ALTERNATIVE METHODS FOR UPDATING WAGE DATA FROM THE OCCUPATIONAL EMPLOYMENT STATISTICS SURVEY

Shail Butani[1], Kenneth Robertson, and George Werking
U.S. Bureau of Labor Statistics, 2 Massachusetts Ave. N.E., Room 4985, Washington, D.C. 20212

Abstract

The Occupational Employment Statistics (OES) survey utilizes three years of data to produce estimates of employment and wage rates by occupation. In order to limit the bias inherent in using older wage data to estimate current period wage rates, we utilize a procedure to update previous year's wage data to the current period. The current procedure uses estimates from the U.S. Bureau of Labor Statistics' Employment Cost Index for the time periods of interest to develop the rate of change in wages at the occupational group level. In this paper we explore alternatives to the current procedure.

Key words: Wage Updating, Bias Reduction

1. Introduction

The Occupational Employment Statistics (OES) Survey is a Federal/State Cooperative Program for which the U.S. Bureau of Labor Statistics has the primary responsibility to develop, oversee, and implement the survey methodology procedures. The OES Survey is designed to measure employment, and mean and median wages per hour, for over 750 occupations. These statistics are developed at detailed levels of industry classification and geographic area . Industry groups are defined by 3-digit Standard Industrial Classification (SIC) codes, 2-digit SIC codes, and industry divisions; there are about 377 3-digit industries. Geographic regions are defined by; 334 metropolitan statistical areas (MSAs), and the residual areas within a state that are not part of an MSA (balance of state); the 50 states plus the District of Columbia, Guam, Puerto Rico, and the Virgin Islands; and the Nation as a whole.

In order to increase the precision of the estimates at very detailed levels, the OES sample design calls for combining sampled data from 1.2 million establishments covering about 70 percent of the employment in the United States across three consecutive years. While there are significant reductions in sampling errors by combining data across years, there are also limitations associated with this estimation procedure in that it requires "wage updating" for the earlier years of data. The Bureau is currently using the over-the-year wage changes in the national Employment Cost Index (ECI) to perform "wage updating".

The focus of this paper is to explore possible data sources and alternative methods that could be used to update prior years wage data to the current year. Five major factors that may have an effect on updating wages for OES are occupation, industry, geography, the size of the local economy, and the size of the employer. Possible data sources for updating wages are the ECI, the Consumer Price Index (CPI), the Covered Employment and Wages (ES-202) Program, the Current Employment Statistics (CES) Survey (commonly known as the payroll survey), and the OES survey; all of these data sources are Bureau programs. Additionally, we explored using OES data at some aggregate level to update the detailed occupational groups.

We start by briefly describing some of the uses and users of OES data. In Section 3, we give an overview of the survey design including data collection procedures. In Section 4, we give a description of the possible data sources; in Section 5, we state alternative methods that can be used to update the prior years wage data. In Section 6, we present the empirical results of our evaluation as measured by relative errors on mean wages and relative errors weighted by occupational employment. In Section 7, we conclude with some recommendations and future research projects.

[1] Any opinions expressed in this paper are those of the authors and do not constitute policy of the U.S. Bureau of the Labor Statistics.

2. Uses and Users of OES data

The two major products of the OES program are occupational staffing patterns and occupational wage rates. Estimates produced by the OES program are used extensively by a wide variety of customers that include Federal, State, and Local Governments, academic researchers, job placement specialists, career counselors, private businesses and employers, job seekers, economic development specialists, etc.

Occupational staffing patterns serve as an input to the National and State industry-occupation employment matrices and projections. The occupational projections are used primarily in planning vocational education and government sponsored training programs and for career guidance.

In recent times, there has been a great demand by employers and workers for occupational wage rates at the state and sub-state (Metropolitan Statistical Area, county, etc.) levels. Additionally, the Alien Labor Certification Program under the U.S. Immigration Act needs statistically sound, comparative, and timely prevailing wages by occupation for states and local areas. The Alien Labor Certification Program is the major sponsor of the redesigned OES Survey.

3. Survey Design

The sampling frame for the survey is derived primarily from the list of establishments reporting to state unemployment insurance (UI) programs. The OES Survey is based on a probability sample; it is stratified by MSA, 3-digit SIC code, and employment size of establishment; the larger establishments are sampled with virtual certainty across a three year cycle. In order to produce MSA-level estimates by 3-digit industry at the desired level of reliability, the OES Survey design requires a sample size of approximately 1.2 million establishments. The sample size was determined by equalizing the targeted relative standard error of the typical occupations across all MSA/SICs. The "balance of state" area, i.e. that part of the state not in an MSA, was over-sampled by 20 percent. This was done to accommodate the request made by the Alien Labor Certification Program to break these large areas into several parts for all-industry occupational wage estimates.

A sample of this magnitude places a large burden on the responding units. Distributing the sample across three consecutive years reduces that burden to a tolerable level. Overlap among the samples drawn during any 3-year cycle is kept to a minimum in order to limit the response burden on any one employer or alternatively to spread the burden more evenly.

The occupational employment and wage data are collected for the reference period of 4^{th} quarter in a matrix format usually via a mailed questionnaire; wage data are collected in 12 non-overlapping intervals (there were 11 intervals prior to 1999 Survey). The State Employment Security Agencies collect the data by conducting an initial and two follow-up mailings of the questionnaire, followed by telephone calls to non-respondents, and in a few cases by personal visit. The States are mostly responsible for data editing.

Estimation in OES begins with an imputation process for non-respondents; these imputation procedures are carried out independently for each survey year. First, the occupational staffing pattern is imputed using a hot-deck "nearest-neighbor" method. This method uses a responding establishment that most closely matches the non-responding establishment for key classification values (Area/SIC/Employment Size) as a donor. Second, missing wage data for the occupations are imputed using the distribution of the reported occupational wage data across wage intervals in the Area/SIC/Size Class. If there are sufficient data at this level, the procedure uses this reported wage distribution. If there are not enough data, the pool of donors is expanded to include adjacent size classes, industries, and areas until a distribution can be determined.

In OES, each single year sample represents one-third of the full 3-year sample plan. Each year's sample is weighted to represent the sample as it appeared at the time the sample was selected. In order to combine the data across years, each unit's weight is divided by the number of years that sample units were selected for that stratum. A ratio estimator is used to develop estimates of occupational employment. The auxiliary variable used is the reference period population value of total employment obtained from state unemployment insurance files. In order to balance the State's need for estimates at different levels of geographic and industrial details, the ratio adjustment process is

applied as a hierarchical series of ratio adjustment factors, or "benchmark" factors. While estimates can be made from a single year of data, the OES survey is designed to produce estimates using the full 3 years of data; this requires "wage updating" for the earlier years of data. For wage updating purposes, BLS is currently using national level over-the-year wage changes based on the fourth quarter (i.e., 4th quarter of year t-1 to 4th quarter of year t) for the nine occupational divisions for which ECI (Employment Cost Index) estimates are available. Such a procedure assumes that each occupation's wages, as measured in the earlier years, moves according to the average movement of its occupational division and that there are no major geographic, industrial, or detailed occupational differences for any given occupation — and this may not be the case.

The purpose of this research, therefore, is to explore possible data sources and alternative methods that would yield smaller errors for updating previous years wage data. Our evaluation of alternatives will be based on empirical results.

4. Possible Data Sources

At present, BLS is using the over-the-year change for nine major occupational groups based on fourth quarter ECI estimates. A major limitation of the ECI is its small sample size; this precludes applying the over-the-year change at sub-national levels and at detailed industry or occupational levels. If there are geographical, industrial, employment size class, or detailed occupational differences, then the potential for bias exists with the current procedure.

The Bureau's CPI is the most widely used statistic to gauge inflation. Operationally, applying a single figure that represents over-the-year change for all geographical, industrial, and occupational groups is easy. The CPI, however, measures the price of a fixed basket of goods. A procedure based on such a measure, therefore, could by its very nature have inherent biases for wage updating.

The primary data source for the Covered Employment and Wages (ES-202) Program is the state unemployment insurance records. Each State file consists of a list of establishments (virtual census of all employers) that have 4-digit SIC codes, County or equivalent code for mapping into a MSA, employment levels for three months of the quarter, and total quarterly wages. The ES-202 files are updated quarterly and are available for use by BLS survey methodologists about 9 months after the end of the quarter. Collectively the files for the 50 States plus the District of Columbia consist of about seven million records. This is a comprehensive database that can be used for measuring over-the-year change in wages at the MSA level for various levels of industry aggregation.

The ES-202 files are not, however, perfect for the purpose of developing rate of change information for the OES survey. First, the ES-202 program does not obtain information by occupation. Second, the definition of wages for the ES-202 Program is different than that used for the OES survey. ES-202 wages include bonuses, contributions to 401K plans in some states, and such items as meals, lodging, tips, and other gratuities, which are excluded by definition from the OES wages. A rate-of-change in wages developed by these files may be less useful for OES purposes because of these conceptual differences.

The Current Employment Statistics (CES) Survey is a monthly survey of about 400,000 establishments that collects information on number of all employees, production or nonsupervisory workers and their associated payroll and hours worked. The estimates are produced for the nation, states, and some MSAs by differing level of industry details. Since it is a monthly survey, it is the most timely data source. It does not, however, collect information by occupation. Moreover, the scope of payroll data is limited to production or nonsupervisory workers. These issues may limit the usefulness of this data source as a comprehensive solution for wage updating.

The final source of data we can consider is the OES survey itself. The survey is not designed to measure change; it is designed to measure levels. The OES survey, therefore, will not be used at fine levels of detail to develop rate of change information. It may be possible, however, to use the OES survey rate of change at moderate levels of detail to supplement the ECI-based procedure, and improve upon the current procedure.

5. Alternative Methods

Basic methods are employed in this analysis.

- We identify the data source whose rate of change most closely matches that of the unadjusted OES data.
- We examine the correlation in the rate of change in wages for these data sources and OES estimates.
- We review graphs of the rate of change.

6. Empirical Results

As indicated above, the first information we developed was the rate of change at an aggregate level from each of several data sources. The table below was constructed by calculating a population level wage rate for 1996, 1997, and 1998 from each of the data sources. The rate of change was then calculated for each period of interest (1996-1997, 1997-1998, and 1996-1998). In the OES program, approximately one-third of the data are from two years ago, one-third from last year, and one-third from the current period. This means that a current period estimate has one-third of the data updated from two years ago, one-third of the data updated from one-year ago, and one-third with no update. The "total" update can then be estimated as follows

$$Update = \frac{1}{3}\left(ROC_{98,98} + ROC_{97,98} + ROC_{96,98}\right) = \frac{1}{3}\left(0 + ROC_{97,98} + ROC_{96,98}\right)$$

Where: $ROC_{x,y}$ is the rate of change in wages from year x to year y.

Rate of Change (x 100) based on 4th Quarter to 4th Quarter data

	Column 1	Column 2	Column 3	
	(97-96) / 96	**(98-97) / 97**	**(98-96) / 96**	**Total Update**
OES	3.3	4.8	8.3	4.4
ECI	3.8	3.7	7.6	3.8
CES	5.3	3.5	9.0	4.1
ES-202	6.5	6.1	13.0	6.4
CPI	1.8	1.6	3.4	1.7

The results in the table above present some interesting information. First, the CPI rate of change (ROC) indicates that this was an extremely stable period for prices. Next, the ES-202 ROC indicates a fairly high ROC in compensation. Again, we remind the reader that the ES-202 wage data includes bonuses and other components of compensation not collected by the OES program. Finally, the data that most closely match the OES ROC are from the ECI and CES programs. For the ECI data this seems reasonable, as the data collected are similar in concept to OES wage data.

The next analysis we conducted was a correlation analysis. For selected data sets we calculated the rate of change at several sub-population, including Occupational Group, State, and Industry Division.

We then examined the correlation of the rate of change developed from other data sources with the rate of change developed from OES data.

Pearson Correlation Coefficients

Factor Level	Data	Correlation	P-Value
Occupational Group	OES vs. ECI	0.7317	0.025
State	OES vs. ES-202	0.2310	0.103
Industry Division	OES vs. ES-202	0.5811	0.101
Industry Division	OES vs. CES	-0.1205	0.776

There are several interesting results presented in this table. The CES ROC, by Industry Division, is negatively correlated with the OES ROC at that level. This indicates that one or two Industry Divisions have substantially

different ROC values for this period in these data sets. We hope to explore this further. The ES-202 data seem to be somewhat correlated with the OES data at the State level, and at the Industry Division level. The correlations , however, are marginally significant given a p-value of ten percent. There is a significant correlation between the OES and ECI data at the Occupational Group level. This is the level at which we currently utilize the ECI estimates to update previous years of OES data.

The next analysis we conducted was to develop two charts (not shown). These charts showed the total update, as defined above, at the state level from the ES-202 data and from the OES data.

The two charts follow the same general pattern. That is, in both charts most of the state updates fall within a narrow band centered around the national total update value. There are, however, some notable exceptions. For example, in the OES data Georgia shows an unusually high ROC value (for the period reviewed), which is not repeated in the ES-202 data. In this case we may be seeing the effect of the sample design on the OES ROC value; that is, the collection of one-third of the large units each year may have contributed to this unusually high ROC value. Likewise, the ES-202 data shows an unusually high ROC value for Washington (for the period reviewed), which is not repeated in the OES data. The ES-202 ROC may by influenced by the bonuses and other non-wage compensation components of large establishments.

The last analysis we conducted was to examine the effect of one year of aging. We graphed the distribution of the following value:

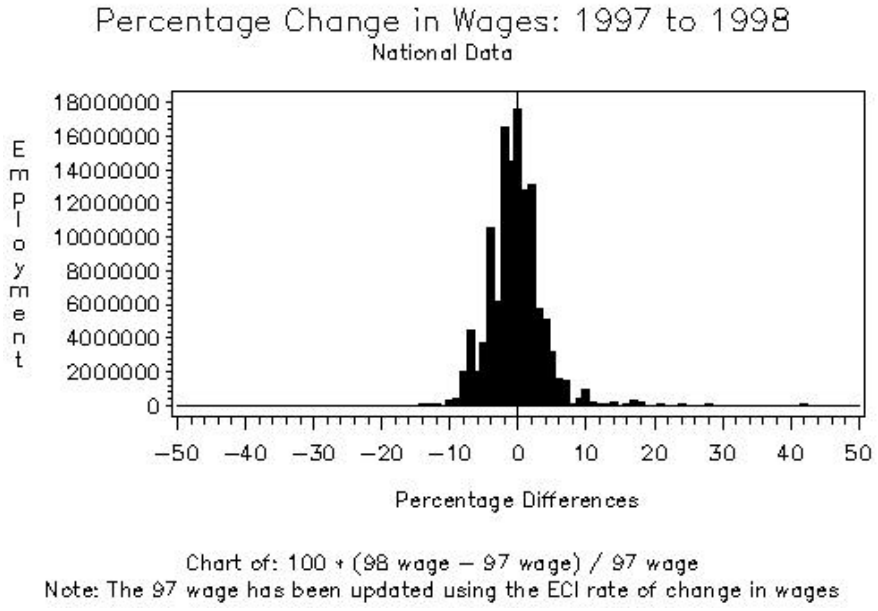
$$d_o = 100 * \left(\frac{\hat{R}_{o,98} - \hat{R}_{o,97A}}{\hat{R}_{o,97A}} \right)$$

Where

o is an occupation, $\hat{R}_{o,98}$ is the 1998 wage rate for occupation o,

$\hat{R}_{o,97A}$ is the updated 1997 wage rate for occupation o, and

δ_o is the centered difference between a 1997 updated wage rate and the 1998 wage rate for an occupation.



The graph above is weighted by occupational employment.

A final quality check that we conducted was to compare the OES wage estimates with estimates from the Bureau's Employer Costs for Employee Compensation (ECEC) program. The ECEC program produces aggregate level wage rates for a number of sup-populations that we can calculate using OES data. The OES estimates were typically comparable to the ECEC estimates that have the same scope.

7. Conclusions and Future Research

Conclusions

In this paper we have described a research project conducted at the BLS. In the current survey, we utilize the national ECI rate of change in wages for nine occupational groups to update previous years wage data. The purpose of the project was to determine if we could develop factors for the rate of change in wages at more detailed levels of geography, industry, employment size, or occupational detail.

In this project we studied geography and industry as classification variables using data from the ES-202 program and industry for the CES survey. The results of this research indicate that the current wage updating procedure is reasonable; that is, the ECI is a good choice for updating OES wages. For the time periods examined in this study, we find that the occupational factors developed from the ECI survey are an improvement over applying an overall ECI factor.

Future Research

It seems intuitive that updating wages at a more detailed level would enhance the bias reduction properties of wage updating. In the future we will try to classify detailed occupations into sub-groups based on their rate of change within the major occupational groups. For example, within a major occupation group, we expect that some detailed occupations have rates of change below the average, near the average, and above the average. We may utilize cluster analysis to develop these sub-groups. We will then assess the utility of these sub-groups in the wage updating process.

8. References

1. Echols, C., Tou, A., and Robertson, K. (1999). "Comparing Alternative Median Wage Rate Estimators for the Occupational Employment Statistics Survey," *1999 Proceedings of the American Statistical Association, Survey Methodology Section.*

2. Robertson, K. and Frugoli, P. (1999). "Statistical Issues for the Redesigned Occupational Employment Statistics Survey," *1999 Proceedings of the American Statistical Association, Government Section.*

3. Rust, R. (1999), "evolution of Expanded Occupational Employment Statistics *and Wage Survey (State Perspective)," 1999* Proceedings of the American Statistical Association, Government Section.

4. U.S. Bureau of Labor Statistics (1997), Handbook of Methods, Washington, DC., Bulletin 249