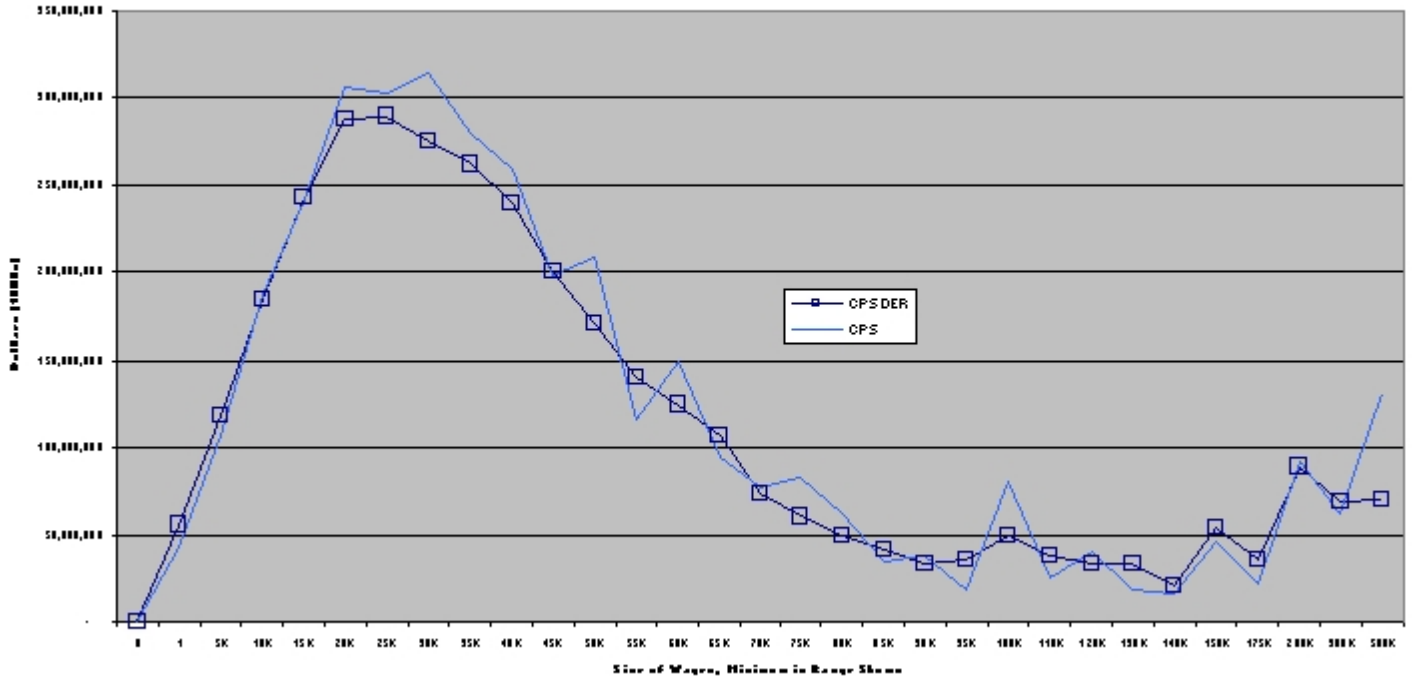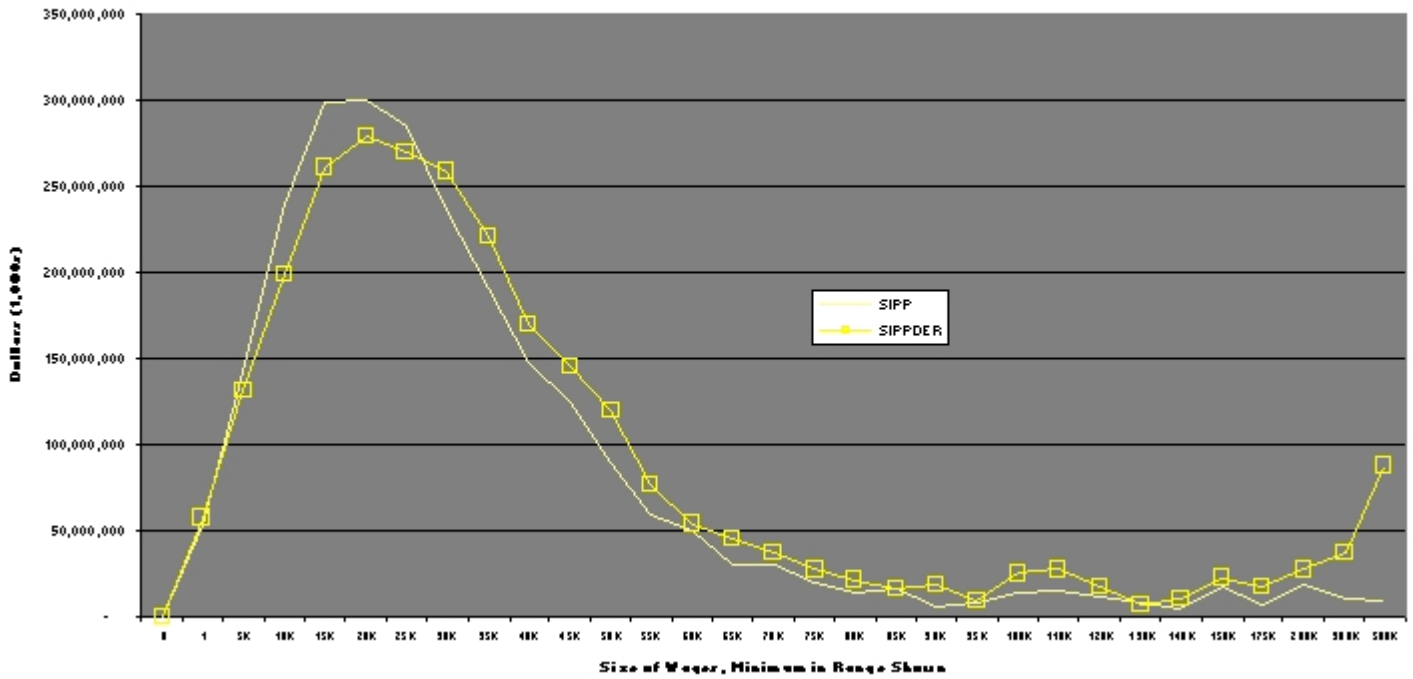Figure J: SIPP 1990 and DER 1990 Wage Distributions

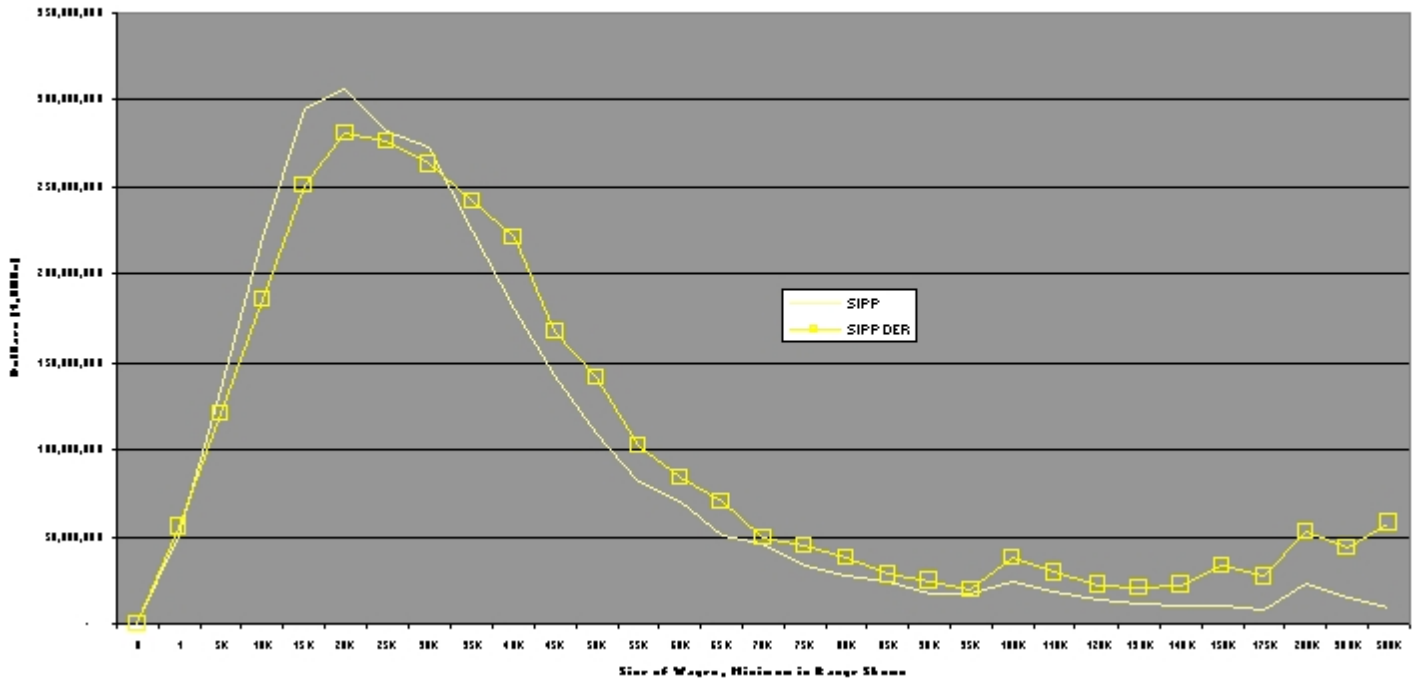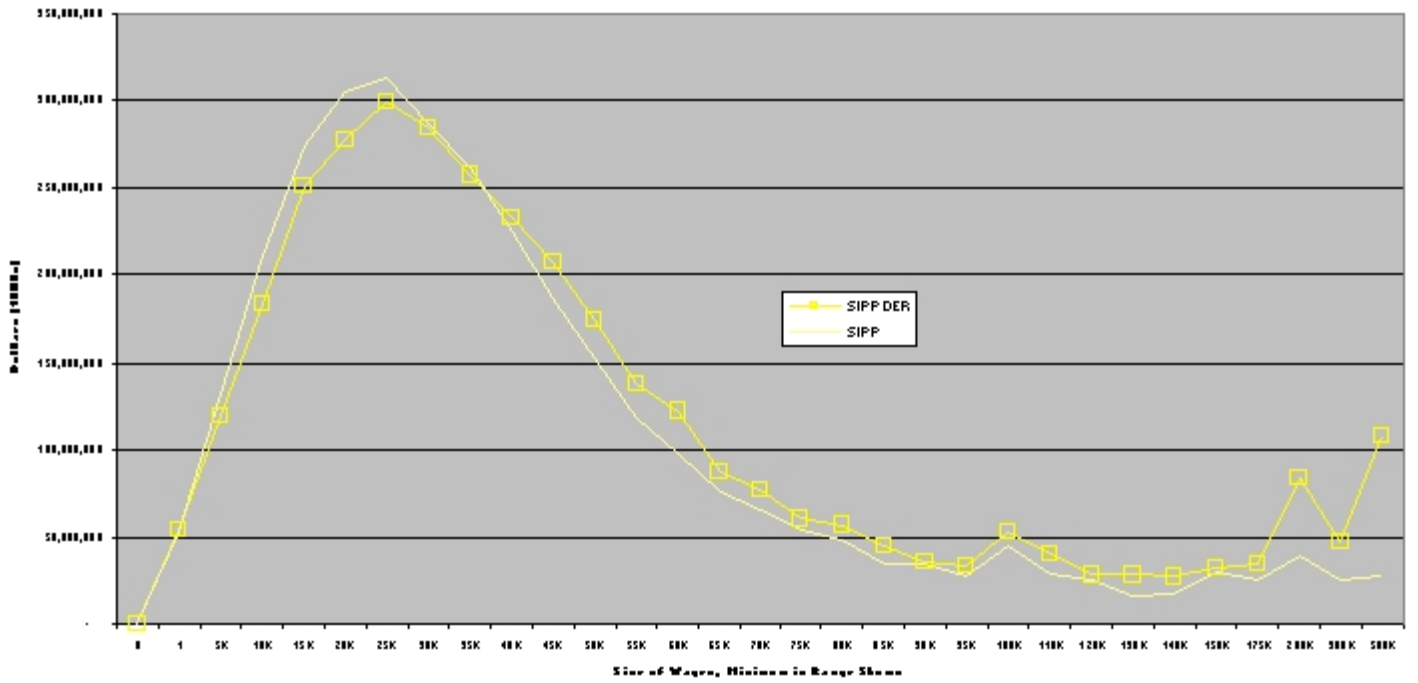Figure K: SIPP 1993 and DER 1993 Wage Distributions
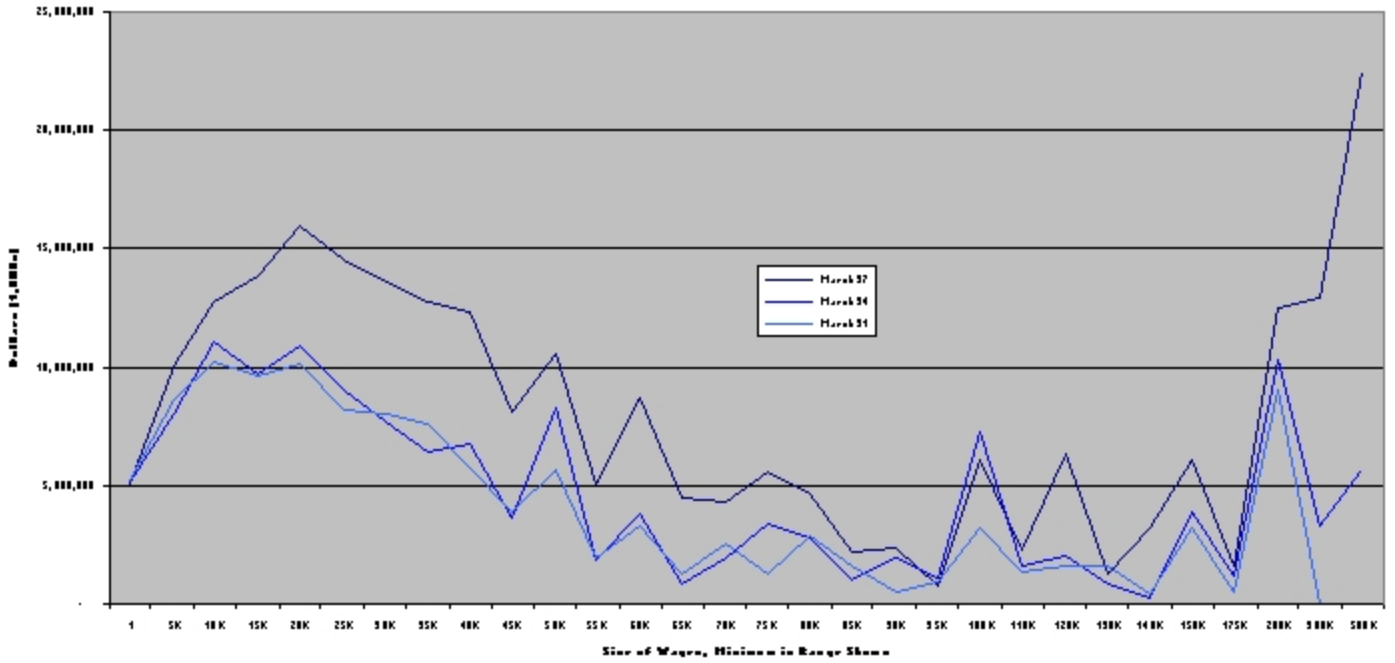

Figure L: SIPP 1996 and DER 1996 Wage Distributions

-22-

Figure M: Size Distribution of March CPS Wages, People with CPS Wages and 0 DER Wages

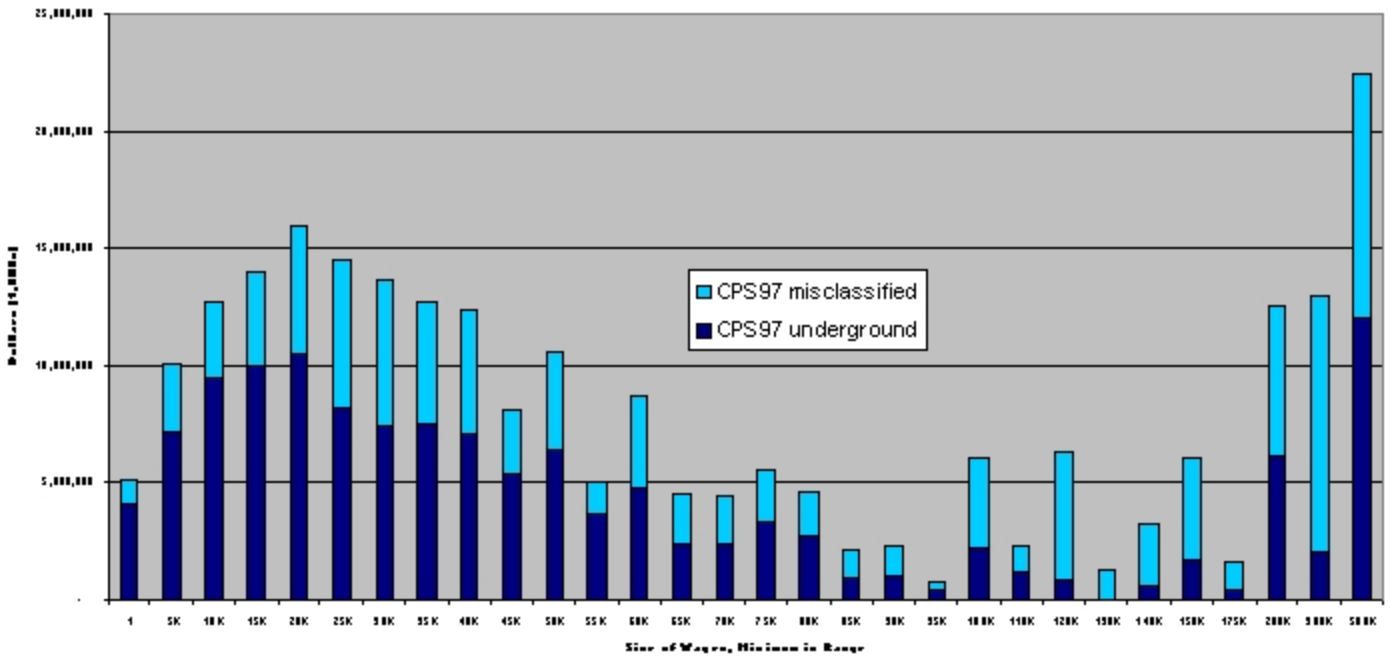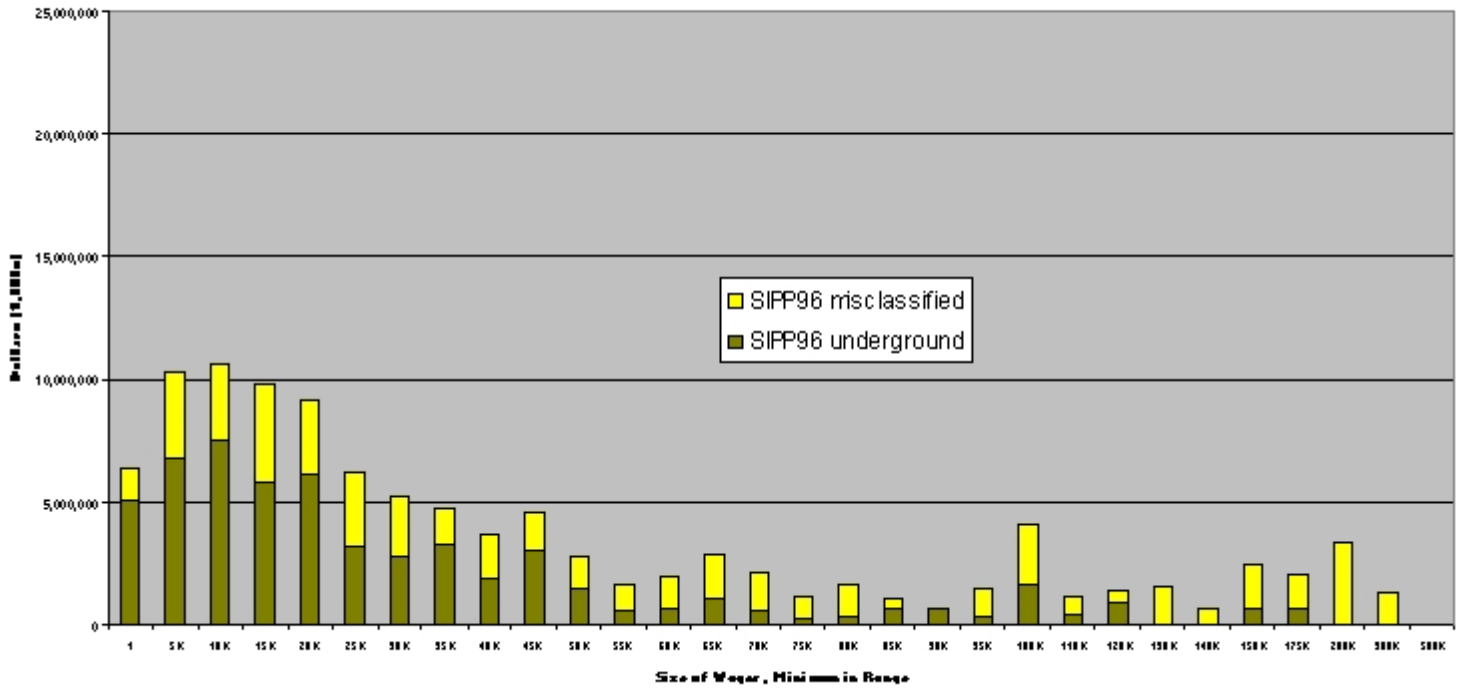**Figure N: Size Distribution of March 1997 CPS Wages Absent from DER**

Figure O: Size Distribution of SIPP 1996 Wages Absent from DER (n=1,558)

Appendix A. Underground and Missclasified Workers by Occupation

Table 9. Likelihood Ratios of People Working Underground in the March 1991, 1994, and 1997 CPS, by Occupation (weighted numbers in 1,000s)

Total	Legitimate Workers		Underground Workers		Occupation
		Percent	Number	Likelihood	
Number	Number	Distribution		Ratio	
1,323	406	0.11	841	39.1	Child care workers, private household
24	11	0.00	12	20.2	Cooks, private household
359	172	0.05	161	17.7	News vendors
1,465	759	0.21	546	13.6	Private household cleaners and servants
92	53	0.01	33	11.8	Family child care providers
1,278	577	0.16	251	8.2	Real estate sales occupations
54	37	0.01	15	7.8	Housekeepers and butlers
100	53	0.01	21	7.6	Barbers
123	78	0.02	28	6.8	Farmers, except horticultural
223	149	0.04	46	5.9	Carpet installers
170	120	0.03	36	5.6	Tile setters, hard and soft
242	180	0.05	52	5.5	Demonstrators, promoters and models, sales
732	528	0.15	146	5.2	Street and door-to-door sales workers
3,135	2,453	0.69	613	4.7	Farm workers
428	323	0.09	79	4.6	Managers, farms, except horticultural
311	235	0.07	53	4.3	Musicians and composers
2,790	2,212	0.62	500	4.3	Groundskeepers and gardeners, except farm
265	214	0.06	48	4.3	Helpers, construction trades
312	252	0.07	57	4.2	Athletes
502	387	0.11	80	3.9	Taxicab drivers and chauffeurs
225	185	0.05	38	3.9	Timber cutting and logging occupations
1,319	1,059	0.30	218	3.9	Painters, construction and maintenance
2,742	2,233	0.63	381	3.2	Construction laborers
192	157	0.04	26	3.2	Dressmakers
1,067	736	0.21	115	2.9	Clergy
267	219	0.06	34	2.9	Artists, performers, and rel. workers, n.e.c.
549	461	0.13	69	2.8	Roofers

Table continues on next page

Table 9 Continued

Total	Legitimate Workers		Underground Workers		Occupation
		Percent		Likelihood	
Number	Number	Distribution	Number	Ratio	
1,181	1,005	0.28	147	2.8	Child care wrkrs, n.e.c.
240	210	0.06	29	2.6	Grader and sorter, agricultural products
3,160	2,637	0.74	359	2.6	Carpenters
298	259	0.07	33	2.4	Animal caretakers, except farm
1,429	1,112	0.31	142	2.4	Hairdressers and cosmetologists
244	215	0.06	26	2.3	Insulation workers
272	243	0.07	29	2.2	Actors and directors
276	243	0.07	28	2.2	Helpers, mechanics, and repairers
453	399	0.11	46	2.2	Automobile body and rel. repairers
1,076	957	0.27	108	2.1	Bartenders
306	274	0.08	31	2.1	Concrete and terrazzo finishers
797	714	0.20	79	2.1	Attendants, amusement/recreation facilities
631	566	0.16	61	2.1	Launderers, ironers, and rel. machine operators
1,291	1,061	0.30	114	2.0	Early childhood teacher's assistants
781	694	0.19	74	2.0	Garage and service station rel. occupations
449	396	0.11	40	1.9	Postmasters and mail superintendents
1,430	1,246	0.35	125	1.9	Designers
958	869	0.24	86	1.9	Vehicle washers and equipment cleaners
7,415	6,631	1.86	646	1.8	Janitors and cleaners
554	501	0.14	46	1.7	Records clerks
831	734	0.21	67	1.7	Managers, properties and real estate
4,414	4,043	1.13	331	1.5	Laborers, except construction
2,516	2,324	0.65	186	1.5	Miscellaneous food preparation occupations
2,298	2,079	0.58	164	1.5	Automobile mechanics
2,293	2,117	0.59	160	1.4	Maids and housemen
1,796	1,674	0.47	122	1.4	Waiters' and waitresses' assistants
5,079	4,685	1.31	337	1.4	Sales workers, other commodities
7,146	6,668	1.87	448	1.3	Cooks
5,020	4,708	1.32	302	1.2	Waiters and waitresses

Universe: all workers in March 1991, 1994, or 1997 with wages from the longest job. Table excludes workers with unincorporated self-employment income from longest job (n=2,478). Shown: all occupations with a likelihood ratio statistically higher than 1.0 (at 90% confidence). Legitimate workers have DER wages (n=164,373). Underground workers have no DER wages or DER self-employment income (n=8,937).

Table 10. Likelihood Ratios of People Having Misclassified Pay in March 1991, 1994, and 1997 CPS, by Occupation (weighted numbers in 1,000s)

Total	Legitimate Workers		Workers with Misclassified Pay		Occupation
		Percent		Likelihood	
Number	Number	Distribution	Number	Ratio	
1,278	577	0.16	450	45.9	Real estate sales occupations
100	53	0.01	26	28.5	Barbers
88	56	0.02	22	23.2	Fishers
41	28	0.01	9	18.6	Social scientists, n.e.c.
164	126	0.04	37	17.4	Dentists
1,067	736	0.21	217	17.3	Clergy
123	78	0.02	17	13.2	Farmers, except horticultural
1,465	759	0.21	160	12.4	Private household cleaners and servants
223	149	0.04	28	11.2	Carpet installers
1,323	406	0.11	76	11.0	Child care workers, private household
1,429	1,112	0.31	175	9.3	Hairdressers and cosmetologists
359	172	0.05	25	8.7	News vendors
1,533	1,303	0.37	185	8.4	Lawyers and Judges
170	120	0.03	15	7.1	Tile setters, hard and soft
1,291	1,061	0.30	115	6.4	Early childhood teacher's assistants
732	528	0.15	57	6.4	Street and door-to-door sales workers
502	435	0.12	45	6.1	Management analysts
311	235	0.07	24	6.0	Musicians and composers
1,418	1,230	0.34	115	5.5	Insurance sales occupations
925	800	0.22	74	5.5	Securities and financial services sales
502	387	0.11	35	5.3	Taxicab drivers and chauffeurs
199	179	0.05	15	4.9	Stenographers
428	323	0.09	26	4.8	Managers, farms, except horticultural
267	219	0.06	15	4.0	Artists, performers, and related workers, n.e.c.
3,160	2,637	0.74	164	3.7	Carpenters
570	522	0.15	31	3.5	Psychologists
1,470	1,365	0.38	79	3.4	Physicians
2,742	2,233	0.63	128	3.4	Construction laborers
299	261	0.07	15	3.3	Photographers
384	341	0.10	18	3.1	Advertising and related sales occupations
1,430	1,246	0.35	59	2.8	Designers
549	461	0.13	20	2.5	Roofers
831	734	0.21	30	2.4	Managers, properties and real estate
1,491	1,357	0.38	55	2.4	Supervisors, construction, n.e.c.
1,319	1,059	0.30	42	2.3	Painters, construction and maintenance
2,790	2,212	0.62	78	2.1	Groundskeepers and gardeners, except farm
8,020	7,395	2.07	241	1.9	Truck drivers
4,219	3,963	1.11	121	1.8	Sales reps, mining, manufacturing, wholesale
1,715	1,567	0.44	47	1.8	Sales occupations, other business services
3,135	2,453	0.69	69	1.7	Farm workers
2,298	2,079	0.58	55	1.6	Automobile mechanics
11,146	10,524	2.95	226	1.3	Supervisors and proprietors, sales

Universe: workers with wages from the longest job in March 1991, 1994, or 1997. Table excludes workers with unincorporated self-employment income from longest job (n=2,478). Shown: all occupations with a likelihood ratio statistically higher than 1.0 (at 90% confidence). Legitimate workers have DER wages (n=164,373). Workers with misclassified pay have DER self-employment income, but no DER wages (n=2,827).

Bibliography

Coder, John, "Comparisons of Alternative Annual Estimates of Wage and Salary Income from SIPP," memorandum for Gordon Green, Assistant Division Chief, Population Division, Bureau of the Census, March 29, 1988.

Coder, John and Lydia Scoon-Rogers, "Evaluating the Quality of Income Data Collected in the Annual Supplement to the March Current Population Survey and the Survey of Income And Program Participation," Housing and Household Economic Statistics Division, Bureau of the Census, July 1996.

DeNavas, Carmen, personal communication, Income Statistics Branch, Housing and Household Economic Statistics Division, Bureau of the Census, October 31, 2002.

Geuzinge, Linda, Johan van Rooijen, and Bart F.F. Bakker, "The use of administrative registers to reduce non-response bias in household surveys," Statistics Netherlands Division of Socio-Cultural Statistics, Department of Integration and Presentation, October 8, 1999.

Roemer, Marc, "Assessing the Quality of March CPS and SIPP Income Estimates, 1990-1996," U.S. Census Bureau, 2000.

Tupek, Alan R., SIPP 1996: Source and Accuracy Statement for 1996 Panel Wave 1 - Wave 12 Public Use File, Memorandum to Walter C. Odom Jr, Administrative and Customer Services Division Chief, April 6, 2001.

U.S. Bureau of the Census, SIPP Quality Profile, SIPP Working Paper Number 320, Third Edition, 1998.

U.S. Bureau of Labor Statistics, Current Population Survey Technical Paper 63RV, Design and Methodology, March 2002.