

Creating a consistent hourly wage series from the Current  
Population Survey's Outgoing Rotation Group, 1979-2002

John Schmitt  
schmitt@cepr.net

Version 0.9  
August 2003

## **Abstract**

This paper summarizes a set of adjustments made to raw earnings data from the Current Population Survey's Outgoing Rotation Group in order to produce a consistent series of hourly earnings for the period 1979-2002. These adjustments include: adjusting for changes in the top-coding of weekly earnings; trimming outliers; including overtime, tips, and commissions for "hourly workers"; and adjusting for the "hours vary" response to the "usual weekly hours", instituted in 1994. The paper concludes that, for the period 1979-2002, the most consistent and robust hourly wage series: (1) uses a log-normal imputation to adjust for top-coding; (2) excludes data below \$1 and above \$100 per hour (in constant 2002 dollars); (3) excludes overtime, tip, and commission earnings for hourly paid workers; and (4) uses a simple procedure to impute usual weekly hours for those who report their "hours vary" after 1994. For the period 1994-2002, the best hourly wage series follows the same procedure, but includes overtime, tips, and commissions for workers that report their earnings by the hour. Both these preferred series (as well as several variations) appear in the Washington-DC-based Center for Economic and Policy Research's Uniform Data Files for the Outgoing Rotation Group and are available through [www.cepr.net](http://www.cepr.net).

## **Introduction**

Since 1979, the Current Population Survey has asked employed adults in one-quarter of the survey's monthly sample –a group often referred to as the "Outgoing Rotation Group" (ORG)– to answer a detailed set of questions about their earnings from work.<sup>1</sup> Several features of the survey and changes over the years in the survey design, however, make it difficult to create a consistent hourly wage series from the raw survey responses. This paper describes a set of procedures designed to overcome several well known obstacles confronting researchers attempting to create a consistent definition of hourly earnings from the ORG data for the period 1979-2002.

The paper addresses problems that arise from: (1) the topcoding of weekly earnings and especially the large, abrupt changes in the topcode in 1989 and 1998; (2) the presence of outliers –implausibly low and less implausibly high– estimated hourly earnings; (3) the uneven treatment of earnings from overtime, tips, and commissions between workers who report their earnings by the hour and those who report their earnings at other intervals; and (4) the introduction, in 1994, of the option to report that usual weekly hours "vary," without having to report a specific number of usual hours. The paper also notes potential problems related to the Bureau of Labor Statistics' (BLS) practice of allocating earnings to respondents who don't provide them or provide responses that the BLS considers implausible.

The next section of the paper provides a brief description of the data set analyzed here. The rest of the paper treats, in turn, each of the preceding potential data problems and then concludes with recommendations for the creation of a consistent hourly wage series.

## **The CEPR ORG Extract**

The version of the CPS ORG data analyzed here is from the Center for Economic and Policy Research (CEPR)'s CPS ORG Uniform Data File. The CEPR ORG extracts use the CPS "Annual Earnings File" data from the National Bureau of Economic Research (NBER) as a starting point. The NBER data sets are a well-documented, broadly consistent, version of the raw CPS data and are available in multiple formats from the NBER for a small fee. The CEPR version of the ORG extracts makes several modifications to the NBER data files and, from 1994 on, also includes additional data from raw CPS files available through the internet from the BLS. All programs used to produce the CEPR extracts, plus additional documentation, are available for download from the CEPR web page ([www.cepr.net](http://www.cepr.net)).<sup>2</sup> The 24 annual CEPR ORG extract data files, covering the years 1979 through 2002 are available, upon request, from CEPR.

## **Adjustments to CPS ORG wage data**

### **Topcoding**

The ORG generally reports earnings information at two intervals: on an hourly basis, for "hourly workers",<sup>3</sup> and, on a weekly basis, for all other, generally non-hourly, workers. (The BLS takes the reported earnings information for non-hourly workers and converts all earnings to a weekly basis.) The first difficulty facing efforts to make a consistent hourly earnings series is that, while hourly workers wages are generally topcoded at \$99.99 per hour (a threshold rarely crossed), weekly earnings are topcoded at much lower thresholds, with an important share of workers above the threshold in some years. Moreover, as the first column of **Table 1** shows, the topcode rose abruptly in 1989 –from \$999 to \$1,923 per week– and again in 1998 –to \$2,884. During periods when the topcode remains unchanged (1979-1988, 1989-1997, and since 1998),

the share of "weekly" workers affected by the topcode grows steadily. Over the course of the 1980s, in particular, the share of topcoded workers grew to represent a fairly large share of the total sample (see column two of Table 1).

Topcoding has important implications for the analysis of ORG wage data. In a single cross-section, topcoding will lower the mean and the variance of the wage data relative to the true mean and variance. Over time, with a fixed nominal topcode and wage inflation, the size of this bias will generally increase, lowering estimates of the change in the mean and variance of wages. With no change in the underlying mean and variance, for example, the measured mean and variance of real wages would generally fall. Irregular and large adjustments to the topcode create further problems by causing sudden jumps in the mean and variance of wages that are not related to actual changes in the true wage distribution.

Researchers have responded to topcoding in several ways, often depending on the purpose of a particular analysis. Some ignore the topcoding issue.<sup>4</sup> An alternative strategy is to attempt to estimate the mean above the topcode and assign this to mean to all topcoded observation. In principle, if all observations above the topcode are assigned the true mean above the topcode, the mean of the resulting distribution would equal the mean of the true distribution. This procedure doesn't produce the correct variance, but, in general would push the variance of the observed distribution closer to the true variance, reducing the downward bias both within a given cross-section and over time. In the absence of data from the BLS on the actual mean above the topcode (and the actual variance of the underlying distribution), researchers must produce their own estimates of the mean above the topcode. Some researchers have taken a simple, arbitrary, approach and multiplied the topcode by 1.3 or 1.4, which they believe to provide estimates of the mean and the variance that are closer to their true values. Others have borrowed

from the literature on topcoding of annual incomes and have estimated the mean above the topcode using the pareto distribution.<sup>5</sup> This section of the paper compares the results of using several different strategies: (1) making no adjustments for topcoding; (2) making adjustments using three different versions of the pareto approach; and (3) making adjustments based on the assumption that the underlying weekly earnings data are log-normally distributed.

The pareto approach is the most technical of the three general strategies. Researchers assume that the top portion of the weekly earnings distribution is well described by the pareto distribution, which is given by:

$$(1) \quad \Pr(y > x) = (k/x)^\alpha$$

where  $y$  is weekly earnings,  $x$  is a given value of weekly earnings, and  $k$  and  $\alpha$  are parameters of the distribution ( $\alpha$  must be greater than 1). Since the distribution has only two parameters, it is possible to estimate the pareto distribution above a given point in the weekly wage data with data from only two points in the observed distribution of wages. Researchers generally choose a point high in the distribution, say the 70th, 80th, or 90th percentile of the observed distribution as the first point, and the topcode as the second point, and then fit the pareto distribution through these two points.<sup>6</sup> The specific pareto distribution calculated using the estimates of  $k$  and  $\alpha$  provide an estimate of the (unobserved) mean of the wage distribution above the topcode.

The log-normal approach is similar in spirit, but somewhat easier to implement in practice. The log-normal procedure models the entire distribution, not just the "top" portion of interest, under the assumption that the entire distribution of weekly earnings is log-normally distributed. Given the mean and the variance of the distribution including the topcode, and information on the share of workers that are above the topcode, the properties of the log-normal allow for the straightforward estimation of the mean and variance of the "true" distribution, that

is, of the distribution without the topcode. This estimate of the "true" distribution, then, allows for the estimation of the mean of weekly earnings distribution above the topcode.<sup>7</sup>

The last four columns of Table 1 display the resulting estimates of the mean above the topcode using the log-normal approach and the pareto approach, implemented separately from the 70th, 80th, and 90th percentile of the entire weekly earnings distribution. The log-normal approach produces estimates of the mean above the top-code that are consistently below those generated by the three versions of the pareto approach. In most years, the three pareto approaches give estimates for the mean above the top-code that are fairly close to one another. Two exceptions are the 90th percentile estimate in 1987 and 1988. In the case of 1988, the technique is not defined because the topcode cuts below the 90th percentile of the weekly earnings distribution. In 1987, the 90th percentile of the weekly earnings distribution is very close to the topcode (which cuts at the 92nd percentile), yielding an estimate of the full distribution that differs markedly from what is obtained using the (arguably more robust) information contained in the 70th and 80th percentile estimates.<sup>8</sup> Given these problems with the 90th percentile distribution and the fairly close match between the 70th and 80th percentile estimates, the remainder of the analysis here will focus on comparing three wage series: (1) one that makes no attempts to adjust for topcoding: `w_nber`, the wage series recommended (with appropriate cautions) by the NBER's Annual Earnings File; (2) one that assigns the mean above the topcode from the log-normal procedure to all topcoded observations: `w_ln_no` (where "ln" indicates the log-normal adjustment, and "no" indicates that the series excludes overtime, tips, and commissions for hourly workers); and (3) one that assigns the mean above the topcode from the 80th percentile pareto procedure to all topcoded observations: `w_p8` (where "p8" indicates

the pareto 80th percentile adjustment, and "no" indicates that the series excludes overtime, tips, and commissions for hourly workers).<sup>9</sup>

**Table 2** and **Figure 1** compare the mean of the three resulting distributions. The series that makes no adjustments for topcoding (*w\_nber*) shows a large jump in the mean wage between 1988 and 1989, when the topcode increased from \$999 to \$1,923 per week (the vertical lines in Figure 1 and subsequent graphs correspond to the years when the topcode changed). The series based on the 80th percentile pareto procedure (*w\_p8\_no*), on the other hand, drops markedly between the two years. One possible explanation for the sharp decline is that the 80th percentile pareto procedure overestimates the mean above the topcode in the period immediately before the 1989 topcode change. Meanwhile, the series based on the log-normal adjustment dips only slightly between 1988 and 1989, with no obvious discontinuity in the series. All three series show remarkably similar behavior thereafter, even through the 1998 topcode change.

**Table 3** and **Figure 3** compare the standard deviations of the same three distributions. The series without a topcode adjustment shows a flat, almost declining, standard deviation over the 1980s, with a sudden jump coinciding with the topcode change in 1989. As before, the 80th percentile pareto distribution falls abruptly across the same change. The log-normal series drops, too, but much less precipitously. As with the mean, the three series respond almost identically to the topcode change in 1998. Figure 3 also highlights another important feature of the time series. The redesign of the CPS in 1994 appears to have led to a marked rise in the standard deviation of all three earnings series between 1993 and 1994.<sup>10</sup>

On the assumption that adjustment techniques that produce smooth changes in the mean and in the standard deviation across major survey changes are best, the evidence in Tables 1 and

2 and Figures 1 and 3 suggests that the log-normal adjustment offers the best method for adjusting topcoded data in the ORG.

### **Trimming**

A concern that is related to the question of topcoding is how best to deal with implausibly low or implausibly high estimates of hourly wages. Despite nearly universal coverage of the minimum wage, for example, a substantial share of workers report hourly wages below the minimum wage in every year of CPS data. Some of these are workers who report a wage that is equal to the minimum wage minus the "tip credit" applied to workers who usually receive tips as a part of their compensation. Many workers with wages below the minimum wage, however, are non-hourly workers whose reported usual weekly earnings and usual weekly hours yield an hourly rate that is very low. At the other end of the distribution, some non-hourly workers report usual weekly earnings and usual weekly hours that put them well over \$100 per hour. Many of these workers undoubtedly earn these wages, but close inspection reveals that many of these workers are in industry and occupation groups where high earnings are rare; these very-high-wage workers also frequently have "ordinary-looking" usual weekly earnings and very low usual weekly hours, suggesting that the source of their high hourly rate may be reporting errors.

In response, many researchers "trim" the wage distributions to eliminate potentially problematic outliers (while, undoubtedly, simultaneously eliminating some legitimate observations, too). **Figure 5** compares the untrimmed hourly wage series with the log-normally adjusted series trimmed so that all observations lie in the range \$1-50 (fairly severe trimming), \$1-100, or \$1-200 per hour (in constant 2002 dollars). As Figure 5 demonstrates, trimming the data at \$1-50 produces a series with a mean that is well below the mean for the other three series (see also **Table 4**). The \$1-50 series also has a noticeable spike coinciding with the 1988-89

change in the topcode for weekly earnings. Trimming the data at \$1-100 or \$1-200 per hour, however, yields series that both track the mean of the untrimmed data closely.

**Figure 6** compares the standard deviations of the same four series (see also **Table 5**). Once again, the \$1-50 trimmed series lies well below the other three, with the same spike associated with the 1988-89 change in the weekly earnings topcode. The standard deviation of the untrimmed series, however, now lies above those of the series trimmed at \$1-100 and \$1-200. Moreover, the untrimmed series exhibits a more jagged pattern than the other three series, suggesting that outliers may be periodically raising and lowering the variance of the wage distribution, relative to the "true" variance. Trimming at \$1-100 and \$1-200 makes little difference through 1994; from 1994 on, the gap between the two series is larger, with most of the discrepancy opening up immediately after the 1994 survey change.

Given that topcoding and trimming both have a potential effect on outliers, **Figures 2 and 4** analyze the interaction of the three main topcoding procedures and trimming at \$1-100. Comparing the series in Figure 1 (topcoding, but no trimming) and Figure 2 (topcoding and trimming) suggests that the trimming procedure has relatively little impact on the effect of the topcoding procedures on the estimated mean. Meanwhile, comparing Figures 3 and 4 reinforces concerns about failing to make adjustments for topcode (`w_nber`) or using the pareto distribution at the 80th percentile (`w_p8_no`).

Summarizing the results of trimming, failing to trim the ORG wage data has little effect on estimates of the mean hourly earnings; failure to trim, however, produces a time series for the standard deviation of wages with an implausibly saw-toothed pattern. Trimming at \$1-50 yields a series that has a substantially lower mean than the untrimmed series and a spike in both the mean and the standard deviation in 1989. Trimming at \$1-100 and \$1-200 produce fairly similar

results, but the \$1-100 series shows a smaller discontinuity before and after the 1994 survey change, which probably makes it the slightly better choice. Given that the increase in the standard deviation is smaller using the \$1-100 series than it is using the \$1-200 series, the choice of the \$1-100 series should produce more conservative estimates of the rise in wage inequality over the span of the data set.

### **Overtime, tips, commissions**

As mentioned earlier, the BLS processing of the CPS treats earnings of "hourly" and "non-hourly" workers differently. In addition to the differential topcodes discussed above, a second important inconsistency is that the BLS reports hourly earnings for hourly workers as "straight time" pay, that is, *excluding* overtime, tips, and commissions, while, the BLS reports weekly earnings *including* overtime, tips, and commissions.

The most straightforward strategy to correct for this inconsistency is to add earnings from overtime, tips, and commissions to the straight hourly pay for hourly workers. In principle, the simplest way to do this would be to create a new hourly wage estimate for hourly workers using the BLS's "usual weekly earnings" (which includes overtime, tips, and commissions) and divide these weekly earnings by "usual weekly hours," in order to produce an hourly wage estimate that includes overtime, tips, and commissions. This is, in effect, the procedure followed for the NBER's Annual Earnings File wage variable. In practice, however, a large share of the resulting estimated hourly wages are actually lower than the reported hourly earnings excluding overtime, tips, and commissions (see Gao, 2003). A second approach, available only after the survey redesign in 1994, is to use new survey questions that specifically ask about earnings from overtime, tips, and commissions to add these earnings elements to straight hourly pay.<sup>11</sup> This procedure appears to work fairly well in practice, but is only available from 1994 on.

**Table 6** reports the results of a hybrid approach to adding overtime, tips, and commissions to straight hourly pay. For the period 1979-1993, the ORG data sets used here calculate hourly wages including overtime, tips, and commissions for "hourly" workers by dividing "usual weekly earnings" (including overtime, tips, and commissions) by "usual weekly hours." The resulting quotient becomes the hourly rate including overtime, tips, and commissions, however, only if this rate exceeds the "straight-time" rate; if the quotient produces an hourly rate below the "straight time" rate, the straight-time rate is taken to be the hourly rate including overtime, tips, and commissions. From 1994 on, the hybrid approach estimates the hourly rate including overtime, tips, and commissions, in two stages. The first stage is identical to the procedure used for 1979 to 1993. The second stage uses new information from the redesigned survey to supplement the first procedure. Specifically, if the reported information on weekly earnings from overtime, tips, and commissions, available in the redesigned survey, suggests a higher wage rate than estimated using the first overtime, tips, and commissions procedure, then the hybrid procedure replaces the hourly rate estimated in the first stage of the procedure with the new, higher, estimate.<sup>12</sup>

Table 6 examines three pairs of wages series; each pair has one series that includes and one series that excludes overtime, tips, and commissions for hourly workers. As a benchmark, the first two columns of Table 6 compare the w\_no\_no series, which makes no adjustment for topcoding and no adjustments for overtime, tips, and commissions, with the w\_nber series, which makes no adjustment for topcoding, but assigns all respondents an hourly wage equal to their usual weekly earnings –including overtime, tips, and commissions– divided by their usual weekly hours. The w\_nber series follows this simple procedure even when the resulting hourly wage estimate is lower than the reported hourly earnings excluding overtime, tips, and

commissions. The third and fourth columns of Table 6 compare the log-normally adjusted wage series with (column four) and without (column three) overtime, tips, and commissions, using the hybrid adjustment procedure described above. Before 1994, the main feature of the this hybrid procedure is that it only replaces the "straight-time" rate if the estimated overtime, tips, and commissions-inclusive rate is higher than the original straight pay. From 1994 on, the procedure also incorporates additional information on weekly earnings from overtime, tips, and commissions available in the redesigned survey. The fifth and sixth columns of the table compare the pareto-distribution-adjusted (80th percentile version) series with (column six) and without the identical hybrid adjustment.

The effects of the two overtime, tips, and commissions (OTC) procedures vary across three distinct periods of the ORG data. Between 1979 and 1988, neither OTC procedure has much impact on mean hourly earnings (see Table 6). Over this period, with few exceptions, the BLS constructs the usual weekly earnings of hourly workers by multiplying their straight-time hourly pay by their usual weekly hours, producing an estimate of usual weekly earnings that effectively excludes overtime, tips, and commissions. As a result, for all three pairs of wage series, the estimated overtime premium is zero or very close to zero (see the last three columns of the table). Between 1989 and 1993, the two OTC procedures consistently add 1.3-1.6% to the mean hourly wage, with little difference between the two approaches. From 1994 on, when the hybrid method begins to incorporate information on overtime, tips, and commissions available in the redesigned survey, however, the two OTC strategies give different results. Using the simple OTC procedure, the OTC premium at the mean varies between 2.7% and 3.1%; using the hybrid OTC method, the premium varies between 3.9% and 4.6% (see the last two columns).<sup>13</sup>

One way to check whether the OTC procedures are yielding sensible results is to compare average hourly earnings with and without overtime, tips, and commissions, across industries where overtime, tips, and commissions have a different weight in overall compensation. **Table 7** displays the average hourly earnings with and without overtime, tips, and commissions, for workers in retail trade, an industry where tips (in restaurants) and commissions are important; manufacturing, where overtime can be important; and education, where overtime, tips, and commissions are generally not important. To sharpen the contrast across the three sets of industries, the table reports the mean wage for hourly workers only, the group whose hourly earnings are affected by the OTC procedures. As with the overall numbers, the OTC effect varies systematically across the periods 1979-88, 1989-93, and 1994-2002. For the period 1979-88, the OTC effect is negligible for all three industry groups. For 1989-93, the OTC effect is largest in retail (about twice the average for manufacturing and education). From 1994 on, the OTC adjustment yields an 11-12% premium in retail, a 9-11% premium in manufacturing, and a 5-6% premium in education. Especially from 1994 on, these results suggest that the procedure is capturing an important aspect of differences in pay structure across industries.<sup>14</sup>

Overtime, tips, and commissions constitute an important part of compensation for many hourly workers. In a given cross-section, failure to include overtime, tips, and commissions lowers the relative wage of hourly workers relative to non-hourly workers, whose earnings generally include overtime, tips, and commissions. Adding overtime, tips, and commissions to straight hourly pay, however, is difficult. For ORG data from 1979 to 1988, standard procedures do not produce meaningful results. From 1989 through 1993, standard procedures appear to work better, but produce time series that are inconsistent with time series before and after the period. From 1994 on, the redesign of the CPS seems to allow for much better tracking of overtime, tips,

and commissions. Given these differences in the CPS over time, the best wage series to use appears to depend on the purpose of the particular analysis. If the analysis is fundamentally concerned with differences within a single cross section, then researchers should probably choose procedures that include overtime, tips, and commissions. The "hybrid" procedure here appears to produce better results than the simple procedure that is often employed. If changes over time are important, and the time frame is short enough to fall within a single overtime, tips, and commissions "regime" (1979-88, 1989-93, or 1994-2002), then the OTC adjusted series are, again, probably best, though the OTC adjustment has no meaningful impact for 1979-88, and only a limited effect for 1989-93. If changes over time are the primary concern and the time frame cuts across overtime, tips, and commission regimes, then, for consistency's sake, researchers should restrict themselves to series that exclude overtime, tips, and commissions for hourly workers.

### **"Hours vary"**

The CPS redesign in 1994 created at least one additional problem for the consistency of hourly wage series over time. Before 1994, CPS respondents were asked for their "usual weekly hours," but not permitted to answer that their hours varied. From 1994 on, the redesigned CPS allowed respondents to indicate that their "hours vary," without being more specific (though the vast majority who say that their hours vary do indicate whether they are usually "part-time" or "full-time"). As **Table 8** reports, from 1994, a sizeable share of workers (typically, 6-7%) chose to report that their hours vary. Since the distribution of hourly earnings for these workers may differ systematically from that of workers whose hours generally do not vary, simply excluding the group of workers whose hours vary may reduce comparability of wage series across the 1994 redesign.<sup>15</sup>

One strategy for reducing the potential for inconsistencies over time is to attempt to impute "usual weekly hours" for those who report that their hours vary.<sup>16</sup> **Table 9** shows the impact across the hourly earnings distribution of implementing a simple hours-imputation procedure.<sup>17</sup> The first two columns of the table display the mean real hourly wage with (column one) and without (column two) the hours imputation; the third column shows the percent difference between the two series. Before 1994, when the "hours vary" response was not permitted, the two series are, of course, identical. From 1994, the mean without adjusting for workers whose hours vary is consistently 0.4-0.8% higher than the series that does attempt to adjust for varying hours. The results in the table suggest that if the adjustment procedure is correct, excluding the hours varying workers might overstate average wage growth before and after the CPS redesign by about 0.5%.<sup>18</sup>

According to the table, wages at the 10th, 50th, and 90th percentiles are consistently slightly higher for the unadjusted series (which simply excludes all observations whose "hours vary") than they are for the wage series that includes the hours-vary observations after imputing usual hours. The differences in wages, however, are typically fairly small and not particularly systematic. At the 10th and 50th percentiles, the discrepancies vary from zero to four percent. At the 90th percentile, the differences are also small and often zero. The erratic nature of the differences, and the tendency for the differences to be slightly larger at the 10th than at the 50th percentiles, suggest that "clumping" of the underlying nominal wage levels (see below) may explain at least part of the discrepancies here.

Overall, the difference between the series that excludes all observations with varying hours and the series that impute usual weekly hours for those whose hours vary appear to be relatively small. These results suggest that researchers should be cautious about potential

problems caused by the introduction of the "hours vary" option in 1994, but that the two very different strategies for dealing with the discontinuity –excluding affected observations or imputing usual hours for them– are likely to yield similar results in most cases.

### **BLS-allocated earnings**

Consistent with general procedures first implemented in the CPS in 1962, the BLS "allocates" earnings responses to individuals in the ORG who have missing (or invalid) values for earnings (and other earnings-related variables). To allocate earnings, the BLS uses a procedure known as "cell hot decking," which involves the creation of a large matrix with cells based on demographic and work-related characteristics. The BLS fills each cell with the earnings of the most recent, valid ORG respondent who has the characteristics that correspond to those of the cell. When a CPS respondent has a missing value for earnings, the BLS assigns that individual the earnings of the entry in the hot-deck cell that corresponds to the individual's demographic and work-related characteristics.

Some researchers worry that using allocated observations may affect results of their analyses, particularly by overstating the wages of nonresponders, who may systematically earn less than the responders who supply entries to the hot-deck matrix. As a result, some researchers choose to exclude allocated wages from their analyses. Other researchers ignore the allocation procedure, essentially arguing that the successful implementation of the BLS's procedure does not change the data in ways that lead to significant biases in subsequent analysis. Unfortunately, neither approach –complete exclusion or complete inclusion– of the allocated earnings observations is entirely satisfactory. While this paper makes no specific recommendations on how researchers should proceed, this last section discusses some of the potential problems posed by the BLS allocation procedures and, especially, by changes in those procedures over time.

**Table 10** displays the share of workers that have had their wages allocated by the BLS.<sup>19</sup> The data in the table show large differences in the share of allocated earnings variables over the 1979-2002 period. The allocation levels group together in several distinct periods, each of which corresponds to different allocation "regimes" at the BLS. From 1979 to 1988, 13-16% of respondents had their earnings allocated by the BLS. During this period, the BLS used the same hot-deck procedure in each year, based on a hot-deck matrix with 11,232 cells defined by gender (2 groups), age (6), race (2), education (3), occupation (13), hours (6), and the receipt of overtime, tips, and commission.<sup>20</sup> If a respondent had missing or invalid responses to *any* earnings or earnings-related question, the BLS used entries in the matrix to impute *all* earnings and earnings-related responses.<sup>21</sup> The rationale for this procedure was to maintain a consistent relationship across earnings and earnings-related variables within each observation. In practice, this procedure meant that the BLS often replaced valid responses for some earnings and earnings-related variables with imputations drawn from the hot-deck matrix.<sup>22</sup>

Beginning in 1989, the BLS changed the hot-decking procedure to preserve all valid responses to earnings and earnings-related questions, and thereafter only allocated responses where information was missing or invalid. At first glance, the procedural change appears to have greatly reduced the share of observations with BLS-allocated earnings information. From 1989 through 1993, the share of workers with allocated earnings apparently never exceeded 3%. The main reason for the sharp decline, however, is that the CPS for these years reports allocations in only about one-fourth of the cases where they occurred (Hirsch and Schumacher, 2001).

The 1994 redesign of the CPS caused additional problems in 1994 and into 1995. The introduction of the "hours vary" response to the usual weekly hours question led the BLS to add two additional categories ("hours vary, usually part-time" and "hours vary, usually full-time) to

the six existing hours categories in the hot-deck matrix. The new categories expanded the matrix from 11,292 to 14,976 cells. The BLS was forced to wait until September 1995, however, before the expanded matrix had valid entries for all cells. As a result, from January 1994 through August 1995, the BLS did not allocate earnings or earnings-related variables to any ORG respondents. Thus, the allocation rate in 1994 was zero; and, in 1995, only about 8% (roughly, a 24% annualized allocation rate applied to only one-third of the annual data).

From 1996, the first full year of allocation using the post-1994 hot-deck matrix, the rate of allocation jumped to about one-quarter of respondents. Hirsch and Schumacher (2001) attribute the rise in allocations to the complexity of the redesigned CPS survey, which allows for complicated skip patterns and multiple consistency checks of responses. No clear explanation exists, however, for the large increase in allocations from 1996 (about 24%) to 2002 (about 34%).

The allocation procedure may cause difficulties for researchers interested in analyzing individual cross-sections of the ORG. In a given cross-section of the ORG, the hot-decking procedure may introduce at least three different kinds of biases: "between-cell nonresponse" bias; "within-cell nonresponse" bias; and "excluded-cell" bias.

Nonresponse bias stems from fundamental differences in the characteristics of nonresponders relative to responders. In the context of the ORG, many researchers fear that nonrespondents will have systematically lower earnings than respondents (see, for example, Lillard, Smith, and Welch, 1986). The hot-decking procedure is specifically designed to address the first type of nonresponse bias, "between-cell nonresponse." If, on average, nonrespondents have lower wages than respondents –because nonrespondents are disproportionately from groups with low average earnings (women, racial and ethnic minorities, the young, or the less educated,

for example)— then the under-representation of these groups in the final sample of earners will bias up average earnings relative to the true average for the complete working population. The hot-decking procedure should reduce this "between-cell" bias by reintroducing these nonrespondents into the earnings distribution with wages that are likely to be "close" to their true earnings (certainly closer to their true earnings than the average wage for respondents, for example).

Hot-decking, however, potentially introduces "within-cell nonresponse" bias. Within a given cell in the hot-deck matrix, nonrespondents may still have systematically lower earnings than nonrespondents. The hot-decking procedure would then, on average, assign too high a wage to nonrespondents in each cell. From the point of view of determining the average wage or the deciles of the overall distribution, within-cell nonresponse is likely to be small relative to untreated between-cell nonresponse bias. From the point of view of measuring inequality of earnings or using regression and other techniques to measure the wage "returns" to particular characteristics, however, "within-cell nonresponse" could bias analysis toward underreporting inequality and overstating the wage returns associated with typically lower-wage characteristics.

Depending on the analytical context, the third type of bias—"excluded-cell" bias—is potentially the most problematic. As Hirsch and Schumacher (2001) have observed, any attempt to measure the impact on earnings of variables that are *not* included in the hot-deck matrix will potentially be subject to significant measurement error. Hirsch and Schumacher motivate their discussion of this form of bias using the example of measuring the union wage premium. Since union status is not a determinant of the cell structure of the hot-deck matrix, any individual with missing earnings—and a valid union-status variable—will be assigned the earnings of an individual with matching hot-deck-matrix characteristics, but without regard to the donor's union

status. If about 10% of the workforce is unionized, this means that 90% of the time a unionized nonrespondent will be assigned earnings from an otherwise "identical" but nonunion worker. Similarly, about 10% of nonunion workers will be assigned earnings from an otherwise identical union worker. Hirsch and Schumacher demonstrate that, under fairly plausible conditions, this measurement error will reduce the measured return to these "excluded" characteristics by  $\Omega\Gamma$ , where  $\Omega$  is the share of allocated earnings in the full ORG and  $\Gamma$  is the "true" return to excluded characteristic. In practice, this means that in 2002, with about one-third of the earnings allocated, the "union wage premium" measured using the ORG, would understate the true wage premium by about 50%.<sup>23</sup> Hirsch and Schumacher emphasize that several features of the CPS ensure that their estimate of the bias probably acts as an upper bound, but the size of the bias is, nevertheless, worrisome. Moreover, such bias is likely to cause serious problems when comparing the union-wage premium (or conducting similar exercises) across periods with shares of allocated earnings variables that range from 0% to 33%. As Hirsch and Schumacher note, union status is not the only variable of interest to researchers that may be subject to this form of bias. Other variables that could potentially suffer from similar biases include: Hispanic status, immigrant status, marital status, presence and number of children, and city size (p. 18).

The preceding discussion of the BLS earnings-allocation procedures does not suggest a clear set of guidelines for researchers attempting to conduct consistent analyses of earnings over time. Researchers interested in analyzing earnings within a given cross-section of the ORG sometimes exclude allocated earnings. This avoids "within-cell nonresponse" bias, but at the cost of accentuating the "between-cell nonresponse" bias, which the hot-deck procedure likely reduces substantially. Since the BLS hot-decking procedure yields very different allocation rates over time, researchers who systematically exclude (or include) allocated observations are not, in

reality, following a "consistent" procedure over time, at least from the point of view of the likely impact on the consistency of the underlying data.

The effect of BLS allocations on standard types of earnings analysis with the CPS is an area that clearly requires further research. In the short-run, the best researchers can probably do is to check the robustness of any conclusions they draw from the CPS data to consideration of these potential biases in each particular circumstance.<sup>24</sup>

## **Conclusions**

This paper has reviewed a series of challenges facing any attempt to construct a consistent hourly wage series from the CPS ORG for the years 1979-2002. These difficulties include: topcoding of the weekly earnings variable; the unequal influence of outliers overtime; the inconsistent treatment of earnings from overtime, tips, and commissions; changes over time in the recording of usual weekly hours worked; and problems associated with "clumping" of reported nominal wages.

The analysis here suggests that the a consistent and robust hourly wage series for the full 1979-2002 period: (1) uses a log-normal imputation to adjust for top-coding; (2) excludes data below \$1 and above \$100 per hour (in constant 2002 dollars); (3) excludes overtime, tip, and commission earnings for hourly paid workers; and (4) uses a simple procedure to impute usual weekly hours for those who report their "hours vary" after 1994 (though the effects of not doing this are likely to be minor). For the period 1994-2002, the best hourly wage series follows the same procedure, but includes overtime, tips, and commissions for workers that report their earnings by the hour. Both these preferred series (as well as several variations) appear in the

Center for Economic and Policy Research's Uniform Data Files for the Outgoing Rotation Group and are available through [www.cepr.net](http://www.cepr.net).<sup>25</sup>

## Notes

I am grateful to Jean Roth of the NBER for assistance with the data and to Daniel Feenberg, Jean Roth, and others at the NBER for their efforts to make their programs, their documentation, and the CPS data all widely available. I am also grateful to Jared Bernstein, Danielle Gao, Larry Mishel, David Webster, and others at the Economic Policy Institute for extensive conversations and assistance over the years with the ORG data. Dean Baker, Heather Boushey, Eric Freeman, and Helene Jorgensen provided extensive comments on the CEPR ORG extract and the programs used to produce it. Finally, I acknowledge generous financial support from the Rockefeller Foundation for the CEPR's CPS Supplements project.

<sup>1</sup> The respondents to the wage question are referred to as the outgoing rotation group because CPS participants are in the survey four consecutive months, out of the survey for eight consecutive months, and then enter the survey again for four consecutive months. Respondents answer earnings-related questions in their fourth and eighth months in the survey. For extensive information on the CPS, see the Bureau of Labor Statistics CPS homepage: [www.bls.census.gov/cps/](http://www.bls.census.gov/cps/), especially Technical Paper 63 Revised, available in pdf format. For more information on the Outgoing Rotation Group, see the National Bureau of Economic Research's "CPS Annual Earnings File" page: <http://www.nber.org/data/morg.html>; the NBER's extract of the ORG is the underlying source for most of the data analyzed here. For a briefer discussion of many of the issues raised here, see Danielle Gao's "Wage Analysis Computations" appendix in Mishel, Bernstein, and Boushey (2003); as well as Webster, 2000. All programs used to create the data analyzed here are available at the Center for Economic and Policy Research web page ([www.cepr.net](http://www.cepr.net)); complete copies of the data on CD are available from the CEPR upon request.

<sup>2</sup> This paper documents key features of the data set pertaining to the hourly wage series. Other documentation on the CEPR web page discusses other aspects of the CEPR extracts used here.

<sup>3</sup> Before 1994, the ORG reports hourly earnings for "hourly paid" workers; from 1994 on, the ORG reports hourly earnings for any worker "hourly paid" or otherwise, for whom it was easiest to report earnings by the hour.

<sup>4</sup> This is the strategy taken by the NBER Annual Earnings File, which warns users about the topcode and changes in it over time, but makes no attempt to adjust for the level or changes in the topcode.

<sup>5</sup> The most prominent user of the pareto distribution for this purpose is the Economic Policy Institute. See, for example, Gao's appendix in Mishel, Bernstein, and Boushey (2003). For use of the pareto distribution in the context of topcoded annual income data, see Kuznets (1953), Feenberg and Poterba (1992), and Saez and Veall (2003).

<sup>6</sup> For a more complete description of the procedure, see Feenberg and Poterba (1992) or Saez and Veall (2003). For the sample program (in Stata) that implements the procedure used here, see `cepr_org_topcode_pareto.do`, available on the CEPR web page.

<sup>7</sup> For a clear discussion and an example, see Greene (2003).

<sup>8</sup> The divergence between the 90th percentile and the other estimates, in fact, begins in 1985, when the topcode cuts at the 94th percentile.

<sup>9</sup> EPI uses the 80th percentile pareto cutoff, but EPI's approach differs from the one used here because EPI estimates separate distribution for men and women above the topcode. While the gender-disaggregated approach makes sense, the procedure can produce the counter-intuitive result that, in some years, women above the topcode have a higher mean than do men above the topcode (see, for example, Mishel, Bernstein, and Boushey, 2003, Table B-2). The approach taken here implicitly assumes that the fundamental difference between men and women's weekly earnings is not their mean conditional on exceeding the topcode, but rather their probability of exceeding the topcode. In any event, the differences by gender in the estimated means above the topcode are generally relatively small.

<sup>10</sup> For detailed discussions of the 1994 CPS redesign, see Polivka and Miller (1995) and Technical Paper Revised (2002).

<sup>11</sup> The relevant variables in the redesigned version of the survey are: `peernuot`, `ptern`, `ptern2`, and `ptot`. These variables are not available in the NBER extracts. The CEPR extracts import these variables from the raw CPS data and match extra data to observations in the NBER extract from 1994 on. For a complete description of the procedure, see the programs `cepr_org_prep_otc.do` and `cepr_org_wages.do`, available on the CEPR's home page.

<sup>12</sup> As a safeguard against replacing valid straight hourly pay with erroneous estimated hourly pay including overtime, tips, and commissions, the procedure used here retains the straight hourly pay if the estimated hourly pay including overtime, tips, and commissions is more than four times the straight hourly pay.

<sup>13</sup> Appendix Tables A1, A2, and A3 present similar analyses of the effect of the two OTC adjustments at the 10th, 50th, and 90th percentiles of the hourly earnings distribution. As at the mean, the data suggest a different impact of the two procedures in each of the three periods 1979-88, 1989-93, and 1994-2002. The effects, however, are not as clear cut as they were with the mean because the tendency for workers to report hourly earnings in round numbers causes percentiles of the distribution to "clump" at whole-dollar increments (an issue addressed below). As would be expected, the effects of OTC adjustments are generally larger (in percentage terms) at the 10th and the 50th percentiles, than they are at the 90th percentile.

<sup>14</sup> Of course, the ORG fails to record non-wage forms of compensation, most importantly health and pension benefits.

<sup>15</sup> The NBER extract excludes workers whose "hours vary" by setting their earnings equal to missing.

<sup>16</sup> This is EPI's approach; see, for example, Webster, 1999, and Gao, 2003.

<sup>17</sup> For complete details, see the programs `cepr_org_hv_prep.do` and `cepr_org_hours.do`, both available on the CEPR web page. The procedure employed here uses information in the CPS to divide workers whose hours vary and whose hours don't vary into four categories: "full-time" men; "part-time" men; "full-time" women; and "part-time" men. The procedure then uses the data for workers whose hours don't vary to fit separate ordinary least square regressions for each of the four groups workers in each calendar year. These regressions estimate "usual weekly hours" as a function of the respondent's age, race, education level, marital status, and immigration and naturalization status. The procedure then uses the coefficients from these four sets of regressions to impute "usual weekly hours" for respondents with varying hours in each of the four groups. The procedure initially included information on the respondents' occupation and industry, but the large number of missing values for these variables and the relatively small additional information provided by the inclusion of the industry and occupation variables suggested that the simpler specification provided better results. Dividing the "hours vary" respondents into four categories (part- and full-time men and women), in fact, appears to cover a large share of the overall variation. Adding additional variables to the imputation regressions generally has the effect of raising the number of "hours varying" observations that can't be processed in the imputation procedure because of missing values for the additional regression variables. The EPI procedure, for example, uses many more variables (personal communication) and reports a much lower share of imputed observations –2.1-2.5% compared to 6-7% here (see Gao, 2003, Table B-3).

<sup>18</sup> Of course, the hours-imputation procedure used here could be flawed. The CPS, however, does not provide independent information that would allow us to assess the impact of the "varying hours" aspect of the redesign.

<sup>19</sup> For a description of the NBER variables used to designate allocated earnings and hours variables, see Feenberg and Roth, 2002. The percentages in Table 10 track those in Hirsch and Schumacher (2001, Table 1) fairly closely. For researchers who link the ORG to the May CPS for analysis of the 1973-1978 period, the BLS did not allocate earnings or earnings-related variables in the May CPS for those years. Observations that would have been allocated in the ORG from 1979 on were set to missing in the May CPS for 1973-1978 (Hirsch and Schumacher, 2001).

<sup>20</sup> This discussion of the history of the BLS's hot-decking procedures here draws heavily on Hirsch and Schumacher's (2001) excellent treatment of these issues.

<sup>21</sup> Each cell in the matrix has one entry, which has been "donated" by the most recent respondent to the CPS that has the characteristics that correspond to the respondent with missing information. Whenever the CPS survey finds a new individual fitting the characteristics of the cell, the cell's contents are updated. Earnings are always stored and assigned in nominal terms, but can correspond to earlier months or years, depending on when the CPS conducted a

successful interview with a person with that cell's characteristics. For a fuller discussion, see Hirsch and Schumacher (2001) and Bishop, Formby, and Thistle (1999).

<sup>22</sup> See Bishop, Formby, and Thistle (1999) for a detailed discussion of the 1979-88 procedure and some of the implications of it for analysis.

<sup>23</sup> The true union wage premium,  $w_u - w_n$ , is equal to  $\Gamma$ , or  $(w_u' - w_n')/(1-\Omega)$ , where  $(w_u' - w_n')$  is the (biased) measured premium.

<sup>24</sup> Another potential area of concern, not addressed here, is the possible impact of the large share of earnings and earnings-related information provided through "proxy" responses.

<sup>25</sup> The preferred series for 1979-2002 is `w_ln_no`; the preferred series for 1994-2002 is `w_ln_ot`.

## References

- Bishop, John A., John P. Formby, and Paul D. Thistle. 1999. "Mitigating Earnings Imputation Bias: Evidence from the CPS," unpublished paper, East Carolina University.
- Bureau of Labor Statistics and US Census Bureau. 2002. *Current Population Survey: Design and Methodology*. Technical Paper 63RV.
- Farber, Henry S. 1995. "Are Lifetime Jobs Disappearing? Job Duration in the United States: 1973-1993," Industrial Relations Section Working Paper No. 341 (January), Princeton University.
- Feenberg, Daniel R. and James M. Poterba. 1992. "Income inequality and the Incomes of Very High Income Taxpayers: Evidence from Tax Returns," NBER Working Paper No. 4229, Technical Appendix, pp. 28-30.
- Feenberg, Daniel and Jean Roth. 2002. "CPS Labor Extracts, 1979-2001," National Bureau of Economic Research, <http://www.nber.org/data/morg.html>
- Greene, William H. 2003. *Econometric Analysis* (5th edition), Upper Saddle River, New Jersey: Prentice Hall.
- Hirsch, Barry T. and Edward J. Schumacher. 2001. "Earnings Imputation and Bias in Wage Gap Estimates," unpublished paper, Trinity University.
- Kuznets, Simon. 1953. *Shares of Upper Income Groups in Income and Savings*, New York: National Bureau of Economic Research.
- Lillard, Lee, James P. Smith, and Finis Welch. 1986. "What Do We Really Know About Wages? The Importance of Nonreporting and Census Imputation," *Journal of Political Economy*, vol. 94, no. 3, (June), pp. 489-506.
- Mishel, Lawrence, Jared Bernstein, and Heather Boushey. 2003. *The State of Working America 2002-2003*. Ithaca, New York: Cornell University Press.
- Polivka, Anne E. and Stephen M. Miller. 1995. "The CPS After the Redesign: Refocusing the Economic Lens," unpublished paper, Bureau of Labor Statistics.
- Saez, Emmanuel and Michael R. Veall. 2003. "The Evolution of High Incomes in Canada," NBER Working Paper No. 9607, Appendix B.1., pp. 38-29.
- Webster, David. 2000. "Economic Policy Institute CPS ORG Labor Extracts, 1973-1999, Version 1.31" Washington, DC: Economic Policy Institute.

**APPENDIX TABLE A1: Effect of different overtime, tips, and commission procedures on 10th percentile real wage**  
(constant 2002 dollars)

Year	All						Overtime, tips, commissions premium (%)		
	w_no_no	w_nber	w_ln_no	w_ln_ot	w_p8_no	w_p8_ot	w_nber	w_ln_ot	w_p8_ot
1979	6.89	6.89	6.89	6.89	6.89	6.89	0.00	0.00	0.00
1980	6.51	6.51	6.51	6.51	6.51	6.51	0.00	0.00	0.00
1981	6.42	6.42	6.42	6.42	6.42	6.42	0.00	0.00	0.00
1982	6.24	6.24	6.24	6.24	6.24	6.24	0.00	0.00	0.00
1983	6.00	6.00	6.00	6.00	6.00	6.00	0.00	0.00	0.00
1984	5.93	5.93	5.93	5.93	5.93	5.93	0.00	0.00	0.00
1985	5.93	5.95	5.93	5.95	5.93	5.95	0.34	0.34	0.34
1986	5.87	5.89	5.87	5.89	5.87	5.89	0.34	0.34	0.34
1987	6.06	6.06	6.06	6.06	6.06	6.06	0.00	0.00	0.00
1988	5.84	5.84	5.84	5.84	5.84	5.84	0.00	0.00	0.00
1989	5.95	6.11	5.95	6.09	5.95	6.09	2.69	2.35	2.35
1990	6.00	6.14	6.00	6.10	6.00	6.10	2.33	1.67	1.67
1991	6.12	6.44	6.12	6.44	6.12	6.44	5.23	5.23	5.23
1992	6.28	6.28	6.28	6.28	6.28	6.28	0.00	0.00	0.00
1993	6.13	6.13	6.13	6.13	6.13	6.13	0.00	0.00	0.00
1994	6.00	6.00	6.00	6.00	6.00	6.00	0.00	0.00	0.00
1995	5.86	6.13	5.86	6.15	5.86	6.15	4.61	4.95	4.95
1996	5.94	6.24	5.94	6.28	5.94	6.28	5.05	5.72	5.72
1997	6.15	6.41	6.15	6.42	6.15	6.42	4.23	4.39	4.39
1998	6.61	6.61	6.61	6.61	6.61	6.61	0.00	0.00	0.00
1999	6.47	6.74	6.47	6.74	6.47	6.74	4.17	4.17	4.17
2000	6.68	6.79	6.68	6.89	6.68	6.87	1.65	3.14	2.84
2001	6.86	7.11	6.86	7.11	6.86	7.11	3.64	3.64	3.64
2002	7.00	7.00	7.00	7.00	7.00	7.00	0.00	0.00	0.00

Source: CEPR ORG extract. All distributions limited to range \$1-100 per hour.

**APPENDIX TABLE A1: Effect of different overtime, tips, and commission procedures on 10th percentile real wage (continued)**  
(constant 2002 dollars)

Year	Women						Men					
	w no no	w nber	w ln no	w ln ot	w p8 no	w p8 ot	w no no	w nber	w ln no	w ln ot	w p8 no	w p8 ot
1979	6.66	6.66	6.66	6.66	6.66	6.66	7.54	7.57	7.54	7.57	7.54	7.57
1980	6.41	6.41	6.41	6.41	6.41	6.41	7.23	7.23	7.23	7.23	7.23	7.23
1981	6.33	6.33	6.33	6.33	6.33	6.33	7.09	7.09	7.09	7.09	7.09	7.09
1982	5.98	5.98	5.98	5.98	5.98	5.98	6.96	6.97	6.96	6.96	6.96	6.96
1983	5.74	5.76	5.74	5.76	5.74	5.76	6.85	6.85	6.85	6.85	6.85	6.85
1984	5.52	5.60	5.52	5.60	5.52	5.60	6.59	6.59	6.59	6.59	6.59	6.59
1985	5.58	5.58	5.58	5.58	5.58	5.58	6.38	6.38	6.38	6.38	6.38	6.38
1986	5.48	5.48	5.48	5.48	5.48	5.48	6.64	6.64	6.63	6.63	6.63	6.63
1987	5.45	5.45	5.45	5.45	5.45	5.45	6.73	6.73	6.73	6.73	6.73	6.73
1988	5.48	5.48	5.48	5.48	5.48	5.48	6.57	6.59	6.57	6.59	6.57	6.59
1989	5.60	5.60	5.60	5.60	5.60	5.60	6.79	7.00	6.79	7.00	6.79	7.00
1990	5.51	5.67	5.51	5.67	5.51	5.67	6.67	6.67	6.67	6.67	6.67	6.67
1991	5.67	5.79	5.67	5.79	5.67	5.79	6.44	6.44	6.44	6.44	6.44	6.44
1992	5.65	5.97	5.65	5.97	5.65	5.97	6.28	6.33	6.28	6.28	6.28	6.28
1993	5.82	6.13	5.82	6.13	5.82	6.13	6.25	6.48	6.25	6.44	6.25	6.44
1994	5.70	6.00	5.70	6.00	5.70	6.00	6.30	6.60	6.30	6.61	6.30	6.61
1995	5.86	5.86	5.86	5.86	5.86	5.86	6.44	6.74	6.44	6.91	6.44	6.90
1996	5.71	5.71	5.71	5.71	5.71	5.71	6.56	6.85	6.56	6.85	6.56	6.85
1997	5.82	5.96	5.82	6.09	5.82	6.09	6.70	6.70	6.70	6.70	6.70	6.70
1998	6.06	6.34	6.06	6.34	6.06	6.34	6.89	7.13	6.89	7.16	6.89	7.16
1999	6.30	6.47	6.30	6.47	6.29	6.47	7.01	7.46	7.01	7.55	7.01	7.55
2000	6.27	6.47	6.27	6.53	6.27	6.53	7.31	7.31	7.31	7.31	7.31	7.31
2001	6.35	6.59	6.35	6.60	6.35	6.60	7.12	7.62	7.11	7.62	7.11	7.62
2002	6.50	6.73	6.50	6.85	6.50	6.85	7.30	7.50	7.30	7.53	7.26	7.50

Source: CEPR ORG extract. All distributions limited to range \$1-100 per hour.

**APPENDIX TABLE A2: Effect of different overtime, tips, and commission procedures on 50th percentile real wage**  
(constant 2002 dollars)

Year	All						Overtime, tips, commissions premium (%)		
	w no no	w nber	w ln no	w ln ot	w p8 no	w p8 ot	w nber	w ln ot	w p8 ot
1979	12.12	12.13	12.12	12.12	12.12	12.12	0.08	0.00	0.00
1980	12.09	12.09	12.09	12.09	12.09	12.09	0.00	0.00	0.00
1981	11.81	11.81	11.81	11.81	11.81	11.81	0.00	0.00	0.00
1982	11.89	11.89	11.89	11.89	11.89	11.89	0.00	0.00	0.00
1983	11.99	11.99	11.99	11.99	11.99	11.99	0.00	0.00	0.00
1984	11.92	11.93	11.92	11.92	11.92	11.92	0.08	0.00	0.00
1985	11.96	11.96	11.96	11.96	11.96	11.96	0.00	0.00	0.00
1986	12.36	12.36	12.36	12.36	12.36	12.36	0.00	0.00	0.00
1987	12.11	12.11	12.11	12.11	12.11	12.11	0.00	0.00	0.00
1988	12.21	12.22	12.21	12.21	12.20	12.21	0.08	0.00	0.08
1989	12.25	12.38	12.25	12.32	12.25	12.32	1.06	0.57	0.57
1990	12.01	12.31	12.01	12.27	12.01	12.27	2.50	2.16	2.16
1991	12.23	12.39	12.23	12.39	12.23	12.39	1.31	1.31	1.31
1992	12.31	12.56	12.31	12.56	12.31	12.56	2.03	2.03	2.03
1993	12.26	12.26	12.26	12.26	12.26	12.26	0.00	0.00	0.00
1994	12.00	12.30	12.00	12.60	12.00	12.60	2.50	5.00	5.00
1995	11.72	12.39	11.72	12.61	11.72	12.61	5.72	7.59	7.59
1996	11.85	12.55	11.85	12.55	11.85	12.55	5.91	5.91	5.91
1997	12.23	12.59	12.23	12.85	12.23	12.85	2.94	5.07	5.07
1998	12.48	13.22	12.45	13.22	12.40	13.22	5.93	6.18	6.61
1999	12.95	13.27	12.95	13.49	12.95	13.49	2.47	4.17	4.17
2000	12.79	13.37	12.79	13.58	12.75	13.58	4.53	6.18	6.51
2001	13.21	13.66	13.21	13.90	13.18	13.82	3.41	5.22	4.86
2002	13.14	13.86	13.14	14.00	13.01	14.00	5.48	6.54	7.61

Source: CEPR ORG extract. All distributions limited to range \$1-100 per hour.

**APPENDIX TABLE A2: Effect of different overtime, tips, and commission procedures on 50th percentile real wage (continued)**  
(constant 2002 dollars)

Year	Women						Men					
	w no no	w nber	w ln no	w ln ot	w p8 no	w p8 ot	w no no	w nber	w ln no	w ln ot	w p8 no	w p8 ot
1979	9.55	9.56	9.55	9.56	9.55	9.56	15.30	15.30	15.30	15.30	15.30	15.30
1980	9.46	9.46	9.46	9.46	9.46	9.46	14.99	15.01	14.99	15.01	14.99	15.01
1981	9.45	9.45	9.45	9.45	9.45	9.45	14.74	14.74	14.74	14.74	14.74	14.74
1982	9.51	9.51	9.51	9.51	9.51	9.51	14.50	14.51	14.50	14.50	14.50	14.50
1983	9.63	9.63	9.63	9.63	9.63	9.63	14.56	14.56	14.56	14.56	14.56	14.56
1984	9.88	9.88	9.88	9.88	9.88	9.88	14.41	14.41	14.41	14.41	14.41	14.41
1985	9.96	9.96	9.96	9.96	9.96	9.96	14.35	14.35	14.35	14.35	14.35	14.35
1986	10.18	10.18	10.18	10.18	10.18	10.18	14.87	14.87	14.87	14.87	14.87	14.87
1987	10.32	10.33	10.32	10.32	10.32	10.32	15.14	15.14	15.14	15.14	15.14	15.14
1988	10.22	10.22	10.22	10.22	10.22	10.22	14.61	14.61	14.61	14.61	14.61	14.61
1989	10.50	10.50	10.50	10.50	10.50	10.50	14.01	14.01	14.01	14.01	14.01	14.01
1990	10.41	10.67	10.41	10.67	10.41	10.67	14.01	14.23	14.01	14.21	14.01	14.20
1991	10.30	10.62	10.30	10.56	10.30	10.56	13.95	14.16	13.95	14.16	13.95	14.16
1992	10.55	10.77	10.55	10.74	10.55	10.73	13.82	14.07	13.82	14.02	13.82	14.01
1993	10.72	10.90	10.72	10.88	10.72	10.88	13.74	13.97	13.74	13.93	13.74	13.93
1994	10.51	10.91	10.51	11.02	10.51	11.02	13.44	14.20	13.44	14.40	13.43	14.40
1995	10.55	10.90	10.55	10.99	10.55	10.99	13.51	14.06	13.51	14.37	13.51	14.36
1996	10.51	10.97	10.51	11.13	10.51	11.13	13.69	14.26	13.69	14.26	13.69	14.26
1997	10.73	11.17	10.73	11.17	10.73	11.17	13.51	14.25	13.51	14.53	13.50	14.53
1998	11.02	11.40	11.02	11.43	11.02	11.43	14.15	14.82	14.12	15.01	14.05	14.87
1999	11.16	11.82	11.14	11.87	11.14	11.87	14.51	15.11	14.51	15.49	14.48	15.43
2000	11.49	11.82	11.49	11.92	11.48	11.90	14.62	15.19	14.62	15.66	14.62	15.62
2001	11.68	12.19	11.68	12.19	11.66	12.19	14.73	15.24	14.73	15.62	14.73	15.58
2002	12.00	12.30	12.00	12.35	12.00	12.31	15.00	15.38	15.00	15.56	15.00	15.50

Source: CEPR ORG extract. All distributions limited to range \$1-100 per hour.

**APPENDIX TABLE A3: Effect of different overtime, tips, and commission procedures on 90th percentile real wage**  
(constant 2002 dollars)

Year	All						Overtime, tips, commissions premium (%)		
	w no no	w nber	w ln no	w ln ot	w p8 no	w p8 ot	w nber	w ln ot	w p8 ot
1979	23.64	23.64	23.64	23.64	23.64	23.64	0.00	0.00	0.00
1980	23.36	23.40	23.38	23.39	23.38	23.38	0.17	0.04	0.00
1981	23.62	23.62	23.62	23.62	23.62	23.62	0.00	0.00	0.00
1982	23.78	23.79	23.79	23.79	23.79	23.79	0.04	0.00	0.00
1983	23.98	23.98	23.98	23.98	23.98	23.98	0.00	0.00	0.00
1984	24.71	24.71	24.71	24.71	24.71	24.71	0.00	0.00	0.00
1985	24.12	24.13	24.26	24.26	24.27	24.27	0.04	0.00	0.00
1986	25.05	25.05	25.05	25.05	25.05	25.05	0.00	0.00	0.00
1987	25.23	25.21	25.61	25.61	25.74	25.74	-0.08	0.00	0.00
1988	25.56	25.56	25.56	25.56	25.56	25.56	0.00	0.00	0.00
1989	25.21	25.77	25.21	25.71	25.21	25.70	2.22	1.98	1.94
1990	25.65	26.02	25.65	26.02	25.65	26.02	1.44	1.44	1.44
1991	25.75	25.75	25.75	25.75	25.75	25.75	0.00	0.00	0.00
1992	25.12	25.56	25.12	25.50	25.12	25.46	1.75	1.51	1.35
1993	25.74	26.12	25.74	26.04	25.74	26.04	1.48	1.17	1.17
1994	26.07	26.95	26.09	27.40	26.05	27.40	3.38	5.02	5.18
1995	26.06	27.02	26.07	27.04	26.05	27.04	3.68	3.72	3.80
1996	26.32	27.10	26.32	27.42	26.32	27.42	2.96	4.18	4.18
1997	26.82	27.07	26.82	27.93	26.82	27.87	0.93	4.14	3.91
1998	27.55	27.93	27.55	28.23	27.23	27.91	1.38	2.47	2.50
1999	28.00	28.78	27.84	29.04	27.66	28.98	2.79	4.31	4.77
2000	28.44	29.30	28.39	29.92	28.11	29.40	3.02	5.39	4.59
2001	29.28	30.02	29.28	30.39	29.28	29.99	2.53	3.79	2.42
2002	29.91	30.59	29.91	30.76	29.22	30.27	2.27	2.84	3.59

Source: CEPR ORG extract. All distributions limited to range \$1-100 per hour.

**APPENDIX TABLE A3: Effect of different overtime, tips, and commission procedures on 90th percentile real wage (continued)**  
(constant 2002 dollars)

Year	Women						Men					
	w_no no	w_nber	w_ln no	w_ln ot	w_p8 no	w_p8 ot	w_no no	w_nber	w_ln no	w_ln ot	w_p8 no	w_p8 ot
1979	17.22	17.22	17.22	17.22	17.22	17.22	27.54	27.54	27.54	27.54	27.54	27.54
1980	17.45	17.46	17.45	17.45	17.45	17.45	26.25	26.25	26.25	26.25	26.29	26.30
1981	17.76	17.76	17.76	17.76	17.76	17.76	26.45	26.45	26.55	26.55	26.57	26.57
1982	17.84	17.84	17.84	17.84	17.84	17.84	26.76	26.76	26.76	26.76	26.76	26.76
1983	18.41	18.41	18.41	18.41	18.41	18.41	27.58	27.58	27.83	27.83	27.83	27.83
1984	18.82	18.82	18.82	18.82	18.82	18.82	27.59	27.59	28.00	28.00	28.00	28.00
1985	19.30	19.31	19.31	19.31	19.30	19.31	27.90	27.90	28.34	28.34	28.37	28.37
1986	19.57	19.57	19.57	19.57	19.57	19.57	28.37	28.26	29.35	29.35	29.35	29.35
1987	20.16	20.16	20.17	20.17	20.15	20.15	28.49	28.39	30.28	30.28	30.28	30.28
1988	20.56	20.58	20.59	20.59	20.58	20.58	28.99	28.83	29.21	29.21	29.65	29.65
1989	21.01	21.01	21.01	21.01	21.01	21.01	28.01	28.29	28.01	28.29	28.01	28.26
1990	21.07	21.35	21.07	21.35	21.05	21.35	28.85	29.11	28.85	29.11	28.85	29.11
1991	21.46	21.89	21.46	21.83	21.46	21.83	28.94	28.97	28.96	28.97	28.89	28.97
1992	21.98	21.98	21.98	21.98	21.98	21.98	28.89	28.99	28.89	28.99	28.89	28.99
1993	22.15	22.65	22.15	22.58	22.15	22.58	29.11	29.41	29.11	29.41	29.11	29.41
1994	22.81	23.08	22.81	23.35	22.81	23.34	28.84	30.01	28.84	30.01	28.84	30.01
1995	22.53	23.08	22.53	23.44	22.53	23.44	29.21	29.49	29.30	30.02	29.30	30.02
1996	22.82	23.02	22.82	23.10	22.82	23.09	28.53	29.67	28.56	30.16	28.53	30.16
1997	23.07	23.46	23.07	23.63	23.04	23.63	29.33	30.11	29.53	30.71	29.50	30.70
1998	23.83	24.24	23.71	24.27	23.63	24.24	30.61	31.76	30.41	31.76	29.91	31.55
1999	24.30	24.88	24.28	24.90	24.28	24.90	31.11	32.28	31.11	32.37	31.11	32.23
2000	24.75	25.09	24.75	25.09	24.59	25.09	32.12	32.63	32.12	33.33	31.37	32.63
2001	25.40	25.40	25.35	25.40	25.11	25.40	32.56	33.86	32.56	34.19	32.08	33.41
2002	25.48	26.00	25.48	26.25	25.17	26.00	33.65	34.18	33.65	34.60	32.68	33.65

Source: CEPR ORG extract. All distributions limited to range \$1-100 per hour.

**TABLE 1: CPS topcode and estimated mean above the topcode**  
(Weekly earnings in nominal dollars)

Year	CPS topcode	Share of all weekly earners topcoded (%)	Imputed mean above the CPS topcode			
			Log-normal	Pareto from:		
				70th	80th	90th
1979	999	1.3	1,242	1,495	1,471	1,462
1980	999	1.6	1,243	1,475	1,454	1,413
1981	999	2.3	1,258	1,529	1,469	1,479
1982	999	3.2	1,276	1,545	1,532	1,443
1983	999	4.2	1,298	1,549	1,547	1,492
1984	999	5.2	1,308	1,647	1,602	1,504
1985	999	6.1	1,316	1,674	1,564	1,806
1986	999	7.4	1,334	1,648	1,581	1,895
1987	999	8.7	1,347	1,700	1,696	2,884
1988	999	10.3	1,367	1,915	1,765	.
1989	1,923	1.1	2,394	2,866	2,801	2,738
1990	1,923	1.3	2,401	2,851	2,848	2,852
1991	1,923	1.6	2,417	2,948	2,855	2,872
1992	1,923	1.8	2,421	2,887	2,805	2,814
1993	1,923	2.0	2,422	2,894	2,854	2,767
1994	1,923	2.8	2,489	3,128	2,992	3,044
1995	1,923	3.1	2,497	3,116	3,064	3,044
1996	1,923	3.3	2,493	3,113	3,016	3,054
1997	1,923	3.9	2,511	3,158	3,085	3,092
1998	2,884	1.5	3,709	4,725	4,652	4,661
1999	2,884	1.7	3,713	4,691	4,611	4,709
2000	2,884	2.0	3,721	4,822	4,700	4,685
2001	2,884	2.2	3,727	4,808	4,627	4,595
2002	2,884	2.5	3,760	4,861	4,790	4,812

Source: CEPR ORG extract.

**TABLE 2: Effect of different top-coding procedures on mean real wage, no trimming**  
(constant 2002 dollars)

Year	All						Annual change (%)					
	w_nber	w_no no	w_ln no	w_p7 no	w_p8 no	w_p9 no	w_nber	w_no no	w_ln no	w_p7 no	w_p8 no	w_p9 no
1979	14.29	14.29	14.36	14.42	14.42	14.42	--	--	--	--	--	--
1980	13.97	13.98	14.06	14.12	14.12	14.12	-2.24	-2.17	-2.09	-2.08	-2.08	-2.08
1981	13.85	13.86	13.97	14.08	14.05	14.05	-0.86	-0.86	-0.64	-0.28	-0.50	-0.50
1982	13.88	13.89	14.04	14.18	14.17	14.18	0.22	0.22	0.50	0.71	0.85	0.93
1983	13.84	13.85	14.05	14.22	14.22	14.22	-0.29	-0.29	0.07	0.28	0.35	0.28
1984	13.84	13.86	14.11	14.37	14.34	14.34	0.00	0.07	0.43	1.05	0.84	0.84
1985	13.95	13.97	14.26	14.58	14.48	14.48	0.79	0.79	1.06	1.46	0.98	0.98
1986	14.18	14.20	14.56	14.89	14.82	14.82	1.65	1.65	2.10	2.13	2.35	2.35
1987	14.19	14.21	14.62	15.04	15.04	15.04	0.07	0.07	0.41	1.01	1.48	1.48
1988	14.19	14.21	14.71	15.46	15.25	.	0.00	0.00	0.62	2.79	1.40	.
1989	14.69	14.42	14.48	14.54	14.53	14.74	3.52	1.48	-1.56	-5.95	-4.72	.
1990	14.69	14.44	14.51	14.58	14.57	14.78	0.00	0.14	0.21	0.28	0.28	0.27
1991	14.68	14.44	14.52	14.62	14.60	14.81	-0.07	0.00	0.07	0.27	0.21	0.20
1992	14.72	14.45	14.54	14.63	14.61	14.82	0.27	0.07	0.14	0.07	0.07	0.07
1993	14.76	14.51	14.60	14.70	14.69	14.90	0.27	0.42	0.41	0.48	0.55	0.54
1994	14.96	14.59	14.74	14.91	14.88	15.55	1.36	0.55	0.96	1.43	1.29	4.36
1995	14.92	14.48	14.64	14.82	14.80	15.46	-0.27	-0.75	-0.68	-0.60	-0.54	-0.58
1996	14.92	14.47	14.64	14.82	14.79	15.43	0.00	-0.07	0.00	0.00	-0.07	-0.19
1997	15.14	14.70	14.91	15.14	15.11	15.74	1.47	1.59	1.84	2.16	2.16	2.01
1998	15.82	15.35	15.47	15.61	15.60	16.23	4.49	4.42	3.76	3.10	3.24	3.11
1999	16.20	15.71	15.84	15.99	15.98	16.63	2.40	2.35	2.39	2.43	2.44	2.46
2000	16.36	15.88	16.03	16.22	16.19	16.86	0.99	1.08	1.20	1.44	1.31	1.38
2001	16.69	16.18	16.34	16.54	16.51	17.19	2.02	1.89	1.93	1.97	1.98	1.96
2002	16.90	16.42	16.61	16.85	16.83	17.49	1.26	1.48	1.65	1.87	1.94	1.75

Source: CEPR ORG extract.

**TABLE 2: Effect of different top-coding procedures on mean real wage, no trimming (continued)**  
(constant 2002 dollars)

Year	Women						Men					
	w_nber	w_no no	w_ln no	w_p7 no	w_p8 no	w_p9 no	w_nber	w_no no	w_ln no	w_p7 no	w_p8 no	w_p9 no
1979	11.07	11.06	11.07	11.07	11.07	11.07	16.76	16.77	16.88	17.00	16.98	16.98
1980	10.92	10.92	10.93	10.93	10.93	10.93	16.40	16.42	16.55	16.67	16.65	16.63
1981	10.92	10.92	10.93	10.94	10.93	10.93	16.23	16.25	16.43	16.62	16.57	16.58
1982	11.09	11.09	11.11	11.13	11.12	11.12	16.20	16.22	16.48	16.72	16.71	16.63
1983	11.18	11.18	11.21	11.23	11.23	11.22	16.07	16.09	16.44	16.74	16.73	16.67
1984	11.28	11.28	11.33	11.38	11.37	11.36	15.99	16.02	16.43	16.88	16.82	16.69
1985	11.44	11.45	11.51	11.58	11.56	11.61	16.08	16.11	16.59	17.13	16.96	17.32
1986	11.73	11.73	11.81	11.89	11.87	11.95	16.30	16.32	16.92	17.48	17.36	17.92
1987	11.89	11.89	11.99	12.10	12.10	12.45	16.19	16.23	16.92	17.62	17.61	19.96
1988	12.01	12.02	12.15	12.35	12.29	.	16.09	16.13	16.96	18.19	17.85	.
1989	12.42	12.12	12.13	12.14	12.14	12.14	16.69	16.46	16.56	16.67	16.65	16.64
1990	12.50	12.23	12.24	12.25	12.25	12.25	16.65	16.40	16.53	16.65	16.65	16.65
1991	12.64	12.40	12.42	12.44	12.44	12.44	16.51	16.28	16.42	16.58	16.55	16.56
1992	12.84	12.56	12.59	12.62	12.62	12.62	16.43	16.17	16.32	16.46	16.44	16.44
1993	12.98	12.70	12.73	12.75	12.75	12.75	16.40	16.16	16.32	16.47	16.46	16.43
1994	13.18	12.89	12.94	13.01	13.00	13.00	16.58	16.13	16.37	16.64	16.58	16.60
1995	13.09	12.72	12.77	12.83	12.83	12.83	16.59	16.07	16.33	16.61	16.59	16.58
1996	13.13	12.77	12.83	12.90	12.89	12.89	16.57	16.02	16.29	16.58	16.53	16.55
1997	13.36	13.00	13.08	13.18	13.17	13.17	16.78	16.25	16.57	16.92	16.88	16.89
1998	13.87	13.47	13.51	13.57	13.57	13.57	17.61	17.07	17.25	17.46	17.45	17.45
1999	14.15	13.74	13.79	13.84	13.83	13.84	18.09	17.53	17.73	17.97	17.95	17.98
2000	14.35	13.91	13.96	14.02	14.01	14.01	18.22	17.70	17.93	18.24	18.20	18.20
2001	14.68	14.27	14.33	14.41	14.40	14.39	18.56	17.94	18.19	18.51	18.45	18.45
2002	14.98	14.56	14.64	14.73	14.72	14.73	18.69	18.14	18.44	18.81	18.79	18.80

Source: CEPR ORG extract.

**TABLE 3: Effect of different top-coding procedures on the standard deviation of the real wage, no trimming**  
(constant 2002 dollars)

Year	All						Annual change (%)					
	w_nber	w_no no	w_ln no	w_p7 no	w_p8 no	w_p9 no	w_nber	w_no no	w_ln no	w_p7 no	w_p8 no	w_p9 no
1979	8.69	8.74	9.05	9.45	9.41	9.39	--	--	--	--	--	--
1980	7.95	8.12	8.44	8.83	8.79	8.72	-8.52	-7.09	-6.74	-6.56	-6.59	-7.14
1981	8.86	8.97	9.35	9.86	9.74	9.76	11.45	10.47	10.78	11.66	10.81	11.93
1982	8.03	8.14	8.69	9.36	9.33	9.10	-9.37	-9.25	-7.06	-5.07	-4.21	-6.76
1983	8.49	8.58	9.28	10.01	10.00	9.83	5.73	5.41	6.79	6.94	7.18	8.02
1984	8.03	8.18	9.01	10.16	10.00	9.65	-5.42	-4.66	-2.91	1.50	0.00	-1.83
1985	8.25	8.40	9.46	10.93	10.45	11.53	2.74	2.69	4.99	7.58	4.50	19.48
1986	8.13	8.21	9.33	10.63	10.33	11.77	-1.45	-2.26	-1.37	-2.74	-1.15	2.08
1987	8.24	8.33	9.60	11.20	11.18	17.67	1.35	1.46	2.89	5.36	8.23	50.13
1988	8.54	8.63	10.01	12.68	11.89	.	3.64	3.60	4.27	13.21	6.35	.
1989	10.28	9.24	9.56	9.96	9.90	9.84	20.37	7.07	-4.50	-21.45	-16.74	.
1990	10.71	9.67	10.01	10.41	10.41	10.41	4.18	4.65	4.71	4.52	5.15	5.79
1991	9.62	9.11	9.54	10.12	10.01	10.03	-10.18	-5.79	-4.70	-2.79	-3.84	-3.65
1992	11.03	10.03	10.44	10.92	10.83	10.84	14.66	10.10	9.43	7.91	8.19	8.08
1993	9.91	9.13	9.56	10.07	10.02	9.93	-10.15	-8.97	-8.43	-7.78	-7.48	-8.39
1994	11.05	10.69	11.51	12.64	12.39	12.48	11.50	17.09	20.40	25.52	23.65	25.68
1995	10.02	9.57	10.28	11.23	11.14	11.11	-9.32	-10.48	-10.69	-11.16	-10.09	-10.98
1996	9.65	9.48	10.17	11.11	10.95	11.01	-3.69	-0.94	-1.07	-1.07	-1.71	-0.90
1997	10.16	9.66	10.48	11.63	11.49	11.50	5.28	1.90	3.05	4.68	4.93	4.45
1998	11.03	10.68	11.34	12.35	12.27	12.28	8.56	10.56	8.21	6.19	6.79	6.78
1999	11.32	11.16	11.83	12.82	12.73	12.84	2.63	4.49	4.32	3.81	3.75	4.56
2000	11.14	11.00	11.75	12.99	12.84	12.82	-1.59	-1.43	-0.68	1.33	0.86	-0.16
2001	12.31	11.98	12.72	13.91	13.70	13.66	10.50	8.91	8.26	7.08	6.70	6.55
2002	12.27	12.00	12.88	14.27	14.18	14.21	-0.32	0.17	1.26	2.59	3.50	4.03

Source: CEPR ORG extract.

**TABLE 3: Effect of different top-coding procedures on mean real wage, no trimming (continued)**  
(constant 2002 dollars)

Year	Women						Men					
	w_nber	w_no no	w_ln no	w_p7 no	w_p8 no	w_p9 no	w_nber	w_no no	w_ln no	w_p7 no	w_p8 no	w_p9 no
1979	7.34	7.35	7.38	7.42	7.42	7.42	8.83	8.91	9.39	10.01	9.95	9.93
1980	5.97	5.99	6.04	6.10	6.09	6.08	8.48	8.74	9.22	9.79	9.73	9.63
1981	6.41	6.43	6.49	6.58	6.56	6.56	9.81	9.96	10.51	11.25	11.07	11.10
1982	6.12	6.14	6.24	6.37	6.37	6.32	8.66	8.84	9.64	10.61	10.56	10.22
1983	7.13	7.17	7.29	7.43	7.43	7.39	8.88	9.01	10.06	11.14	11.13	10.88
1984	6.50	6.56	6.78	7.10	7.06	6.96	8.54	8.75	9.94	11.56	11.33	10.85
1985	6.45	6.53	6.82	7.26	7.12	7.45	8.98	9.18	10.68	12.73	12.07	13.55
1986	6.72	6.73	7.10	7.54	7.44	7.95	8.64	8.75	10.32	12.11	11.71	13.67
1987	7.10	7.13	7.71	8.47	8.46	11.71	8.63	8.76	10.46	12.58	12.55	20.95
1988	7.89	7.93	8.45	9.53	9.21	.	8.62	8.77	10.71	14.35	13.30	.
1989	9.87	7.67	7.73	7.81	7.80	7.79	10.23	10.01	10.50	11.10	11.01	10.93
1990	9.88	8.03	8.12	8.21	8.21	8.21	11.04	10.53	11.05	11.65	11.65	11.65
1991	8.58	7.62	7.77	7.97	7.93	7.94	10.12	9.91	10.54	11.37	11.21	11.24
1992	10.87	9.95	10.12	10.32	10.29	10.29	10.89	9.79	10.41	11.12	10.99	11.00
1993	9.37	7.75	7.90	8.08	8.06	8.02	10.11	9.95	10.58	11.31	11.24	11.10
1994	10.60	10.26	10.87	11.70	11.51	11.58	11.19	10.83	11.82	13.21	12.89	13.01
1995	8.85	8.66	9.00	9.47	9.43	9.41	10.71	10.07	11.04	12.33	12.22	12.17
1996	8.39	8.36	8.72	9.22	9.13	9.17	10.42	10.15	11.07	12.33	12.12	12.20
1997	8.77	8.72	9.18	9.83	9.75	9.76	11.05	10.19	11.29	12.79	12.61	12.63
1998	9.62	9.24	9.62	10.21	10.16	10.16	11.91	11.57	12.44	13.77	13.67	13.68
1999	9.63	9.48	9.83	10.36	10.31	10.37	12.39	12.24	13.13	14.45	14.33	14.47
2000	9.65	9.44	9.82	10.47	10.39	10.38	12.07	11.98	12.99	14.65	14.45	14.42
2001	11.92	11.65	12.02	12.62	12.51	12.49	12.38	12.01	13.06	14.74	14.44	14.38
2002	11.19	10.86	11.30	12.03	11.98	11.99	12.94	12.73	13.94	15.83	15.70	15.74

Source: CEPR ORG extract.

**TABLE 4: Effect of different trimming procedures on the mean of the real wage for all workers**  
(constant 2002 dollars)

Year	w nber			w ln no			w p8 no		
	\$1-50	\$1-100	\$1-200	\$1-50	\$1-100	\$1-200	\$1-50	\$1-100	\$1-200
1979	14.05	14.26	14.28	14.00	14.32	14.35	13.98	14.38	14.41
1980	13.79	13.95	13.97	13.74	14.02	14.05	13.70	14.08	14.11
1981	13.78	13.82	13.84	13.65	13.92	13.95	13.56	14.00	14.04
1982	13.82	13.86	13.88	13.65	14.01	14.04	13.56	14.14	14.18
1983	13.77	13.81	13.83	13.62	14.01	14.05	13.45	14.18	14.21
1984	13.79	13.83	13.85	13.66	14.07	14.11	13.46	14.29	14.34
1985	13.89	13.94	13.95	13.75	14.22	14.26	13.71	14.44	14.48
1986	14.13	14.18	14.19	13.97	14.53	14.56	13.91	14.79	14.82
1987	14.13	14.17	14.19	13.99	14.59	14.62	13.67	15.00	15.03
1988	14.11	14.16	14.17	14.54	14.67	14.70	13.60	15.20	15.24
1989	14.27	14.63	14.66	14.00	14.46	14.48	14.00	14.51	14.54
1990	14.25	14.61	14.64	14.00	14.47	14.49	13.99	14.53	14.56
1991	14.37	14.64	14.67	14.05	14.51	14.52	14.04	14.58	14.60
1992	14.38	14.66	14.68	14.04	14.51	14.53	14.03	14.58	14.60
1993	14.46	14.73	14.75	14.16	14.60	14.61	14.10	14.68	14.69
1994	14.56	14.90	14.96	14.02	14.66	14.75	13.94	14.78	14.88
1995	14.61	14.92	14.95	14.04	14.63	14.68	13.95	14.77	14.84
1996	14.66	14.92	14.94	14.05	14.63	14.67	13.97	14.78	14.82
1997	14.82	15.13	15.16	14.21	14.88	14.93	14.11	15.07	15.13
1998	15.10	15.78	15.83	14.64	15.32	15.48	14.64	15.18	15.60
1999	15.43	16.14	16.21	14.93	15.66	15.85	14.93	15.67	15.98
2000	15.53	16.33	16.39	15.05	15.98	16.05	15.04	15.86	16.21
2001	15.86	16.64	16.70	15.31	16.28	16.35	15.30	16.13	16.52
2002	16.05	16.85	16.91	15.51	16.55	16.61	15.51	16.37	16.83

Source: CEPR ORG extract.

**TABLE 5: Effect of different trimming procedures on standard deviation of real wage for all workers**  
(constant 2002 dollars)

Year	w nber			w ln no			w p8 no		
	\$1-50	\$1-100	\$1-200	\$1-50	\$1-100	\$1-200	\$1-50	\$1-100	\$1-200
1979	7.34	7.94	8.10	7.26	8.28	8.47	7.21	8.66	8.85
1980	7.27	7.69	7.84	7.19	8.04	8.26	7.12	8.40	8.62
1981	7.46	7.61	7.78	7.25	8.05	8.33	7.08	8.50	8.77
1982	7.61	7.76	7.91	7.39	8.34	8.58	7.23	9.00	9.23
1983	7.68	7.84	7.99	7.56	8.60	8.83	7.25	9.34	9.58
1984	7.67	7.81	7.89	7.62	8.64	8.89	7.28	9.60	9.86
1985	7.70	7.85	7.97	7.67	8.80	9.06	7.74	9.67	9.93
1986	7.79	7.97	8.07	7.82	9.09	9.26	7.85	10.08	10.27
1987	7.75	7.89	8.02	7.85	9.16	9.32	7.39	10.67	10.84
1988	7.69	7.85	7.99	8.99	9.33	9.49	7.36	11.19	11.39
1989	8.00	8.93	9.14	7.92	9.22	9.37	7.91	9.54	9.72
1990	7.99	8.89	9.08	7.93	9.22	9.38	7.90	9.60	9.80
1991	8.21	8.91	9.06	7.99	9.29	9.41	7.98	9.71	9.86
1992	8.24	8.92	9.08	8.01	9.30	9.39	7.99	9.68	9.80
1993	8.29	8.97	9.07	8.16	9.36	9.44	8.05	9.81	9.91
1994	8.43	9.25	9.60	8.19	10.00	10.47	8.03	10.63	11.15
1995	8.40	9.14	9.34	8.18	9.81	10.09	8.01	10.59	10.93
1996	8.46	9.08	9.25	8.17	9.71	9.92	8.03	10.47	10.70
1997	8.47	9.19	9.41	8.18	9.93	10.20	8.02	10.88	11.17
1998	8.47	10.17	10.42	8.34	10.24	11.00	8.33	9.86	11.85
1999	8.60	10.30	10.62	8.46	10.43	11.21	8.46	10.67	12.12
2000	8.65	10.49	10.77	8.54	11.09	11.43	8.54	10.88	12.43
2001	8.94	10.73	11.06	8.71	11.32	11.69	8.70	11.03	12.68
2002	9.10	10.89	11.25	8.88	11.64	12.01	8.87	11.34	13.25

Source: CEPR ORG extract.

**TABLE 6: Effect of different overtime, tips, and commission procedures on mean real wage**  
(constant 2002 dollars)

Year	All						Overtime, tips, commissions premium (%)		
	w_no_no	w_nber	w_ln_no	w_ln_ot	w_p8_no	w_p8_ot	w_nber	w_ln_ot	w_p8_ot
1979	14.26	14.26	14.32	14.32	14.38	14.38	0.00	0.00	0.00
1980	13.95	13.95	14.02	14.02	14.08	14.08	0.00	0.00	0.00
1981	13.81	13.82	13.92	13.92	14.00	14.00	0.07	0.00	0.00
1982	13.86	13.86	14.01	14.01	14.14	14.14	0.00	0.00	0.00
1983	13.82	13.81	14.01	14.02	14.18	14.18	-0.07	0.07	0.00
1984	13.84	13.83	14.07	14.07	14.29	14.30	-0.07	0.00	0.07
1985	13.94	13.94	14.22	14.22	14.44	14.44	0.00	0.00	0.00
1986	14.19	14.18	14.53	14.54	14.79	14.80	-0.07	0.07	0.07
1987	14.19	14.17	14.59	14.59	15.00	15.00	-0.14	0.00	0.00
1988	14.18	14.16	14.67	14.68	15.20	15.20	-0.14	0.07	0.00
1989	14.40	14.63	14.46	14.67	14.51	14.71	1.60	1.45	1.38
1990	14.40	14.61	14.47	14.66	14.53	14.72	1.46	1.31	1.31
1991	14.42	14.64	14.51	14.71	14.58	14.78	1.53	1.38	1.37
1992	14.43	14.66	14.51	14.72	14.58	14.78	1.59	1.45	1.37
1993	14.50	14.73	14.60	14.81	14.68	14.89	1.59	1.44	1.43
1994	14.51	14.90	14.66	15.33	14.78	15.45	2.69	4.57	4.53
1995	14.47	14.92	14.63	15.28	14.77	15.42	3.11	4.44	4.40
1996	14.47	14.92	14.63	15.26	14.78	15.41	3.11	4.31	4.26
1997	14.68	15.13	14.88	15.51	15.07	15.70	3.07	4.23	4.18
1998	15.33	15.78	15.32	15.95	15.18	15.81	2.94	4.11	4.15
1999	15.66	16.14	15.66	16.32	15.67	16.32	3.07	4.21	4.15
2000	15.85	16.33	15.98	16.64	15.86	16.52	3.03	4.13	4.16
2001	16.14	16.64	16.28	16.95	16.13	16.81	3.10	4.12	4.22
2002	16.38	16.85	16.55	17.19	16.37	17.02	2.87	3.87	3.97

Source: CEPR ORG extract. All distributions limited to range \$1-100 per hour.

**TABLE 6: Effect of different overtime, tips, and commission procedures on mean real wage (continued)**  
(constant 2002 dollars)

Year	Women						Men					
	w_no no	w_nber	w_ln no	w_ln ot	w_p8 no	w_p8 ot	w_no no	w_nber	w_ln no	w_ln ot	w_p8 no	w_p8 ot
1979	11.01	11.02	11.02	11.02	11.02	11.03	16.74	16.75	16.85	16.85	16.95	16.96
1980	10.89	10.89	10.90	10.90	10.90	10.91	16.38	16.38	16.50	16.51	16.61	16.61
1981	10.89	10.89	10.89	10.90	10.90	10.91	16.18	16.19	16.36	16.36	16.51	16.51
1982	11.07	11.07	11.08	11.09	11.10	11.10	16.19	16.19	16.44	16.44	16.67	16.67
1983	11.15	11.14	11.17	11.17	11.19	11.20	16.06	16.06	16.40	16.40	16.69	16.69
1984	11.26	11.26	11.30	11.30	11.34	11.34	16.00	15.99	16.39	16.39	16.77	16.77
1985	11.43	11.43	11.49	11.50	11.54	11.54	16.07	16.06	16.53	16.53	16.90	16.90
1986	11.73	11.73	11.80	11.81	11.86	11.86	16.31	16.29	16.89	16.89	17.32	17.32
1987	11.87	11.87	11.97	11.97	12.06	12.06	16.21	16.18	16.89	16.89	17.57	17.57
1988	11.98	11.97	12.11	12.11	12.24	12.24	16.11	16.07	16.93	16.93	17.80	17.80
1989	12.09	12.33	12.10	12.31	12.11	12.32	16.44	16.66	16.54	16.76	16.63	16.84
1990	12.19	12.41	12.20	12.40	12.21	12.41	16.36	16.58	16.48	16.68	16.59	16.79
1991	12.38	12.61	12.40	12.60	12.42	12.62	16.27	16.48	16.41	16.61	16.53	16.73
1992	12.52	12.76	12.55	12.76	12.56	12.77	16.16	16.39	16.31	16.52	16.42	16.63
1993	12.70	12.93	12.72	12.93	12.75	12.95	16.15	16.38	16.31	16.52	16.44	16.66
1994	12.79	13.09	12.84	13.37	12.88	13.41	16.07	16.55	16.30	17.11	16.49	17.30
1995	12.69	13.06	12.75	13.26	12.79	13.30	16.07	16.62	16.33	17.10	16.57	17.34
1996	12.78	13.13	12.83	13.31	12.87	13.36	16.01	16.57	16.27	17.04	16.52	17.28
1997	12.98	13.35	13.05	13.53	13.12	13.60	16.24	16.76	16.55	17.32	16.85	17.62
1998	13.44	13.82	13.42	13.90	13.37	13.85	17.04	17.58	17.05	17.81	16.84	17.60
1999	13.70	14.11	13.67	14.19	13.66	14.18	17.47	18.03	17.50	18.27	17.53	18.30
2000	13.89	14.32	13.93	14.45	13.86	14.38	17.66	18.20	17.87	18.66	17.70	18.49
2001	14.21	14.61	14.25	14.78	14.17	14.70	17.92	18.53	18.15	18.96	17.94	18.76
2002	14.51	14.91	14.56	15.09	14.47	14.99	18.11	18.65	18.38	19.14	18.14	18.90

Source: CEPR ORG extract. All distributions limited to range \$1-100 per hour.

**TABLE 7: Effect of different overtime, tips, and commision procedures on the mean of the real wage for hourly workers**  
(constant 2002 dollars)

Year	Retail			Manufacturing			Education		
	w ln no	w ln ot	Premium (%)	w ln no	w ln ot	Premium (%)	w ln no	w ln ot	Premium (%)
1979	8.96	8.97	0.1	13.36	13.36	0.0	9.53	9.54	0.1
1980	8.77	8.78	0.1	13.18	13.18	0.0	9.44	9.45	0.1
1981	8.54	8.55	0.1	13.08	13.08	0.0	9.50	9.51	0.1
1982	8.34	8.34	0.0	13.08	13.08	0.0	9.40	9.41	0.1
1983	8.17	8.18	0.1	12.97	12.97	0.0	9.62	9.64	0.2
1984	8.02	8.02	0.0	12.83	12.83	0.0	9.87	9.88	0.1
1985	7.97	7.97	0.0	12.93	12.93	0.0	10.22	10.23	0.1
1986	8.02	8.02	0.0	13.05	13.05	0.0	10.11	10.12	0.1
1987	8.02	8.03	0.1	12.86	12.86	0.0	10.32	10.33	0.1
1988	7.97	7.97	0.0	12.78	12.79	0.1	10.52	10.53	0.1
1989	7.94	8.40	5.8	12.66	12.98	2.5	10.67	10.89	2.1
1990	7.96	8.41	5.7	12.48	12.76	2.2	10.49	10.73	2.3
1991	7.96	8.41	5.7	12.50	12.78	2.2	10.80	11.01	1.9
1992	8.02	8.46	5.5	12.50	12.79	2.3	10.54	10.80	2.5
1993	8.04	8.47	5.3	12.48	12.78	2.4	10.87	11.09	2.0
1994	8.35	9.37	12.2	12.67	14.01	10.6	11.54	12.22	5.9
1995	8.13	9.10	11.9	12.59	13.88	10.2	11.46	12.10	5.6
1996	8.18	9.13	11.6	12.54	13.78	9.9	11.23	11.86	5.6
1997	8.25	9.14	10.8	12.70	13.98	10.1	11.36	11.97	5.4
1998	8.66	9.58	10.6	13.04	14.29	9.6	11.90	12.50	5.0
1999	8.83	9.83	11.3	13.20	14.47	9.6	12.02	12.72	5.8
2000	8.83	9.89	12.0	13.22	14.59	10.4	12.11	12.81	5.8
2001	9.07	10.11	11.5	13.44	14.72	9.5	12.58	13.28	5.6
2002	9.14	10.19	11.5	13.63	14.89	9.2	12.64	13.36	5.7

Source: CEPR ORG extract. All distributions limited to range \$1-100 per hour.

**TABLE 8: Share of workers reporting that usual "hours vary"**  
(Percent)

Year	All	Women	Men
1979	--	--	--
1980	--	--	--
1981	--	--	--
1982	--	--	--
1983	--	--	--
1984	--	--	--
1985	--	--	--
1986	--	--	--
1987	--	--	--
1988	--	--	--
1989	--	--	--
1990	--	--	--
1991	--	--	--
1992	--	--	--
1993	--	--	--
1994	5.7	5.3	6.1
1995	6.1	5.7	6.4
1996	6.9	6.2	7.6
1997	6.8	6.4	7.2
1998	6.9	6.5	7.3
1999	6.6	6.3	6.9
2000	6.5	6.1	6.9
2001	6.7	6.3	7.1
2002	6.6	6.3	7.0

Source: CEPR ORG extract. Limited to wages in the range \$1-100 per hour. Reporting "hours vary" not a survey option 1979-93. Results are for w\_ln\_no wage variable, but are similar for all other wage variables.

**TABLE 9: Effect of "hours vary" adjustment on the distribution of real wage for all workers**  
(constant 2002 dollars)

Year	Mean			10th percentile			50th percentile			90th percentile			Standard deviation		
	Adj.	Excl.	Diff (%)	Adj.	Excl.	Diff (%)	Adj.	Excl.	Diff (%)	Adj.	Excl.	Diff (%)	Adj.	Excl.	Diff (%)
1979	14.32	14.32	0.00	6.89	6.89	0.00	12.12	12.12	0.00	23.64	23.64	0.00	8.28	8.28	0.00
1980	14.02	14.02	0.00	6.51	6.51	0.00	12.09	12.09	0.00	23.38	23.38	0.00	8.04	8.04	0.00
1981	13.92	13.92	0.00	6.42	6.42	0.00	11.81	11.81	0.00	23.62	23.62	0.00	8.05	8.05	0.00
1982	14.01	14.01	0.00	6.24	6.24	0.00	11.89	11.89	0.00	23.79	23.79	0.00	8.34	8.34	0.00
1983	14.01	14.01	0.00	6.00	6.00	0.00	11.99	11.99	0.00	23.98	23.98	0.00	8.60	8.60	0.00
1984	14.07	14.07	0.00	5.93	5.93	0.00	11.92	11.92	0.00	24.71	24.71	0.00	8.64	8.64	0.00
1985	14.22	14.22	0.00	5.93	5.93	0.00	11.96	11.96	0.00	24.26	24.26	0.00	8.80	8.80	0.00
1986	14.53	14.53	0.00	5.87	5.87	0.00	12.36	12.36	0.00	25.05	25.05	0.00	9.09	9.09	0.00
1987	14.59	14.59	0.00	6.06	6.06	0.00	12.11	12.11	0.00	25.61	25.61	0.00	9.16	9.16	0.00
1988	14.67	14.67	0.00	5.84	5.84	0.00	12.21	12.21	0.00	25.56	25.56	0.00	9.33	9.33	0.00
1989	14.46	14.46	0.00	5.95	5.95	0.00	12.25	12.25	0.00	25.21	25.21	0.00	9.22	9.22	0.00
1990	14.47	14.47	0.00	6.00	6.00	0.00	12.01	12.01	0.00	25.65	25.65	0.00	9.22	9.22	0.00
1991	14.51	14.51	0.00	6.12	6.12	0.00	12.23	12.23	0.00	25.75	25.75	0.00	9.29	9.29	0.00
1992	14.51	14.51	0.00	6.28	6.28	0.00	12.31	12.31	0.00	25.12	25.12	0.00	9.30	9.30	0.00
1993	14.60	14.60	0.00	6.13	6.13	0.00	12.26	12.26	0.00	25.74	25.74	0.00	9.36	9.36	0.00
1994	14.66	14.75	0.61	6.00	6.00	0.00	12.00	12.00	0.00	26.09	26.19	0.38	10.00	10.14	1.40
1995	14.63	14.74	0.75	5.86	5.86	0.00	11.72	11.86	1.19	26.07	26.28	0.81	9.81	9.95	1.43
1996	14.63	14.73	0.68	5.94	6.00	1.01	11.85	11.98	1.10	26.32	26.32	0.00	9.71	9.80	0.93
1997	14.88	14.98	0.67	6.15	6.26	1.79	12.23	12.29	0.49	26.82	26.82	0.00	9.93	10.05	1.21
1998	15.32	15.38	0.39	6.61	6.61	0.00	12.45	12.67	1.77	27.55	27.55	0.00	10.24	10.27	0.29
1999	15.66	15.74	0.51	6.47	6.53	0.93	12.95	12.95	0.00	27.84	27.88	0.14	10.43	10.47	0.38
2000	15.98	16.05	0.44	6.68	6.79	1.65	12.79	13.05	2.03	28.39	28.43	0.14	11.09	11.15	0.54
2001	16.28	16.35	0.43	6.86	7.11	3.64	13.21	13.21	0.00	29.28	29.28	0.00	11.32	11.40	0.71
2002	16.55	16.62	0.42	7.00	7.00	0.00	13.14	13.40	1.98	29.91	29.91	0.00	11.64	11.69	0.43

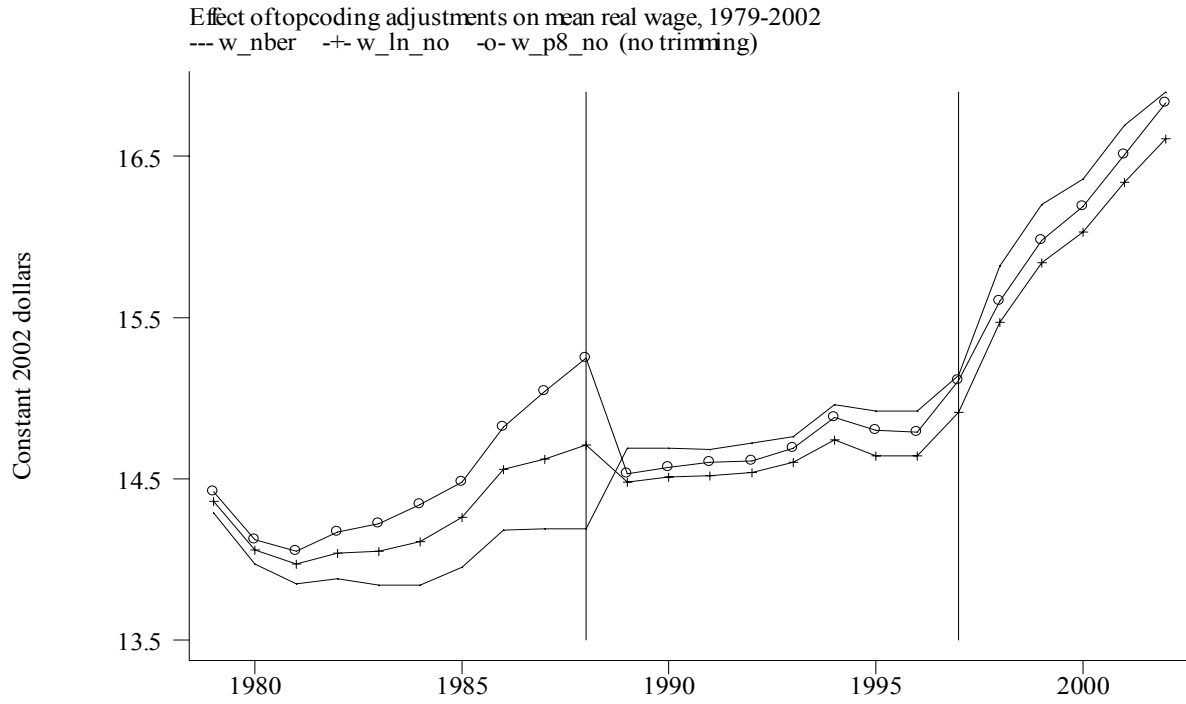
Source: CEPR ORG extract. All distributions limited to range \$1-100 per hour. "Adj." indicates that data include adjustments for observations from 1994) on where "hours vary." "Excl." indicates that data exclude all observations (from 1994) where "hours vary." Results are for w\_ln\_no wage variable, but are similar for all other wage variables.

**TABLE 10: Share of workers with wage allocated by BLS**  
(Percent)

Year	All	Women	Men
1979	16.5	14.7	17.9
1980	15.9	14.2	17.3
1981	15.3	13.8	16.6
1982	13.8	12.4	15.0
1983	13.9	12.6	15.1
1984	14.8	13.3	16.0
1985	14.1	12.8	15.2
1986	10.4	9.5	11.1
1987	13.4	12.5	14.2
1988	14.5	13.5	15.3
1989	2.1	2.1	2.1
1990	2.3	2.3	2.2
1991	2.5	2.4	2.6
1992	2.4	2.5	2.3
1993	2.6	2.6	2.6
1994	0.0	0.0	0.0
1995	8.3	7.9	8.7
1996	24.3	23.3	25.2
1997	24.6	23.4	25.7
1998	25.8	24.8	26.8
1999	29.6	28.4	30.7
2000	32.0	30.6	33.3
2001	33.7	32.6	34.7
2002	33.7	32.7	34.7

Source: CEPR ORG extract. Limited to wages in the range \$1-100 per hour. Results are for w\_ln\_no wage variable, but are similar for all other wage variables.

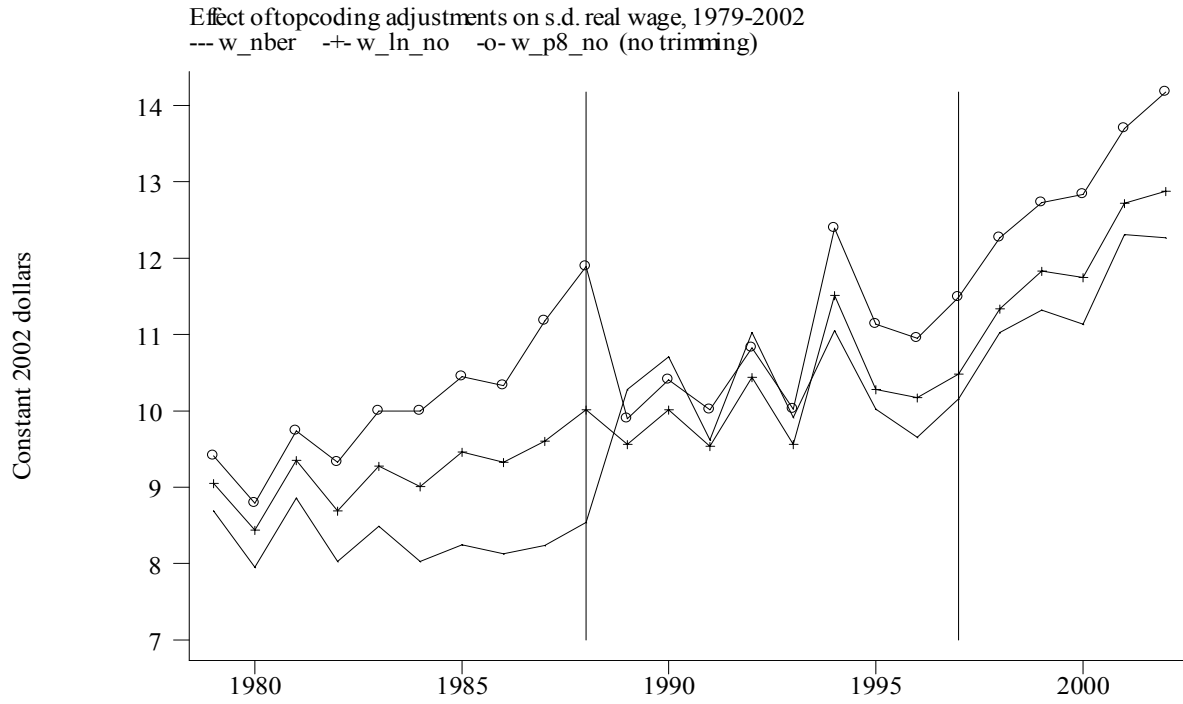
**Figure 1**



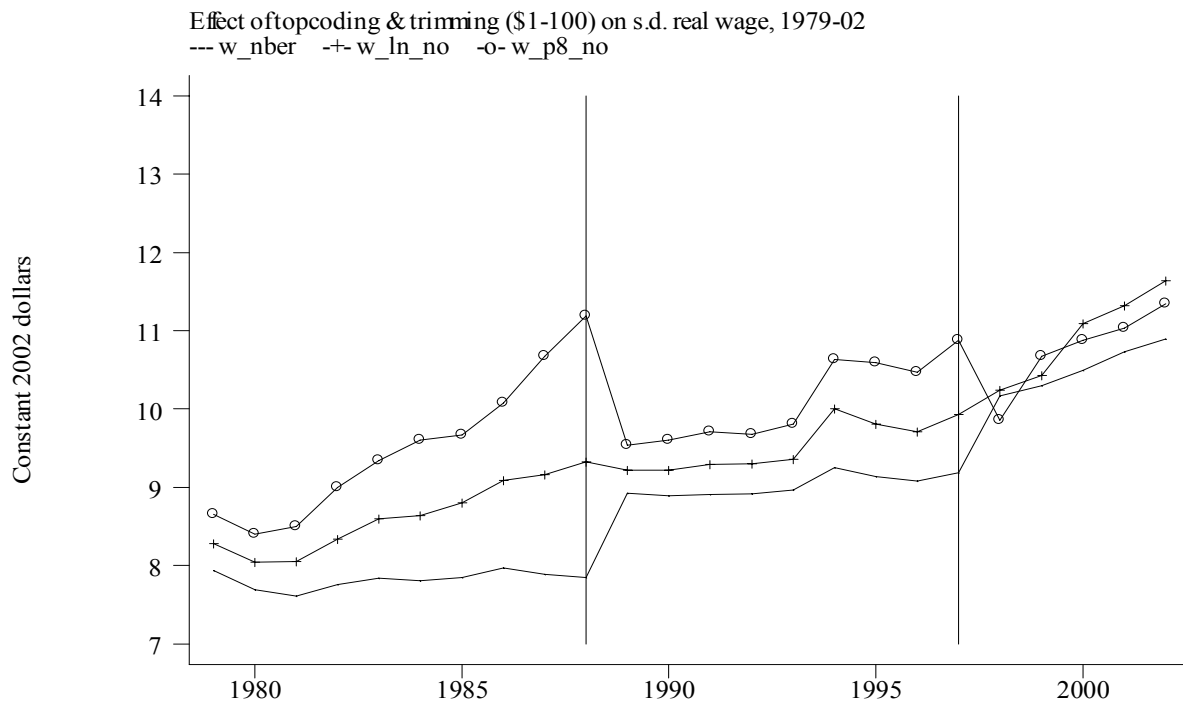
**Figure 2**



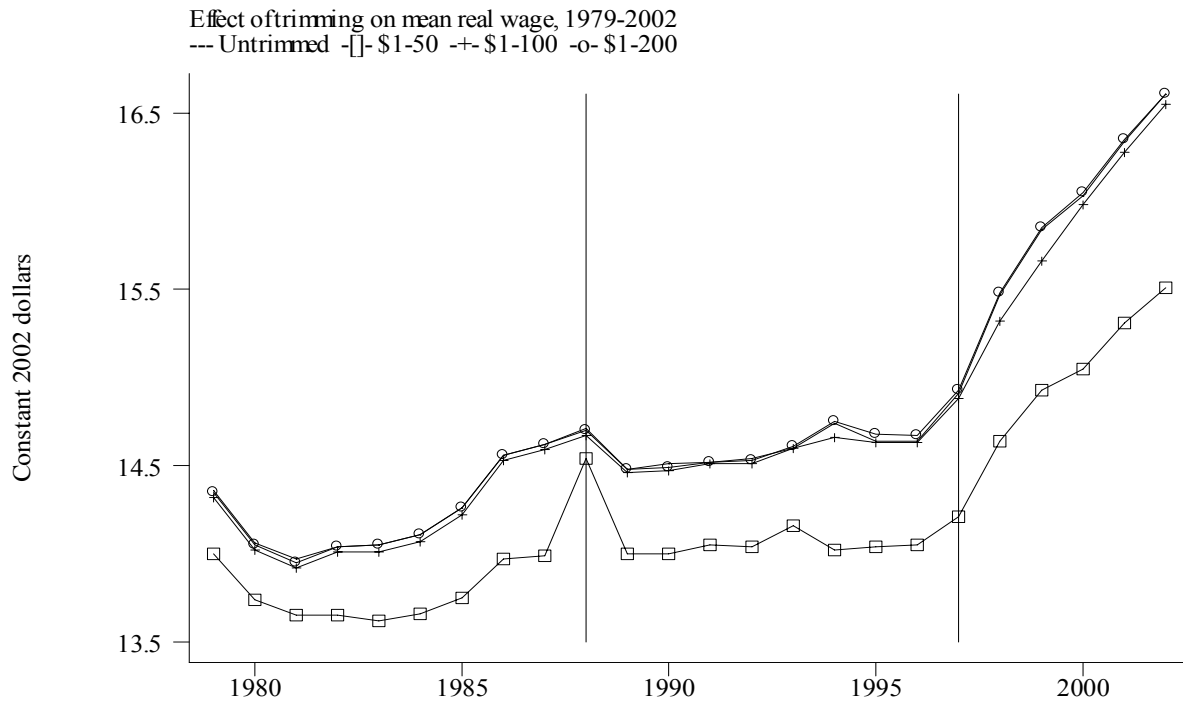
**Figure 3**



**Figure 4**



**Figure 5**



**Figure 6**

