APPENDIX B

Wage analysis computations

by Danielle Gao

This appendix provides background information on the analysis of wage data from the Current Population Survey (CPS), which is prepared by the Bureau of the Census for the Bureau of Labor Statistics (BLS). Specifically, for 1979 and beyond, we analyze microdata files provided by the BLS that contain a full year's data on the outgoing rotation groups (ORG) in the CPS. (For years prior to 1979, we use the CPS May files; our use of these files is discussed below.) We believe that the CPS ORG files allow for a timely, up-to-date, and accurate analysis of wage trends keeping within the familiar labor force definitions and concepts employed by BLS.

The sampling framework of the monthly CPS is a "rolling panel," in which households are in the survey for four consecutive months, out for eight, and then back in for four months. The ORG files provide data on those CPS respondents in either the fourth or eighth month of the CPS (i.e., in groups four or eight, out of a total of eight groups). Therefore, in any given month the ORG file represents a quarter of the CPS sample. For a given year, the ORG file is equivalent to three months of CPSs (one-fourth of 12). For our analysis, we use a sample drawn from the full-year ORG sample, the size of which ranges from 160,000 to 180,000 observations during the 1979 to 1995 period. Due to a decrease in the overall sample size of the CPS, the ORG was shrunk to 145,000 cases from 1996 to 1998, and our current sample comes in at about 170,000 cases.

Changes in annual or weekly earnings can result from changes in hourly earnings or from more working time (either more hours per week or weeks per year). Our analysis is centered around the hourly wage, which represents the pure price of labor (exclusive of benefits), because we are interested in chang-

ing pay levels for the workforce and its sub-groups. We do this to be able to clearly distinguish changes in earnings resulting from more (or less) work rather than more (or less) pay. Most of our wage analysis, therefore, does not take into account that weekly or annual earnings may have changed because of longer or shorter working hours or lesser or greater opportunities for employment. An exception to this is Table 2.1, where we present annual hours, earnings, and hourly weighted wages from the March CPS.

In our view, the ORG files provide a better source of data for wage analysis than the traditionally used March CPS files. In order to calculate hourly wages from the March CPS, analysts must make calculations using three retrospective variables: the annual earnings, weeks worked, and usual weekly hours worked in the year prior to the survey. In contrast, respondents in the ORG are asked a set of questions about hours worked, weekly wages, and (for workers paid by the hour) hourly wages in the week prior to the survey. In this regard, the data from the ORG are likely to be more reliable than data from the March CPS. See Bernstein and Mishel (1997) for a detailed discussion of these differences.

Our subsample includes all wage-and-salary workers with valid wage and hour data, whether paid weekly or by the hour. Specifically, in order to be included in our sub-sample, respondents had to meet the following criteria:

- age 18-64;
- employed in the public or private sector (unincorporated self-employed were excluded);
- hours worked within the valid range in the survey (1-99 per week, or hours vary—see discussion below); and,
- either hourly or weekly wages within the valid survey range (top-coding discussed below).

For those who met these criteria, an hourly wage was calculated in the following manner. If a valid hourly wage was reported, that wage was used throughout our analysis. For salaried workers (those who report only a weekly wage), the hourly wage was their weekly wage divided by their hours worked. Outliers, i.e., persons with hourly wages below 50 cents or above $100 in 1989 CPI-U-X1-adjusted dollars, were removed from the analysis. Starting from year 2002, we use CPI-RS adjusted dollars instead. These yearly upper and lower bounds are presented in **Table B-1**. CPS demographic weights were applied to make the sample nationally representative.

TABLE B-1 Wage earner sample, hourly wage upper and lower limits, 1973-2003

Year	Lower	Upper
1973	$0.19	$38.06
1974	0.21	41.85
1975	0.23	45.32
1976	0.24	47.90
1977	0.25	50.97
1978	0.27	54.44
1979	0.30	59.68
1980	0.33	66.37
1981	0.36	72.66
1982	0.39	77.10
1983	0.40	80.32
1984	0.42	83.79
1985	0.43	86.77
1986	0.44	88.39
1987	0.46	91.61
1988	0.48	95.40
1989	0.50	100.00
1990	0.53	105.40
1991	0.55	109.84
1992	0.57	113.15
1993	0.58	116.53
1994	0.60	119.52
1995	0.61	122.90
1996	0.63	126.53
1997	0.65	129.54
1998	0.66	131.45
1999	0.67	134.35
2000	0.69	138.87
2001	0.71	142.82
2002	0.70	140.05*
2003	0.72	143.26*

* adjusted by CPI_RS

Source: Authors' analysis.

The hourly wage reported by hourly workers in the CPS is net of any overtime, tips, or commissions (OTTC), thus introducing a potential undercount in the hourly wage for workers who regularly receive tips or premium pay. OTTC is included in the usual weekly earnings of hourly workers, which raises

the possibility of assigning an imputed hourly wage to hourly workers based on the reported weekly wage and hours worked per week. Conceptually, using this imputed wage is preferable to using the reported hourly wage because it is more inclusive. We have chosen, however, not to use this broader wage measure, because the extra information on OTTC seems unreliable. We compared the imputed hourly wage (reported weekly earnings divided by weekly hours) to the reported hourly wage; the difference presumably reflects OTTC. This comparison showed that significant percentages of the hourly workforce appeared to receive negative OTTC. These error rates range from a low of 0% of the hourly workforce in the period 1989-93 to a high of 16-17% in 1973-88, and persist across the survey change from 1993 to 1994. Since negative OTTC is clearly implausible, we rejected this imputed hourly wage series and rely strictly on the hourly rate of pay as reported directly by hourly workers, subject to the sample criteria discussed above.

For tables that show wage percentiles, we "smooth" hourly wages to compensate for "wage clumps" in the wage distributions. The technique involves creating a categorical hourly wage distribution, where the categories are 50-cent intervals, starting at 25 cents. We then find the categories on either side of each decile and perform a weighted, linear interpolation to locate the wage precisely on the particular decile. The weights for the interpolation are derived from differences in the cumulative percentages on either side of the decile. For example, suppose that 48% of the wage distribution of workers by wage level are in the $9.26-9.75 wage "bin," and 51% are in the next higher bin $9.76-10.25. The weight for the interpolation (in this case the median or 50th percentile) is $\frac{(50\text{-}48)}{(51\text{-}48)}$ or 2/3. The interpolated median equals this weight, times the width of the bin ($.50), plus the upper bound of the previous bin ($9.75), or $10.08 in this example.

For the survey years 1973-88, the weekly wage is top-coded at $999.00; an extended top-code value of $1,923 is available in 1986-97; the top-code value changes to $2,884.61 in 1998-2003. Particularly for the later years, this truncation of the wage distribution creates a downward bias in the mean wage. We dealt with the top-coding issue by imputing a new weekly wage for top-coded individuals. The imputed value is the Pareto-imputed mean for the upper tail of the weekly earnings distribution, based on the distribution of weekly earnings up to the 80th percentile. This procedure was done for men and women separately. The imputed values for men and women appear in **Table B-2**. A new hourly wage, equal to the new estimated value for weekly earnings, divided by that person's usual hours per week, was calculated.

TABLE B-2 Pareto-imputed mean values for top-coded weekly earnings, and share top coded, 1973-2003

	Share			Value	
Year	All	Men	Women	Men	Women
1973	0.11%	0.17%	0.02%	$1,365	$1,340
1974	0.16	0.26	0.01	1,385	1,297
1975	0.21	0.35	0.02	1,410	1,323
1976	0.30	0.51	0.01	1,392	1,314
1977	0.36	0.59	0.04	1,384	1,309
1978	0.38	0.65	0.02	1,377	1,297
1979	0.57	0.98	0.05	1,388	1,301
1980	0.72	1.23	0.07	1,380	1,287
1981	1.05	1.82	0.10	1,408	1,281
1982	1.45	2.50	0.18	1,430	1,306
1983	1.89	3.27	0.25	1,458	1,307
1984	2.32	3.92	0.42	1,471	1,336
1985	2.78	4.63	0.60	1,490	1,343
1986	0.80	1.37	0.15	2,435	2,466
1987	1.06	1.80	0.20	2,413	2,472
1988	1.30	2.19	0.29	2,410	2,461
1989	0.48	0.84	0.08	2,710	2,506
1990	0.60	1.04	0.11	2,724	2,522
1991	0.71	1.21	0.17	2,744	2,553
1992	0.77	1.28	0.22	2,727	2,581
1993	0.86	1.43	0.24	2,754	2,580
1994	1.25	1.98	0.43	2,882	2,689
1995	1.34	2.16	0.43	2,851	2,660
1996	1.41	2.27	0.46	2,863	2,678
1997	1.71	2.67	0.65	2,908	2,751
1998	0.63	0.97	0.24	4,437	4,155
1999	0.71	1.18	0.21	4,464	4,099
2000	0.83	1.37	0.24	4,502	4,179
2001	0.91	1.44	0.33	4,477	4,227
2002	1.05	1.66	0.38	4,555	4,252
2003	1.07	1.69	0.40	4,546	4,219

Source: Authors' analysis.

In January 1994, a new survey instrument was introduced into the CPS; many labor force items were added and improved. This presents a significant challenge to the researcher who wishes to make comparisons over time. The most careful research on the impact of the survey change has been conducted

by BLS researcher Anne Polivka (1996, 1997). Interestingly, Polivka does not find that the survey changes had a major impact on broad measures of unemployment or wage levels, though significant differences did surface for some sub-groups (e.g., weekly earnings for those with less than a high school diploma and those with advanced degrees, the unemployment rate of older workers). However, a change in the reporting of weekly hours did call for the alteration of our methodology. In 1994 the CPS began allowing people to report that their usual hours worked per week vary. In order to include non-hourly workers who report varying hours in our wage analysis, we estimated their usual hours using a regression-based imputation procedure, where we predicted the usual hours of work for "hours vary" cases based on the usual hours worked of persons with similar characteristics. An hourly wage was calculated by dividing weekly earnings by the estimate of hours for these workers. The share of our sample that received such a wage in the 1994-97 period is presented in **Table B-3**. The reported hourly wage of hourly workers was preserved.

BLS analysts Ilg and Hauzen (2000), following Polivka (1999), do adjust the 10th percentile wage because "changes to the survey in 1994 led to lower reported earnings for relatively low-paid workers, compared with pre-1994 estimates." We make no such adjustments for both practical and empirical reasons. Practically, the BLS has provided no adjustment factors for hourly wage trends that we can use—Polivka's work is for weekly wages. More importantly, the trends in 10th percentile hourly wages differ from those reported by Ilg and Hauzen for 10th percentile weekly earnings. This is perhaps not surprising, since the composition of earners at the "bottom" will differ when measured by weekly rather than hourly wages, with low-weekly earners being almost exclusively part-timers. Empirically, Ilg and Hauzen show the unadjusted 50/10 wage gap jumping up between 1993 and 1994, when the new survey begins. In contrast, our 50/10 wage gap for hourly wages falls between 1993 and 1994. Thus, the pattern of wage change in their data differs greatly from that in our data. In fact, our review of the 1993-94 trends across all of the deciles shows no discontinuities whatsoever. Consequently, we make no adjustments to account for any effect of the 1994 survey change. Had we made the sort of adjustments suggested by Polivka, our measured 1990s' fall in the 50/10 wage gap would be even larger and the overall pattern—falling 50/10, rising 90/50, and especially the 95/50 wage gaps—would remain the same.

When a response is not obtained for weekly earnings, or an inconsistency is detected, an "imputed" response is performed by CPS using a "hot deck" method, whereby a response from another sample person with similar demographic and economic characteristics is used for the nonresponse. This procedure for imputing missing wage data appears to bias between union and non-

TABLE B-3 Share of wage earners assigned an hourly wage from imputed weekly hours, 1994-2003

Year	Percent hours vary
1994	2.0%
1995	2.1
1996	2.4
1997	2.4
1998	2.5
1999	2.4
2000	2.4
2001	2.5
2002	2.5
2003	2.5

Source: Authors' analysis.

union members. We restrict our sample to the observations with non-imputed wages only for union wage premium analysis (table 2.35).

Demographic variables are also used in the analysis. Our race variable comprises four mutually exclusive categories:

- white, non-Hispanic;
- black, non-Hispanic;
- Hispanic, any race;
- all others.

Beginning in 1992, the CPS employed a new coding scheme for education, providing data on the respondent's highest degree attained. The CPS in earlier years provided data on years of schooling completed. The challenge to make a consistent wage series by education level is to either make the new data consistent with the past or to make the old "years of schooling" data consistent with the new, educational attainment measures. In prior versions of *The State of Working America*, we achieved a consistent series by imputing years of schooling for 1992 and later years, i.e., making the "new" consistent with the "old." In this version, however, we have converted the "old" data to the new coding following Jaeger (1997). However, Jaeger does not separately identify four-year college and "more than college" categories. Since the wages of these sub-

groups of the "college or more" group have divergent trends, we construct pre-1992 wages and employment separately for "four-year college" and "advanced." To do so, we compute wages, wage premiums, and employment separately for those with 16, 17, and 18-plus years of schooling completed. The challenge is to distribute the "17s" to the 16 years (presumably a four-year degree) and 18-plus years (presumably advanced) groups. We do this by using the share of the "17s" that have a terminal four-year college degree, as computed in the February 1990 CPS supplement that provides both education codings: 61.4%. We then assume that 61.4% of all of the "17s" are "college-only" and compute a weighted average of the "16s" and 61.4% of the "17s" to construct "college-only" wages and wage premiums. Correspondingly, we compute a weighted average of 38.6% (or 1 less 61.4%) of the "17s" and the "18s" to construct advanced "wages and wage premiums." Distributing the "17s" affects each year differently depending on the actual change in the wages and premiums for "17s" and the changing relative size of the "17s" (which varies only slightly from 2.5% of men and women from 1979 to 1991).

We employ these education categories in various tables in Chapter 2, where we present wage trends by education over time. For the data for 1992 and later, we compute the "some college" trends by aggregating those "with some college but no degree beyond high school" and those with an associate or other degree that is not a four-year college degree.