Using Point and Intervalized Data in Occupational Employment Statistics Survey Estimates October 2010

David Piccone, Teresa E. Hesley
US Bureau of Labor Statistics
2 Massachusetts Ave NE, Suite 4985, Washington DC, 20212

Abstract
The Occupational Employment Statistics (OES) survey conducted by the U.S. Bureau of Labor Statistics produces detailed occupational employment and wage estimates. To reduce response burden, the OES form asks respondents to report the number of employees by wage interval. Many of OES's larger respondents, such as the Federal Government, report data electronically, providing wage rates for each of their employees, or "point data", which OES then reclassifies into wage intervals for estimation. A mean wage rate is calculated for each interval from a secondary source and given to all employees within the interval. Capturing and integrating point-data wage distributions into estimation could increase the accuracy of estimates. This paper describes the research and implementation of mean and percentile wage estimators, along with variances estimators, using available point data wage data.

Key Words: Establishment survey, estimation, Occupational Employment Statistics, wage point data

1. Introduction

The Occupational Employment Statistics (OES) survey is a Federal/State Cooperative program conducted by the Bureau of Labor Statistics[1] (BLS) in partnership with the 50 States, the District of Columbia, and three US territories (Guam, Puerto Rico, and the Virgin Islands). OES is an establishment survey that produces cross-industry employment and wage estimates for over 800 detailed occupations by area, specifically metropolitan statistical areas (MSA) along with the residual areas within the states called the balance of state (BOS) areas. OES also produces national employment and wage estimates for detailed occupations by industry. In order to reduce response burden, OES collects employment data from responding establishment in a matrix format, where the rows are the different occupations found at the establishment and the columns are wage intervals that each employee can be put into. For example, if there are three nurses making $25.00 an hour at a hospital, a 3 would be entered in the entry where the nurse occupation row intersects with the wage interval column containing $25.00 an hour. An example of an OES survey form can be found in appendix A.

In order to produce mean wage estimates using data reported by interval, OES assigns a mean wage rate to all employees in a given wage interval. The mean wage rates for each interval are calculated using point wage data from another BLS survey, the National Compensation Survey (NCS). To produce percentile wage estimates, OES uses a linear interpolation method that assumes a uniform distribution within each interval. These methods work well for most occupational wage estimates, especially if the wage data are normally distributed across several different wage intervals. However, when an

[1] Views expressed in this paper are those of the authors and do not necessarily reflect the views of policies of the Bureau of Labor Statistics

occupation’s employment is concentrated in only a few wage intervals, or is bimodal, the estimates may not be as reliable.

Many larger respondents find that it is easier to send OES an electronic file containing each of their employee’s wage rates, rather than intervalizing each employee’s wage on the survey form. This is the case for the Federal government, where OES receives a file from the US Office of Personnel (OPM) and the US Postal Service (USPS) which contains a “point-data” wage rate for every Federal employee. Prior methodology had OES reclassifying these Federal employees into wage intervals for estimation. In this paper we will describe how OES incorporated Federal point data into their mean and percentile wage estimates, along with current research we are doing to incorporate State government and private sector point data into our wage estimates.

2. Prior Methodology

Starting with the May 2009 estimates, OES began using new mean and percentile wage estimators that utilize both Federal government point wage data along with non-Federal intervalized wage data. Prior to this, OES reassigned Federal point wage data into 12 consecutive non-overlapping wage intervals, making all wage data appear as interval data, which is consistent with the usual interval data reported in OES. To produce mean and percentile wage estimates OES assumes a normal distribution of the wage data across wage intervals and a uniform distribution within intervals to use interval-base estimation techniques. The following sections will describe OES’s methodology for estimating mean and percentile wage rates for intervalized data.

2.1 Percentile Wage Estimator

In order to estimate a percentile wage rate for an occupation, OES must know how the employment is distributed across all of the wages for a particular occupation. Since it is unknown how this employment is distributed within each wage interval, OES makes the assumption that the wage data are distributed uniformly. This assumption sometimes fails for certain occupations, as we will explain later in the paper.

For intervalized data, OES uses a linear interpolation approach for estimating wage percentiles. The first step is finding the occupational weighted employment that falls into each wage interval.

$$\hat{x}_{r,o} = \sum_{i=1}^{n}\left(w_i \cdot x_{i,r,o}\right) \qquad (2.1)$$

Where,

$\hat{x}_{ro}$ = weighted employment estimate for occupation o in interval r

w_i = sample weight from establishment i

$x_{i,r,o}$ = reported employment for occupation o in wage interval r from establishment i

The next step is to find which worker corresponds to the percentile wage estimate. In other words, if we are trying to find the 50th percentile wage estimate we would find the employee who is in the very center of the employment data and estimate their wage rate. This employee is called the „target employee’, and for the $(100 \cdot p)^{th}$ percentile estimate it is calculated by:

$$target\ employee_{p,o} = p\sum_{r=1}^{12}\hat{x}_{r,o} \tag{2.2}$$

Once the target employee is identified, the next step is to find in which wage interval this employee is located. To do this, OES calculates the cumulative employment across the wage intervals. The first interval with a cumulative employment greater than the target employee is where this employee could be found.

$$\hat{x}_{r,o}^{cuml} = \sum_{r=1}^{r}\hat{x}_{r,o} \tag{2.3}$$

Where, $\hat{x}_{r,o}^{cuml}$ is the cumulative employment for wage interval *r*. Also let's denote the wage interval containing the target employment as *r* = *target*.

Once OES knows the wage interval containing the target employee, linear interpolation can be used to find the wage rate for this employee, i.e. to find the estimate for the $(100 \cdot p)^{th}$ percentile wage [OES State Operations Manual].

$$\hat{v}_{p,o} = LB_{target} + \frac{\left(target\ employee_{p,o} - \hat{x}_{target,o}^{cuml}\right)}{\hat{x}_{target,o}} \cdot \left(LB_{target+1} - LB_{target}\right) \tag{2.4}$$

Where,

$\hat{v}_{p,o}$ = the p^{th} percentile wage estimate for occupation *o*

LB_{target} = lower wage bound of the target interval

2.2 Mean Wage Estimator

Occupational mean wage estimates cannot be calculated using just the OES grouped wage data, since each worker's wage rate within a wage interval is unknown. OES assigns each worker an interval mean wage rate calculated from NCS data for estimation. NCS collects individual wage rates for the private sector and state and local government employees [Survey Methods and Reliability Statement, 2009].

OES previously researched two different methods of assigning wage rates to the grouped employees by using:

1. Arithmetic mean of interval bounds: $c_r = \dfrac{LB_r + LB_{r+1}}{2}$
2. Geometric mean of interval bounds: $c_r = (LB_r \times LB_{r+1})^{1/2}$

   Where, c_r would be the wage rate given to each employee found in interval r.
   Note: the arithmetic mean method is a very common approach for interval-based estimation, and works well when data within the intervals are distributed uniformly.

The research showed that the NCS-based mean, the Geometric mean, and the Arithmetic mean for each of the 12 intervals were very similar, the largest differences occurring in the higher wage intervals. The empirical results showed that estimates produced using the Arithmetic mean had a bias of about two percent, which indicates that the OES wages within an interval are not distributed uniformly. Using the Geometric mean for each interval reduced the bias to about one percent, and the NCS-based means dropped the bias to nearly zero [Analysis of Alternative Procedures for Interval Mean Wage Rates, 2000].

The mean hourly wage estimate for an occupation is the total weighted hourly wages divided by the weighted survey employment. Estimates of mean hourly wages are calculated using a standard grouped data formula:

$$\hat{R}_o = \frac{\sum_{i=1}^{n}\left(w_i \cdot \hat{y}_{i,o}\right)}{\sum_{i=1}^{n}\left(w_i \cdot x_{i,o}\right)} \tag{2.5}$$

$$\hat{y}_{i,o} = \sum_{r=1}^{12}\left(c_r \cdot x_{i,r,o}\right) \tag{2.6}$$

$$x_{i,o} = \sum_{r=1}^{12} x_{i,r,o} \tag{2.7}$$

Where,

$\hat{R}_o$ = mean hourly wage rate for occupation *o*
w_i = sample weight from establishment *i*
$\hat{y}_{i,o}$ = unweighted total hourly wage estimate for occupation *o* from establishment *i*
c_r = NCS-based mean hourly wage for wage interval *r*
$x_{i,r,o}$ = reported employment for occupation *o* in establishment *i* in wage interval *r*

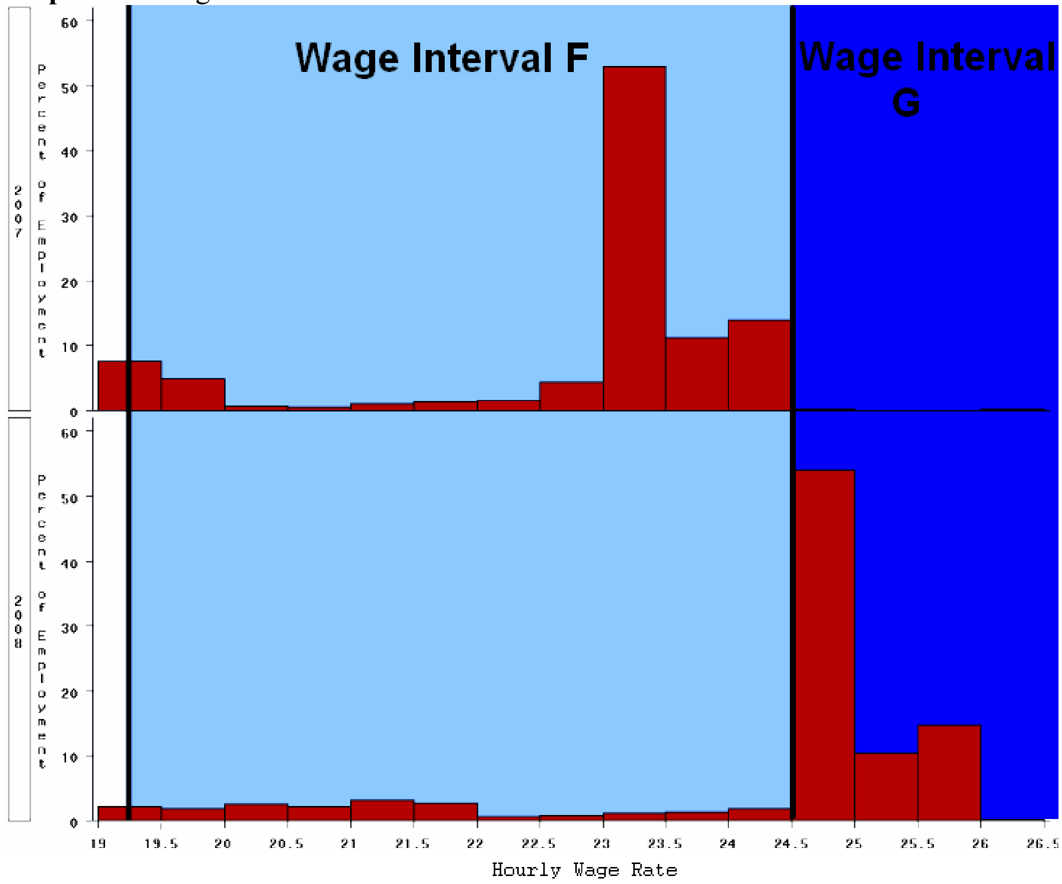
The difference between this methodology and OES's current methodology is rather than using formula 2.5 for all the survey data, the current methodology uses each Federal government employee's actual wage rate during estimation. We will discuss this in more detail later on in this paper.

2.2.1 Issues with the Prior Mean Wage Rate Estimator

This prior methodology gives practically unbiased mean hourly wage estimates for most occupations, especially when the wage data are found in several different wage intervals. When an occupation's wage data are concentrated in only a few wage intervals or show a multi-modal distribution, there is potential for this methodology to produce bias estimates. This tends to happen more in the government sector since many government employees are paid based on salary tables. This also happens when an occupation has a small amount a wage data for the estimate.

An example of the prior methodology producing bias estimates happened in OES's May 2008 estimates. The wage distribution for the Postal Service Clerk occupation was extremely narrow, with nearly all of these workers earning between $23.00 and $24.50 per hour. These wages are at the top of the 6th OES wage interval, which we call wage interval F. In 2008 USPS employees received a 5% cost of living adjustment, meaning these same workers were now earning between $24.50 and $25.50 per hour. The lower bound of the 7th OES wage interval, or wage interval G, is $24.50, meaning nearly all wage data for postal service clerks shifted up one wage interval from 2007 to 2008. This 5% increase in wages resulted in a 23% increase in the mean hourly wage estimate for this occupation. OES receives a census of point wage data for postal service clerks from the USPS, which includes a point-data wage rate for every worker. OES had the ability to report the true mean hourly wage rate instead our prior methodology gave almost all these employees the NCS-base mean wage rate for wage interval F, in 2007, and NCS-base mean wage rate for interval G in 2008, which tend to be towards the middle of the wage interval. This means in 2007 OES under-estimated the mean hourly wage rates for postal service clerks and in 2008 these wage rates were over-estimated. The following graph illustrates this issue:

Graph 2.1 – Wage Data Distribution for Postal Service Clerks – In 2007 and 2008



This same issue happened in 2008 for two other postal-service specific occupations, Postal Service Mail Carriers and Postal Mail Sorters, Processors and Processing Machine Operators [Warren, 2010].

2.3 Mean Wage Variance Estimator

Since the mean wage rate estimator uses data from another survey, the variance estimator uses a model approach to account for the added variability associated with using the NCS interval means for *all* OES data [Miller, 2005]. OES's variance estimator has four components, three of which account for the variability of using interval means.

$$v(\hat{R}_o) = VC_{s,o} + VC_{c,o} + VC_{e,o} + VC_{w,o} \tag{2.8}$$

Where,

$v(\hat{R}_o)$ = variance estimator for the wage rate estimate

$VC_{s,o}$ = design-based variance component

$VC_{c,o}$ = variance component for the variability of using the overall NCS mean estimates for each wage interval

$VC_{e,o}$ = variance component for the variability of establishment level differences in wages within wage intervals

$VC_{w,o}$ = variance component for the variability of worker level differences in wages within wage intervals

Since all OES wage data is still put into wage intervals for the purpose of variance estimation, every unit contributes to the overall variance of wage rate $\hat{R}_o$. For instance, the Federal government data is still contributing to the overall mean wage variance because even though we receive a census of this data, we intervalize it for the use in our variance estimator. The design-base variance component would be zero, since we are receiving all Federal data, but the other three components would be non-zero. Incorporating point data into our variance estimator is described below as part of our ongoing research.

Between the mean wage bias and the added variability to our estimates, OES wanted to find a way to incorporate point data wage rates into our mean and percentile wage estimates. The most logical first step was creating estimators that could use the census of Federal government point data OES receives yearly.

3. Incorporating Federal Point Data

OPM and USPS send OES a census of Federal employment and wage data once a year to be included in our yearly estimates. These data are self representative and complete, meaning there are no weighting or nonresponse issues that need to be considered.

3.1 Percentile Wage Estimator

OES implemented a new percentile estimator in May of 2009, which combines the Federal point data wage rates and non-Federal interval data wages. This estimator is similar to the original estimator, except for the assumption made about the wage data distribution within the wage intervals. Previously OES assumed that all the wage data are distributed uniformly within each wage interval. The new estimator still assumes the non-Federal, intervalized wage data are uniformly distributed, but since the federal point data wage rates are known we can place the Federal employment at its exact location within the wage intervals.

The first step is to examine the wage distribution of the non-Federal interval data for the percentile you are trying to estimate. OES still assumes that the non-Federal data are distributed uniformly within each wage interval, but instead of being continuously uniform OES now assumes this data are now discretely uniform by penny value. This will make incorporating the Federal point data easier. The calculation of non-Federal employment per penny is as follows:

$$\widehat{EP}_{r,o} = \frac{\sum_{i \in NonFed}\left(w_i \cdot x_{i,r,o}\right)}{100 \cdot (LB_{r+1} - LB_r)} \tag{3.1}$$

Where,

$\widehat{EP}_{r,o}$ = non-Federal employment per penny for occupation *o*, in wage interval *r*

Note: $i \in NonFed$ means all establishments that are non-Federal

Once the non-Federal interval employment per penny is known for all wage intervals, the total employment can be calculated for every penny across the wage intervals. This is the non-Federal plus the Federal employment per penny.

$$\hat{x}_{j,r,o} = \widehat{EP}_{r,o} + X_{j,o} \tag{3.2}$$

Where,

$\hat{x}_{j,r,o}$ = occupation o's employment on penny j, found in wage interval r
Note: The penny value j has the constraints: : $LB_r \leq j \leq (LB_{r+1} - 0.01)$

$X_{j,o}$ = the number of Federal employees found on penny j

Just as before, we must find the employee that corresponds to the percentile wage estimate. We call this the target employee. The next step is to find which penny interval this employee is found in. To do this, the cumulative employment per penny must be calculated.

$$\hat{x}_{j,r,o}^{cuml} = \sum_{j=1}^{j} \hat{x}_{j,r,o} \qquad (3.3)$$

Where, $\hat{x}_{r,o}^{cuml}$ is the cumulative total employment for penny j.

The target employee will be found in the first penny interval with a cumulative employment greater than the target employee. For our current percentile estimator this penny value will be our percentile estimate, unless the target employee is exactly equal to the cumulative employment for a penny value. In that case we take the midpoint wage value between the penny that has a cumulative employment equal to the target employee, and the next penny with a cumulative employment greater than our target employee.

$$\hat{v}_{p,o}^{*} = \begin{Bmatrix} j & where\ the\ first\ \hat{x}_{j,k,o}^{cuml} \geq target\ employee_{p,o}\ is > \\ \dfrac{j + j^{*}}{2} & where\ the\ first\ \hat{x}_{j,k,o}^{cuml} \geq target\ employee_{p,o}\ is = \end{Bmatrix}$$

Where,

$\hat{v}_{p,o}^{*}$ = the current p^{th} percentile wage estimate for occupation *o*

j^{*} = next penny with a cumulative employment greater than our target employee, if *j*'s cumulative employment is equal to the target employee

3.2 Mean Wage Estimator

Beginning in 2009 OES has been using a mean wage estimator that uses each Federal employee's actual wage rate, instead of putting these employees into wage intervals. This decreases the chance of introducing bias into our estimates, as described in section 2.2.1.

Since the Federal point wage data and non-Federal interval wage data are disjoint, it is easy to divide the two data types out when estimating occupational mean wage rates. The estimate is calculated by dividing the total wages for an occupation by the total employment. This could be broken out by dividing the occupation's total Federal wages plus its total non-Federal wages by total Federal and non-Federal employment.

$$\hat{R}_{o}^{*} = \frac{\sum_{i \in Fed} Y_{i,o} + \sum_{i \in NonFed}(w_i \cdot \hat{y}_{i,o})}{\sum_{i \in Fed} X_{i,o} + \sum_{i \in NonFed}(w_i \cdot x_{i,o})} \qquad (3.4)$$

Where,

$\hat{R}_{o}^{*}$ = current mean hourly wage rate for occupation *o*

$Y_{i,o}$ = Federal hourly wages (calculated from point-data wage rates) for occupation *o*, from establishment *i*

$X_{i,o}$ = Federal employment for occupation *o*, from establishment *i*

Please note that the first summation in the numerator and denominator are constants, since OES receives a census of Federal data.

3.3 Mean Wage Variance Estimator

Since there is a new mean wage rate estimator that can incorporate Federal point data, OES has begun researching a new mean wage variance estimator that could also incorporate Federal point data. Using Taylor Series variance estimation, OES found a new wage variance estimator.

$$v(\hat{R}_o^*) = \frac{1}{\left(X_{P,o} + \hat{X}_{I,o}\right)^2} \cdot \left\{ v(\hat{Y}_{I,o}) + \hat{R}_o^{*2} \cdot v(\hat{X}_{I,o}) - 2 \cdot \hat{R}_o^* \cdot cov(\hat{Y}_{I,o}, \hat{X}_{I,o}) \right\} \qquad (3.5)$$

Where,

$X_{P,o} = \sum_{i \in Fed} X_{i,o}$
= Federal point data total employment constant for occupation o

$\hat{X}_{I,o} = \sum_{i \in NonFed} (w_i \cdot x_{i,o})$
= non-Federal interval data total employment estimate for occupation o

$\hat{Y}_{I,o} = \sum_{i \in NonFed} (w_i \cdot \hat{y}_{i,o})$
= non-Federal interval data total wage estimate for occupation o

From examination you can see that only the intervalized non-Federal data is contributing to the variance components for the new mean wage variance estimator. This is ideal since these data are not random, but constant, since we receive a census of it. OES already produces an estimate for $v(\hat{X}_{I,o})$, and is currently researching ways to estimate $v(\hat{Y}_{I,o})$ and $cov(\hat{Y}_{I,o}, \hat{X}_{I,o})$ using a model-base approach similar to that mentioned in section 2.3.

4. Other Current and Future Research

Currently OES is researching ways to incorporate State government point wage data into our wage estimators. Like the Federal data, OES receives a census of State government data. However, not all States can easily give us its wage rates as point-data, so instead of having all State wage data in the same format, there will be a mix of interval and point data. Also there is a chance for nonresponse for State data, so imputation must be researched.

OES will also like to research using private sector point wage data. Many larger units respond to OES by sending data-dump files that contain point wage data for all of their employees. Currently, these employees are reclassified into wage intervals. OES's wage estimates could improve if we can incorporate these data into our estimators. Careful thought will be needed since private data are weighted and have a chance of nonresponse.

Acknowledgements

Marie C. Stetser, Bureau of Labor Statistics
Martha A. Duff, Bureau of Labor Statistics
Laura Train, Bureau of Labor Statistics
Sertan Akinci, Bureau of Labor Statistics
Laurie Salmon, Bureau of Labor Statistics
George Stamas, Bureau of Labor Statistics

References

Miller, Steve. "A Simple Model for Variances of OES Wage Rate Estimates." 25 Aug. 2005. Bureau of Labor Statistics. Print.

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

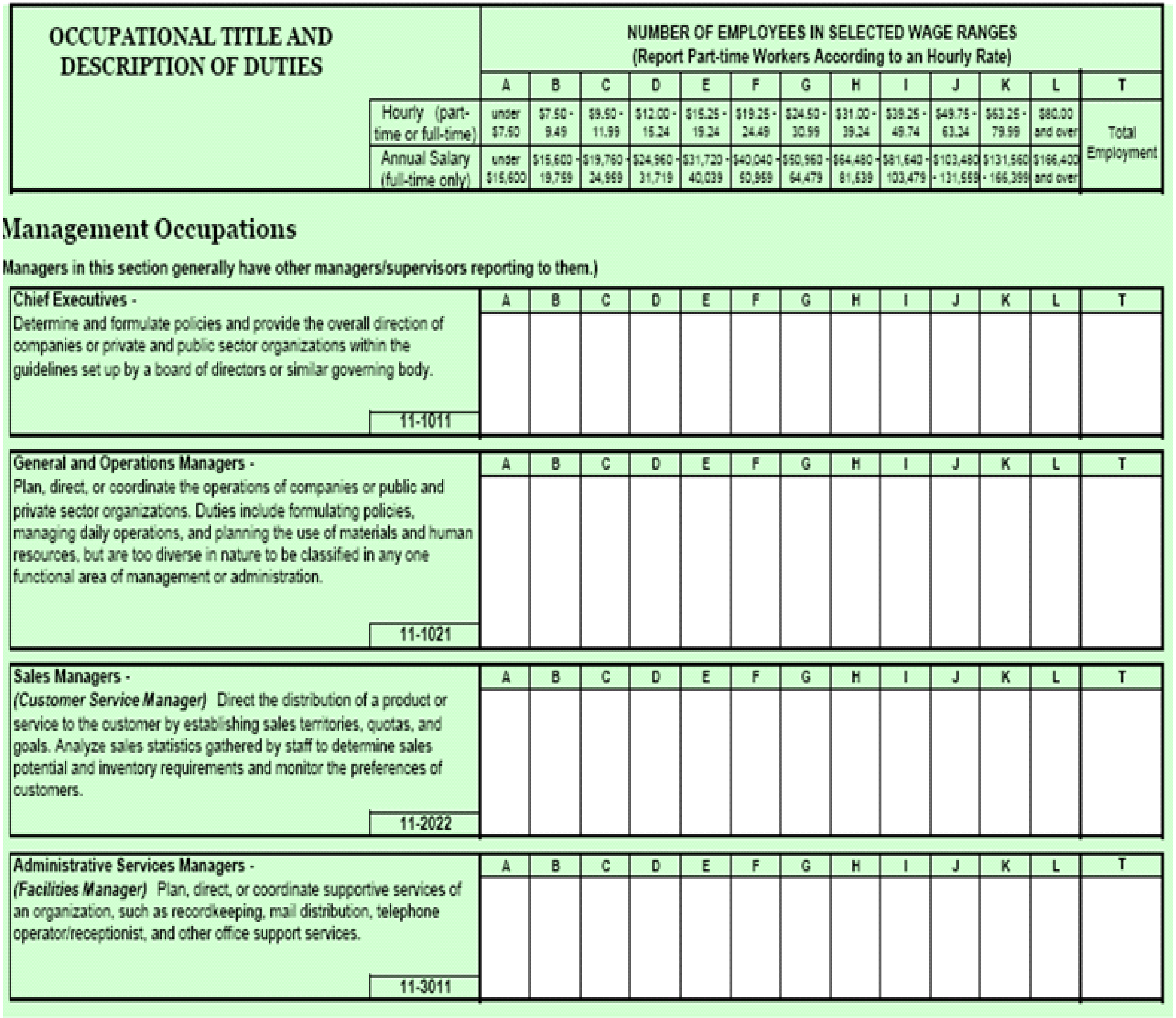
"Occupational Employment Statistics Survey: Analysis of Alternative Procedures for IntervalMean Wage Rates, Supplement II". December 2000. Bureau of Labor Statistics. Print.

OES State Operations Manual. Bureau of Labor Statistics Reference Manual. Print.

"Survey Methods and Reliability Statement for the May 2009 Occupational Employment Statistics Survey." Web. <http://www.bls.gov/oes/2009/may/methods_statement.pdf>.

Warren, Zachary. Internal Memo Summarizing the Postal Service Issue. Bureau of Labor Statistics. 2010.

Appendix A: Example of an OES Survey Form



OCCUPATIONAL TITLE AND DESCRIPTION OF DUTIES		NUMBER OF EMPLOYEES IN SELECTED WAGE RANGES (Report Part-time Workers According to an Hourly Rate)												
		A	B	C	D	E	F	G	H	I	J	K	L	T
	Hourly (part-time or full-time)	under $7.50	$7.50 - 9.49	$9.50 - 11.99	$12.00 - 15.24	$15.25 - 19.24	$19.25 - 24.49	$24.50 - 30.99	$31.00 - 39.24	$39.25 - 49.74	$49.75 - 63.24	$63.25 - 79.99	$80.00 and over	Total Employment
	Annual Salary (full-time only)	under $15,600	$15,600 - 19,759	$19,760 - 24,959	$24,960 - 31,719	$31,720 - 40,039	$40,040 - 50,959	$50,960 - 64,479	$64,480 - 81,639	$81,640 - 103,479	$103,480 - 131,559	$131,560 - 166,399	$166,400 and over	

Management Occupations

Managers in this section generally have other managers/supervisors reporting to them.)

	A	B	C	D	E	F	G	H	I	J	K	L	T
Chief Executives - Determine and formulate policies and provide the overall direction of companies or private and public sector organizations within the guidelines set up by a board of directors or similar governing body. 11-1011													

	A	B	C	D	E	F	G	H	I	J	K	L	T
General and Operations Managers - Plan, direct, or coordinate the operations of companies or public and private sector organizations. Duties include formulating policies, managing daily operations, and planning the use of materials and human resources, but are too diverse in nature to be classified in any one functional area of management or administration. 11-1021													

	A	B	C	D	E	F	G	H	I	J	K	L	T
Sales Managers - *(Customer Service Manager)* Direct the distribution of a product or service to the customer by establishing sales territories, quotas, and goals. Analyze sales statistics gathered by staff to determine sales potential and inventory requirements and monitor the preferences of customers. 11-2022													

	A	B	C	D	E	F	G	H	I	J	K	L	T
Administrative Services Managers - *(Facilities Manager)* Plan, direct, or coordinate supportive services of an organization, such as recordkeeping, mail distribution, telephone operator/receptionist, and other office support services. 11-3011													