**6/12/2007**

# Methodologies for Estimating Mean Wages for Occupational Employment Statistics (OES) Data

Mallika Kasturirangan, Shail Butani, Tamara Sue Zimmerman

Bureau of Labor Statistics

Office of Employment and Unemployment Statistics

Statistical Methods Staff

April 2007

**ABSTRACT**

The Occupational Employment Statistics (OES) is a joint Federal/State partnership program with a sample size of 1.2 million establishments over a 3-year period (six semi-annual panels each consisting of 200,000 establishments). The OES collects occupational employment and wage data for approximately 800 occupations at the MSA by 4-5 digit industrial (NAICS) level. Because of the burden on respondents, this survey is designed to collect wage data in intervals rather than exact wages for individual employees. In this talk, we will present the previous research work on the construction of lower and upper bounds of the intervals; alternative methods for estimating mean wages—arithmetic, geometric, and NCS mean wages; updating of wages from prior panels; and calculation of mean wages for the upper open-ended Interval L (i.e., employees making $70 or more per hour in the years 2003-2005). This study further examines several methods for approximating mean wages for Interval L for occupations that have significant employment (>5%) in Interval L and validates the OES methodology on independent data sets from the Current Population Survey for years 2003, 2004, and 2005.

**KEY WORDS:** Open-ended interval wage estimation, geometric mean, kernel density

### I. The Occupational Employment Statistics Survey

The Occupational Employment Statistics (OES) program is a Federal/State partnership program between the Bureau of Labor Statistics (BLS) and State Workforce Agencies (SWAs). OES collects data from approximately 1.2 million establishments over a three-year period. The data is collected in six semi-annual panels, with about 200,000 establishments surveyed per panel. Every effort is made to survey an establishment only once in each three-year period.

The OES survey is designed to cover all full- and part-time salary workers in non-farm industries; it does not cover the self-employed, owners and partners in unincorporated firms, or household workers. OES is designed to produce over 800 occupational employment and wage estimates. Occupations are classified using the Office of Management and Budget (OMB) Standard Occupational Classification (SOC) system. OES is designed to produce these estimates by geographic area—National, State, and Metropolitan Area levels—and for over 450 industry classifications corresponding to the 3-, 4- and 5-digit North American Classification System (NAICS).

OES data are available for a wide variety of uses. One of the OES's biggest clients is the Foreign Labor Certification program by the Employment Training Administration (ETA). OES data are also used in analysis of occupational employment and wages, development of occupational projections, vocational counseling and planning, industry skills and technology studies, and market analysis.

**II. OES Wage Intervals**

OES requests a large amount of data from respondents. In order to ease the burden on responding establishments, OES does not collect data on exact wages. Instead, establishments report the number of workers in a certain occupation earning within each of twelve wage intervals, denoted "A" through "L". Figure 1 (all figures are at the end of this document) shows an example of the form given to respondents of the OES Survey. Each row represents a certain occupation: chief executives, general and operations managers, etc. Each column represents a wage interval. Interval A represents those earning less than $6.75 an hour, and so on. All a respondent has to do is report the number of chief executives earning wages within Interval K ($55.50-$69.99), for example.

Extensive research was conducted as to how the upper and lower bounds of these wage intervals should be constructed. Several factors were considered prior to interval construction. Research was done by Dr. Sandra West on reported weekly wages from the Current Population Survey. [1] Figure 2 shows some results of her research. Reported weekly earnings spiked at clean values like $200 ($5.00 per hour), $300, $400, etc. West concluded that this data does partially represent the true distribution of wages, but it also represents reporting error. Interval methodology could eliminate some of this error.

Earlier simulations showed that mean squared errors were less when the lower bound of an interval was a factor of $0.25, rather than $0.25 plus a penny. The general method for determining the bounds of the intervals was to attempt to equalize the percent relative errors and coefficients of variation within each interval. This method is applied to

---

[1] West, Sandra A. "Standard Measures of Central Tendency for Censored Earnings Data from the Current Population Survey"

Intervals B-K. Construction of Interval A is a function of the federal minimum wage, and Interval L is a function of inflation.[2] Figure 3 shows the OES interval bounds for 2003-2005, along with the percent relative errors and coefficients of variation. Notice the percent relative errors are similar for Intervals B-K (11.5-11.8%) as are the coefficients of variation (6.5-6.8%). Even Interval A is fairly close to these numbers. Interval L cannot be compared because it is an open-ended interval.

There have been several proposed methods for estimating mean wages within each of these intervals, including the arithmetic mean (midpoint), geometric mean, and National Compensation Survey (NCS) mean.

**Arithmetic**: (Lower Bound + Upper Bound) / 2

**Geometric**: (Lower Bound x Upper Bound) $^{1/2}$

The NCS mean is calculated using point data from the National Compensation Survey, weighted by number of workers earning that wage. The arithmetic mean was the method used previously. The geometric mean was never used, but suggested by a participating state office. The NCS mean is the method currently used to estimate mean wages within intervals. Research showed that the arithmetic mean worked well, the geometric was better, and the NCS mean performed best of the three.[3] Figure 4 shows a comparison of these three methods as well as the weighted mean using Current Population Survey (CPS) point data for 2004. For Intervals B-K, the interval means are very similar across all

---

[2] Tou, Albert, Christina Chiu, and Kenneth Robertson. "An Evaluation of the Occupational Staffing Pattern and Mean Wage Rates of State Government Workers"

[3] "Occupational Employment Statistics Survey: Analysis of Alternative Procedures for IntervalMean Wage Rates, Supplement II". December 2000

methods, which is why all three methods performed well in testing. The NCS Mean is generally lower than the arithmetic mean (the midpoint) suggesting that the data are not uniformly distributed within each interval, but tend to be slightly more concentrated in the lower half of the interval.

It is also important to examine the data across years. Figure 5 shows interval means using NCS and CPS data across the years 2003-2005. The interval means do not seem to change across years, yet the overall means do increase about 2-3% each year. Figure 6 gives us the reason for that 2-3% increase in overall mean wages. In all three surveys, NCS, CPS, and OES, we see a shift in percent of employment from the lower Intervals A-D to the upper Intervals E-L. It is also important to notice in Figure 5, there is a $60 difference between the mean wages for Interval L in CPS and NCS, with CPS being higher. If we consider that Interval L makes up a little over 1% of employment in CPS (Figure 6), this would create about a $0.60 difference in the overall means between NCS and CPS, which we can also see in Figure 5.

### III. OES Mean Wage Estimation

OES does not publish estimates for interval means. These are simply used in calculations of overall mean wages for occupations. Currently, to determine occupational mean wages, exact data from NCS are used to determine a wage for each interval. This interval mean is then applied to the OES sample. For Intervals A and B, if the State-specific minimum wage is higher than the calculated NCS mean for that interval, then the State minimum wage is applied. Otherwise, the NCS mean will be applied as with the other intervals. Interval L, the open-ended interval ($70 per hour or more), has a special

procedure. Because this interval makes up only about 1% of the population, data in Interval L has little or no impact on mean wages for most occupations. The interval means are then used to compute an overall mean wage for each occupation. To take into account the 3-year period over which data are collected, previous panels' data are updated using the employment cost index (ECI) by nine major occupational groups. The results of a study conducted by Kirk Wolter and Rashna Ghadialy of National Opinion Research Corporation (NORC) found this procedure performed the best of all the alternative methods used for updating prior year wage data.[4]

In order to ascertain the appropriateness of using NCS data to compute mean wages for each interval for OES, a Wage Comparison Study was done in 2003.[5] When comparing occupational wages between the two surveys it was found that about 70% of the detailed occupations did not have statistically significant differences at the 10% significance level. About one-third of the other 30% of occupations did not have economically meaningful differences between mean wages in each survey. This left about 20% of occupations that had both statistically significant differences and economically meaningful differences. These differences could arise from a number of factors including: small sample size in either survey; conceptual differences regarding definitions of occupations; occupational coding differences; and differences in methods of measurement of mean wages in open-ended Interval L for high paying occupations (the focus of this research), etc. When OES interval methodologies were applied to the NCS

---

[4]"Occupational Employment Statistics Survey: Analysis of Alternative Procedures for IntervalMean Wage Rates, Supplement II". December 2000
[5] Barkume, Tony. Matt Dey, Larry Ernst, Maury Gittleman, Anne Polivka. "Comparing OES and NCS Wage Estimates. Phase I: Comparison of National Estimates." May 2006.

data, the above results did not change, suggesting it is appropriate to use NCS point data to estimate mean wages for OES.

## IV. Purpose of Research

It is important to note that while OES uses interval mean methodology, the program is not designed to make occupational estimates for any specific interval. This study is not exploring differences in wage rates by area occupation. We assume a priori that wages do, in fact, differ by area and occupation. The assumption that occupational mean wage rates vary by these factors does not imply, however, that wage rates vary within a wage interval. Virtually all occupations are reported across a subset of the complete set of wage intervals, low paying occupations concentrated in the lower intervals, and high paying occupations in the higher intervals. Because of this fact, this study will concentrate only on mean wages for occupations overall. The following research has two goals: 1) Evaluate the effectiveness of the OES estimation procedures used in occupational mean wage rates, across all intervals, on an independent data set. For this study we have used data from the Current Population Survey (CPS) for 2003-2005 as our independent sample. 2) Develop alternative procedures for estimating mean wages for Interval L, specifically for those occupations having significant employment in this open-ended wage interval.

## V. Validation of OES Methodology

To check the accuracy and effectiveness of the OES methods, data from CPS was tabulated and used to compute mean wages using the OES methodology as described

previously. CPS and OES use the same occupational coding system, which allows for some comparison. However, the CPS only offers data at the major occupational group level. In each year, about one-third of the 22 major occupational groups had statistically significant differences between the actual mean computed using CPS reported data, and the mean computed using OES methodology at the 10% significance level. Also consistently from 2003 to 2005 only one major occupational group or 5% of the groups had an economically meaningful difference of more than 10%.

Figure 7 displays the results of this validation procedure for 2005. The table shows first the mean wages computed using CPS data weighted by employment, followed by the unweighted sample size for a point of reference as to relative sample size of each occupational group. For each occupational group, a 90% confidence interval was computed using standard errors produced from CPS replicate weights. Notice that the confidence intervals around the CPS computed mean are very tight. This suggests that there is very small sampling error in CPS. Next on this table is the mean wage using OES methodology. That is, CPS data were dropped into the OES wage intervals. The wage values were then replaced with mean wages for each interval computed from NCS data. State-specific minimum wages were applied when applicable for Intervals A and B. The overall mean wage for each occupation group was then calculated from the CPS sample. Means using OES methodology tended to be lower than those using CPS reported data. Some of this difference is attributable to the difference in mean wages between NCS and CPS within Interval L. As mentioned before, CPS wages in Interval L were about $60 higher than those in NCS for 2003-2005. However, only one group, as mentioned before, had an economically meaningful difference in this study. For 2005, as shown in Figure 7,

this group is arts, design, entertainment, sports and media occupations. Results for the years 2003 and 2004 were generally similar, leading to the conclusion that the OES methodology is fairly accurate, with a portion of the bias being attributable to Interval L.

## VI. Wage Interval L

OES Interval L accounts for those earning at least $70 per hour. A relatively small portion of the OES sample falls within this interval. For the years 2003-2005 less than 1.5% of the OES sample was in Interval L. What complicates OES mean wage estimation the most is the open-ended nature of this interval. An observation in Interval L could make $72 per hour or $5000 per hour, but this cannot be discerned from the collected data. This leads to hard-to-answer questions. Is it possible that OES is currently underestimating or overestimating the mean wage within Interval L? About 40 occupations have significant employment (over 5%) in the Interval L range. This is just a handful of the 800 detailed occupations dealt with in OES, but it is for these occupations that Interval L estimation has a meaningful effect. It is therefore necessary, especially for these specific occupations, to create a special methodology for accurately estimating mean wages in Interval L.

Figures 8, 9, and 10 show occupations that have employment of 5% or more in Interval L for 2003-2005. Notice that most of these occupations consist of physicians, managers, sales occupations and postsecondary teachers. These are the groups of high-paying occupations that are most effected by Interval L. Also of note here is the difference between percent employment in Interval L and employment level in Interval L. It is important to be aware of both numbers as each has its own merit. For example, in

Figure 9, podiatrists have 27% of their employment in Interval L, compared to 25% for lawyers. However, actual employment in L is only 2,200 for podiatrists, compared to 134,000 for lawyers. Which of the two measures is more important depends on the question we are attempting to answer.

## Computing Mean Wages for Interval L for Detailed Occupations

For this study it was necessary to examine several different data sources and assess the appropriateness of each to this topic. Mean and median wages from the NCS were examined for OES occupations with significant employment in Interval L. It was found that NCS data tended to be quite variable, and that the percent employment within Interval L by occupation was often quite different between the NCS and OES surveys. For example, in OES about 45% of pilots earn wages in Interval L, while in NCS about 70% of pilots fall in Interval L. Also overall pilots comprise about 2% of the total employment in Interval L for OES, compared to about 10% in NCS.

CPS data were also examined, but the lack of detailed occupational coding and large difference in mean wages for Interval L between CPS and NCS were flagged as possible issues. Data from the American Community Survey were also considered, but issues with quality of data and mean wages similar to CPS eliminated this option.

There is a methodology already in place for computing mean wages in Interval L. Currently, NCS data are used to compute a mean wage across all occupations excluding pilots for each of the years or panels in the sample. This mean wage across occupations excludes pilots because they have very high hourly wages (NCS data shows $145 per hour) and work a relatively low number of hours per year (1100 per year compared to the

average 2080).  Then the simple average of the yearly or panels estimates is computed.

The mean wage is not updated using ECI because it is difficult to justify updating wages

in this group when the Interval L mean wages for both all occupations and that excluding

pilots are trending downwards or possibly fluctuating around a true mean.


**<u>Pilots</u>**

Figure 11 shows the Interval L mean wages for 1999-2002. For each of the four

years, the mean wage excluding pilots is lower than the mean for all occupations. Taking

a simple average of the four years' data shows still a $10 difference between the two

means. pilots' data tends to bias the Interval L data upwards.

Figure 12 examines pilots mean wages more closely. NCS publishes a mean

hourly wage for pilots as shown in this table, close to $100 for each year (notice the

variability across years). OES used a fairly conservative hourly estimate of about $95 per

hour according to the above methodology for the 2003-2005 years, yet the OES published

annual wage for pilots is still higher than that calculated using NCS wages and mean

annual hours for each year. Some of this difference could be explained by the fact that, in

OES about 45% of pilots earn wages in Interval L, while in NCS about 70% of pilots fall

in Interval L.  However, most of this difference is arising from the number of hours

worked by pilots.  In NCS, the number of hours worked in a year by pilots is

approximately 1100, whereas, in OES the assumption is 2080 hours.

## Options for Estimating Mean Wages in Interval L

Along with the current methodology, four other options have been proposed for computing mean wages for Interval L.

1. Compute mean wages across all occupations

2. Compute mean wages for NCS major occupational groups (see Figure 13)

3. Compute mean wages for two major groupings: sales and non-sales occupations

4. Compute mean wages for a combination of specific occupations and major occupational groups

Options 1 and 2 are fairly straightforward. When looking at results for Option 2 in Figure 14, it is apparent that, within Interval L, Occupational Groups D-K are negligible. These are groups that do not tend to have high employment in Interval L. It is also apparent that Occupational Groups A and B are very similar in both size and mean wages in Interval L. The most noticeable of these occupational groups is Group C, sales occupations. It is then reasonable to collapse Groups A, B, and D-K into a major grouping of non-sales and compare to the sales occupations, resulting in Option 3.

Option 4 can also be considered in this light. The occupational groups with the most employment in Interval L are Groups A, B, and C. These are professional and technical occupations, executive, administrative, and managerial occupations, and sales occupations. In other words, these are comparable to the handful of detailed occupations with significant employment in Interval L shown in Figures 8, 9, and 10. Ideally, these

detailed occupations could be handled individually, to compute the most accurate overall mean wage for each occupation. Option 4, consequently, would compute a mean wage in Interval L for specific occupations—physicians, post-secondary teachers, lawyers & judges, and pilots; and major groups—managers, sales occupations, and all others. The specific occupations chosen are all in Occupational Group A, while managers is Group B, sales Group C, and the rest of Group A and D-K are grouped in "all others."

The current option, as well as the three new options, were tested. (Note: option 2 is equivalent to option 3) Figure 15 details those occupations that failed tests for each of the options. For an occupation to fail when tested with each option, it had to have at least 1000 employed (weighted) in Interval L and the given option had to produce a relative error on mean wages per hour of 10% or more for at least 2 of the 3 years of 2003-2005. All options worked fairly well. In total, 25 occupations had at least 1000 employed in Interval L. Out of this only three occupations (or about 12%) failed for the current option and Option 4, while in Option 1 five occupations (or about 20%) failed.

Any option chosen would still need to be tested on OES data. For example, let us choose Option 4. In testing on NCS data, Option 4 forces pilots to pass because a separate mean wage for Interval L is computed for pilots using NCS data. If the interval hourly mean wage for pilots from NCS ($145 per hour) is applied to the OES sample, it would further increase the OES published as $95 per hour for Interval L and 2080 hours are used for annual wages. However, when we adjust NCS mean wages for Interval L to take into account the NCS annual hours worked and OES annual hours worked assumption (i.e. Interval L mean wages in OES context are equal to about $77 or $145 \times 1100 \div 2080$), then this bring the OES annual wages close to the NCS calculated annual wages, as

shown in Figure 16. This suggests that even for specific occupations, special methods may need to be applied.

## VII. Future Research

More research on this topic is definitely required. As mentioned before, each proposed option for estimating mean wages in Interval L should be tested further. This should include testing the options on NCS data after it has changed to using the Standard Occupational Coding system currently used by OES. This should also include testing the options on OES data with the new Interval L lower bound of $80 per hour. It would also be beneficial to test the use of special procedures for certain occupations, as in the case of pilots as well as post-secondary teachers, who often earn wages on an annual basis rather than hourly (as a result of unusual annual hours). In question as well is whether this procedure should utilize one year's data in NCS or the average of three years.

To even further examine the distribution of wages within intervals, it may be useful to explore a kernel density model.[6] A non-parametric kernel density function could be used to estimate mean wages for wage intervals by occupation. This approach is particularly suitable for wage data collected from the NCS because the point data are available. It can also be useful in estimating percentiles from OES data.

---

[6] Lehman.  <u>Elements of Large Sample Theory</u>, Springer 1999, pp. 406-417

# References

1. Butani, Shail. "Mean OES Wages--$60.00". October 17, 1990. Electronic message with attachment

2. Butani, Shail J., and Michael McElroy. "Managing Various Customer Needs for Occupational Employment Statistics Survey"

3. Butani, Shail, Kenneth Robertson, and George Werking. "Alternative Methods for Updating Wage Data from the Occupational Employment Statistics Survey"

4. Echols, Carrae, Albert Tou, and Kenneth W. Robertson. "Comparing Alternative Median Wage Rate Estimates for the Occupational Employment Statistics Survey"

5. Robertson, Kenneth W., Albert Tou, and Larry Huff. "A Study of Donor Pools and Imputation Methods for Missing Employment Data"

6. Robertson, Kenneth W., and Pamela L. Frugoli. "Statistical Issues for the Redesigned Occupational Employment Statistics Survey"

7. Tou, Albert, Christina Chiu, and Kenneth Robertson. "An Evaluation of the Occupational Staffing Pattern and Mean Wage Rates of State Government Workers".

8. West, Sandra A. "Estimation of the Mean from Censored Income Data".

9. West, Sandra A. "Standard Measures of Central Tendency for Censored Earnings Data from the Current Population Survey".

10. West, Sandra A., Shail Butani, and Michael Witt. "Alternative Imputation Methods for Wage Data"

11. West, Sandra, Diem-Tran Kratzke, and Shail Butani. "Measures of Central Tendency for Censored Wage Data"

12. West, Sandra, Diem-Tran Kratzke, and Kenneth Robertson. "Alternative Procedures for Item Non-response from New Establishments in the Universe"

13. Wolter, Kirk, and Rashna Ghadialy. "Preliminary Report on Evaluation of the Occupational Employment Statistics Estimation and Alternative Methods of Updating Wage Estimates"

14. "Occupational Employment Statistics Survey: Analysis of Alternative Procedures for IntervalMean Wage Rates". October 2000

15. "Occupational Employment Statistics Survey: Analysis of Alternative Procedures for IntervalMean Wage Rates, Supplement II". December 2000

16. Lehman.  <u>Elements of Large Sample Theory</u>, Springer 1999, pp. 406-417

All of the figures mentioned in this research paper can be found at

http://www.bls.gov/ore/pdf/st070010_figures.pdf