Using OES Data in Federal Pay Comparability: A Regression-Based Approach

March 2009

A report of the Wage Estimates for Federal Pay Comparability Team

Team Members

Matt Dey
Maury Gittleman
Mike Lettau
Steve Miller

Introduction

Under the Federal Employees Pay Comparability Act (FEPCA), the President's Pay Agent must use salary surveys conducted by Bureau of Labor Statistics (BLS) to set locality pay. At present, the sole survey that is used in this endeavor is the National Compensation Survey (NCS). Given the possibility that future budget shortfalls may lead to cuts in the wage sample of the NCS, a team was formed to investigate whether one can use data from the Occupational Employment Statistics (OES) program instead of or in addition to data from the NCS.

To this end, the team proposed two alternative estimators – one called an interval method and the other a regression method -- that combine NCS and OES data in different ways. In an internal research paper, we compare and contrast the results under these two methods to the current approach that is used to generate pay gaps.[1] Four main themes emerged from this analysis. First, we demonstrate that it is, indeed, feasible to use OES data for pay comparability, but the OES data cannot be used by itself; it must be combined with NCS data. Second, the proposed methods both appear to be capable of estimating reasonable-looking pay gaps with greater precision than does the current approach. Third, the proposed methods are more robust to cuts in the NCS sample, assuming the OES sample sizes would remain constant, than is the current approach. Fourth, the proposed methods can both be used to extend the estimation of pay gaps to areas that are not present in the NCS sample.

After careful investigation of the advantages and disadvantages of the two proposed methods, the team believes that the regression method is clearly better suited to produce the non-Federal salary estimates required to calculate area pay gaps.

The remainder of the report proceeds as follows. In the next section, we describe the current NCS-based approach to Federal pay comparability. After describing the potential benefits of incorporating OES data, we present the details of the regression method combining both OES and NCS data. Following that, we turn to the actual numbers and present our analysis of the current method versus the proposed approach. The final section provides a restatement of our main themes and some indication of where we think further investigation may be warranted.

Current Approach Using NCS Data Only

For the 2008 Pay Agent's Report to the President on Locality Pay for 2010, the Pay Agent was interested in non-Federal pay levels and non-Federal/Federal pay gaps for 31 areas, including one residual category for the Rest of the United States (RUS).[2] BLS provided, for each of these areas, non-Federal salary estimates for up to 15 grades for the five PATCO job families (professional, administrative, technical, clerical and officer),

[1] This document is available upon request.
[2] If one includes Raleigh, the number of pay agent areas is actually 32, but the NCS sample did not contain sufficient data for this locality for estimates to be produced.

resulting in 67 estimates per area.[3] Using local Federal employment weights by grade and PATCO, the Office of Personnel Management (OPM) then averaged the estimates provided by BLS to arrive at a single number for non-Federal pay in each area.[4] These numbers were then compared to the average salary for Federal white-collar employees in each locality, and a non-Federal/Federal pay gap was then calculated.

To help evaluate the proposed methods that make use of OES data, it is useful to take a few steps back, to see how NCS data are used to arrive at the 67 estimates per area. For the purposes of the latest Pay Agent's report, there were nearly 5,000 different jobs held by Federal white-collar employees, where jobs are defined by GS series, grade and whether the job is supervisory or not. Because jobs in the NCS are not defined in the same manner, a crosswalk was created that maps these 5,000 jobs to jobs defined by six-digit Standard Occupational Classification (SOC) code and grade. As many Federal jobs map to the same SOC code-grade combination, the creation of the crosswalk resulted in the identification of about 2,000 unique SOC code-grade pairs. The current methodology asks BLS to come up with an estimate of average pay in each of these jobs in each area, and then to use national Federal employment weights to arrive at the 67 estimates for each locality. Multiplying the number of jobs by the number of localities reveals that more than 60 thousand estimates were required.

To compute these estimates, BLS made use of the data in the 2007 NCS sample.[5] To get an idea of the strain that these calculations place on the NCS data, it may be worth noting that, taking all the localities together, there were approximately 23,000 establishments that responded to the NCS that year. Each establishment provided information on anywhere from one to 32 different jobs, with most reporting data for eight or fewer jobs. In order for the NCS sample to better correspond to jobs that are relevant to pay comparisons for the Federal white-collar workforce, a number of restrictions were then applied to the resulting samples for each locality. Only those jobs that are full-time, have a grade attached, and have valid wage information were included. The sample was then further limited to jobs in the crosswalk file provided by OPM and to jobs that would not be classified above GS-15 in the Federal Government. After applying these restrictions, a national sample of 41,250 jobs remained. Put differently, the number of jobs in this sample was less than the number of job-average estimates needed for the calculations.

The estimates were computed in two steps, one involving direct estimates at the level of the locality, the other involving indirect estimates generated by a national regression model. If data meeting publication criteria were available to calculate average pay in an SOC code-grade combination in a given area, a direct estimate was computed. In cases where a direct estimate was not available, a regression model, which will be described below, was used to estimate average pay in area-SOC code-grade cells.[6] Then, using

[3] There is no federal employment for eight of the grade-PATCO job family combinations.
[4] OPM aged the data by area, so that the reference date became March 2008. Unless otherwise noted, the estimates presented in this report have this reference date as well.
[5] Establishments located in Alaska and Hawaii were excluded, as the Pay Agent's interest was limited to jobs in the contiguous United States.
[6] Because 25 occupations did not appear in the national sample used for the regression model, indirect estimates could not be produced for 118 cells in each area.

national Federal employment weights, the direct and modeled estimates were averaged up to job family and grade, to produce the 67 numbers needed for each area.

It is worth stressing that, as a result of the relatively small size of the usable NCS sample and the fact that there was Federal white-collar employment in numerous SOC code-grade combinations that are unlikely to occur in the non-Federal sector, the majority of the estimates, as weighted by national Federal employment, come from the model. In the past year, the percentage modeled ranged from 33 percent for RUS to 86 percent for Milwaukee. But this range is misleading because the percentage for RUS is an outlier, as it is the only case where the majority of estimates was not modeled. Three areas had between 50 and 60 percent of the estimates modeled, 9 were between 60 and 70 percent, 11 were between 70 and 80 percent, and the remaining 7 were above 80 percent.

As a point of comparison with the approach using OES data that will be discussed subsequently, it may be useful to briefly discuss the regression model that is currently used to produce the indirect estimates.

The regression model is of the following form:

[1]

$$\ln W_m = \alpha + \sum_{a=1}^{A-1} \beta_a \cdot AREA_{ma} + \sum_{o=1}^{O-1} \chi_o \cdot OCCUP_{mo} + \delta \cdot XGRADE_m + \phi \cdot XGRADESQ_m + \varepsilon_m$$

where $\ln W_m$ is the natural log of the average hourly wage rate of the m^{th} job, *AREA* is a vector of dummy variables indicating locality, *OCCUP* is a vector of dummy variables for occupation as defined by six-digit SOC code (XX-YYYY), *XGRADE* is a transformation of grade[7], and *XGRADESQ* is the square of *XGRADE*. β, χ, δ, and ϕ are the corresponding coefficients, α is a constant term, and ε is the error term. Areas are indexed by *a* and are numbered from one to *A*, occupations are indexed by *o* and are numbered from one to *O*.

The functional form of the model, which was chosen in line with OPM's preferences, embodies some strong assumptions. First, the three components of the model – area, occupation and grade (xgrade) – enter directly without any interactions. Put differently, estimated area pay differentials will be the same regardless of occupation and grade combination, occupational pay differentials will not vary by area and grade, and the returns to grade will be the same regardless of area and occupation. Second, returns to grade are assumed to be quadratic in xgrade.

Clearly, it is possible to come up with examples where one would not expect the first set of restrictions to hold, owing to differences across local labor markets. Past research suggests, however, that owing to the relatively small size of the sample that is used for modeling, any reduction in bias from having a less restrictive model with more

[7] xgrade is the same as grade if grade is less than 12. For grade 12, xgrade=13, for grade 13, xgrade=15, for grade 14, xgrade=17 and for grade 15, xgrade=19.

parameters would be offset by increases in the variance of the estimates. As an alternative to the assumption about the functional form of grade (xgrade), one can imagine using a set of dummy variables for each grade. While such an approach has the advantage of allowing a freer estimation of returns to grade, it has the disadvantage of increasing the number of parameters that need to be estimated with a small sample, as well as necessitating the estimate of a parameter for grades that are not well represented in the regression sample. Nonetheless, even if these assumptions are reasonable for a model estimated only on NCS data, starting afresh with OES data, one would not start out by imposing these restrictions.

Potential Benefits of Incorporating OES Data

Before proceeding to the details of the proposed method that combines both NCS and OES data, we will quickly discuss why incorporating OES data into the process to produce wage estimates for the President's Pay Agent potentially provides important benefits.

First, pinning down area-SOC code mean wages with the OES data frees up the NCS data to allow a richer specification of the grade level effects. As discussed above, the current regression has a constant term (1 parameter), occupation dummy variables (258 parameters), area dummy variables (31 parameters), and a quadratic grade level specification (2 parameters). Therefore, the regression has a total of 292 parameters, but only two of these parameters determine the behavior of mean wages across grades. By pinning down a majority of these parameters with the OES data, we could allow a more robust specification of how grade affects wages.

Second, since the OES sample is much larger than the NCS sample, we would expect, overall, efficiency gains in the estimates of mean wages by occupation and area. In some cases, because the current OPM model borrows strength by pooling data across occupations and areas and because a direct estimate of mean wages from the OES may be based on a small sample, there may not be gains in precision. But, at the very least, one could run a similar (to the current OPM model) log wage regression with only main effects on the OES data and expect to obtain more precise estimates. Moreover, since the OES sample is larger, we could perhaps allow a richer specification of how area and occupation interact in the wage process (i.e., we could relax the strong assumption that, regardless of occupation, an area will always be high wage and another area will always be low wage).

Third, since the OES samples establishments in all metropolitan areas of the country, we may be able to extrapolate the level effects from the NCS to the "unsampled" areas to generate Federal pay gaps for all metropolitan areas. Of course, this requires that the estimated grade effects do not vary across (detailed) areas and that we are comfortable in the extrapolation, but if we have sufficient confidence in the robustness of these estimates, the incorporation of the OES potentially increases the number of localities for which estimates can be provided to the Pay Agent.

Finally, with the possibility of a sample cut to the NCS, the question of whether OES can provide some support for the Pay Agent estimates seems quite natural. If the NCS sample is reduced to the index only portion of the sample, for example, can we maintain the quality of the Pay Agent products by augmenting the remaining NCS data with the OES data?

The Regression Method: An Estimation Method that Combines OES and NCS data

This method is somewhat similar to the procedure used currently when the sample size from the NCS is not sufficient to provide a direct estimate of the average wage rate for an area-SOC code-grade combination. As noted, in this case, the current procedure uses the prediction from a log-wage regression with main effects for area, occupation, and grade, but with no interactions among them. The proposed regression method uses a regression equation with an additive effect on job grade on the expected log wage rate. However, the proposed method applies this grade effect to the average wage rate by area and occupation from the OES data, instead of relying exclusively on data from the NCS.

Define $\ln W_i$ as the natural log of the hourly wage rate for the ith individual in the NCS sample[8], and define $\ln W_{oa}^{OES}$ as the natural log of the average wage rate for occupation o in area a from the OES sample. The regression method uses the following equation:

[2]
$$\ln W_i - \ln W_{oa,i}^{OES} = \theta + \sum_{v=0}^{1}\sum_{g=1}^{G}\kappa_{gv}\cdot Leveled_{v,i}\cdot Group_{g,i} + \sum_{g=1}^{G}\lambda_g\cdot FT_i\cdot Group_{g,i} + \sum_{l=1}^{15}\pi_l\cdot Level_{l,i} + \nu_i$$

where $Leveled_{v,i}$ is a vector of indicator variables for whether the i^{th} individual is in a job that is leveled or not, FT_i is an indicator variable for whether the individual's job is full-time, $Group_{g,i}$ is a vector of expected grade level group indicator variables, and $Level_{l,i}$ is a vector of grade indicator variables.[9] The term ν_i is the residual for the regression equation.

The estimate for the average full-time wage rate for grade l of occupation o in area a is then equal to the following.

[3]
$$\hat{W}_{oal} = W_{oa}^{OES} \times \Delta_{ol}$$

[8] The sample used for the regression method is significantly bigger than that used under the current approach, as it includes those who work full-time, those in an SOC code that is in the crosswalk but in a grade that is not, and those with missing grade information.

[9] The expected grade level groups are roughly defined by the midpoint of the expected minimum and maximum grade levels defined by the NCS, plus a separate category for nurses.

where

$$\Delta_{ol} = \Delta_{g(o)}^{-1} \times \exp\left(\hat{\theta} + \hat{\kappa}_{g(o),1} + \hat{\lambda}_{g(o)} + \hat{\pi}_{l}\right)$$

$$\Delta_{g(o)} = \frac{\sum_{i \in g(o)} \omega_i \times \exp\left(\hat{\theta} + \sum_{v=0}^{1} \hat{\kappa}_{g(o)v} \cdot Leveled_{v,i} + \hat{\lambda}_{g(o)} \cdot FT_i + \sum_{l=1}^{15} \hat{\pi}_l \cdot Level_{l,i}\right)}{\sum_{i \in g(o)} \omega_i}$$

and ω_i represents the NCS weight of the i^{th} individual. The term $\Delta_{g(o)}$ is an adjustment to ensure that the detailed mean wages for the area-SOC code-grade-hours combinations add up to the overall wage rate for occupation o in area a from the OES sample.

The regression equation includes expected grade group indicator variables as an alternative to estimating the coefficients of the regression separately for each expected grade group. This is done for two key reasons. First, some area-SOC code-grade combinations of interest to the Pay Agent would not be covered if we estimated the grade effects separately for each group. For example, the Pay Agent wants a wage estimate for grade 9 security guards. According to the NCS, the expected grade range for security guards is 1 to 5, so the Pay Agent is asking for data that the NCS expects there to be little chance of getting. For the security guard example, there are only a handful of observations in the group that contains security guards that are in grade 9.

Second, if we do not somehow correct for the expected grade group that an occupation falls into, we are comparing apples and oranges. For example, security guards who are paid close to the average wage are likely to be in grade 3 (security guards are expected to be in grades 1 to 5), while paralegals who are paid close to the average wage are likely to be in grade 7 (they are expected to be in grades 5 to 9). By correcting for expected grade group in the regression we are able to shift the wage differentials by group for each grade, while nevertheless being able to produce the wage estimates that the Pay Agent desires.

Analysis

Our analysis can be roughly divided into three main subsections, with corresponding tables. The first subsection addresses the feasibility of the proposed approach and how the results of the current method versus proposed method compare and contrast. In the second subsection, we turn our analysis of the impact of a reduction in the NCS sample from the current, full sample to the current, index sample. Finally, we explore the issue of whether having the broader geographic coverage of the OES can be used to extend pay gap calculations to areas not in the NCS.

Table 1 (Non-Federal Salaries) & Table 2 (Pay Gaps)

Tables 1 and 2 present very similar information, except that Table 1 is presented in terms of estimated non-Federal salaries for the Pay Agent areas, while Table 2 provides information in terms of pay gaps. The two tables provide an immediate answer to the question of whether the OES data alone can be used to calculate non-Federal salaries and pay gaps. A comparison of the results in the columns for "Current Method" with those from the columns for "OES Only" indicates that the answer is "no", at least if an important criterion is that the results should be similar to those under the present approach. Using the OES data alone results in much lower estimates of non-Federal salaries and correspondingly narrower pay gaps. The stark differences do not seem to be the result of fundamental differences between the OES and NCS, but rather to the fact that grade information is not present in the OES. When grade information in the NCS is ignored and calculations are made in a manner similar to what one does in using the OES alone, the estimates of non-Federal salaries and pay gaps are again much lower than under the current method, as shown in the column labeled "NCS with No Grade Information".

The remaining columns of the two tables provide means, standard errors, and relative standard errors (RSEs) for estimates from an NCS-only approach that relies exclusively on modeled data ("OPM Model") and the regression method. Ideally, we would have standard errors and RSEs from the current method, but these are not currently produced. Owing to the complex way the current approach combines local direct estimates and national indirect estimates, it was deemed too difficult to use this approach in our current analysis, not least because it would have required the development of a methodology to compute standard errors. Instead, much of the analysis revolves around comparisons between the OPM model and the regression method. As can be seen from a comparison of the means from the OPM model with those from the current method, these two NCS-based approaches are very similar in terms of estimates of non-Federal salaries and of pay gaps.[10] This closeness should not be surprising given the widespread use of the modeled estimates noted above. Thus, any estimates for the OPM Model are likely to be very similar to what would have been generated by the current method, could that have been replicated for this project.

How does the attempt to combine the two datasets fare? A glance at the tables suggests that, unlike the case with OES data alone, the estimates from the NCS-OES regression method are in the same ballpark as those from the current method and OPM model. The unweighted correlation between the non-Federal salaries estimated by the OPM model and the regression method is 0.97, indicating consistency across the methods in terms of which areas are high-paying and which are low-paying. Though the corresponding correlation for the pay gaps themselves is lower (0.90), the pay gaps are also fairly close between the two methods. With each area given equal weight, the average of the pay gap over the areas is 45.5 percent for the OPM model, while it is 49.9 percent for the regression method.

[10] The unweighted correlation for non-Federal salaries is 0.99 and that for the pay gaps is 0.97. The average pay gap under the current method is 46.1 percent versus 45.5 percent for the OPM model.

A pictorial look at how close the regression method is to the current method in terms of the ranking of pay gaps is provided in Figure1. If the pay gaps had exactly the same ordering under both methods, all localities would be on the 45 degree line. That is clearly not the case, but the Figure does demonstrate a high degree of similarity between the two sets of rankings.

How does the regression method fare in terms of the precision of its estimates? Though the standard errors and RSEs from this approach should be viewed as lower bounds because in calculating them the OES data were treated as fixed, this approach appears to have greater precision than the OPM model. On average, the RSEs for the pay gaps from the OPM model are about 177 percent higher than those from the regression method. Thus, it appears that using the OES data to pin down occupation-area mean wages does, in fact, play a useful role in improving the efficiency of the estimates.[11]

Table 3 (Non-Federal Salaries) & Table 4 (Pay Gaps)

Next, we turn to our analysis of the impact of a reduction in the NCS sample from the current, full sample to the current, index sample, which cuts the sample roughly in half. Obviously, this exercise is meant only to gauge how sensitive the various estimators are to a significant reduction in the size of the NCS sample and should in no way be interpreted as a recommendation about the scope and manner an NCS sample reduction would or should take.

Not surprisingly, the sample cut has a bigger impact on the estimates that rely exclusively on the NCS data, though, even here, the impact is not huge. For the OPM model, the estimates from the index sample are highly correlated with those from the full sample (0.97 for the non-Federal salaries and 0.90 for the pay gaps), but there is a tendency for the non-Federal salaries and pay gaps to be somewhat higher for the reduced sample. The standard errors do go up substantially, moreover, suggesting that, with a smaller sample, estimates will bounce around more from year to year.

For the regression method, it is striking how little the estimates change (the correlations are nearly perfect for both the non-Federal salaries and the pay gaps). In no case is the absolute value of the change in non-Federal salary estimates greater than 1 percent, and the largest change in pay gap is 1.23 percentage points. Unavoidably, the precision of the estimates is reduced by the sample cut, though it still compares favorably with the precision of the estimates from the OPM model using the full sample.

Table 5 & Table 6

[11] It may be worth noting that the greater precision for the methods that include the OES data does not appear to be attributable to the fact that these methods use a greater portion of the NCS sample – unlike the OPM model, they include part-timers, those with missing grades, and those with SOC code-grade combinations where the occupation was in the crosswalk, but the grade does not. Estimation of an OPM-style model that included the larger sample did not lead to lower standard errors.

Finally, in this third subsection, we turn to the question of whether the greater geographic coverage of the OES can lead to an expansion of the number of Pay Agent areas that can be considered. As a first exercise, summarized in Table 5, we construct new estimates of non-Federal salaries and pay gaps for each Pay Agent area after eliminating NCS data from that particular area. This exercise allows us to assess how sensitive the proposed estimators are to the absence of area-specific NCS data.[12] It is striking to note how little estimates from the regression method change when the area-specific NCS data are removed.

A second exercise is to estimate non-Federal salaries and pay gaps for nine areas that wish to be considered Pay Agent areas, but either are not sampled by the NCS or do not presently have a sample of sufficient size for Federal pay comparability. Under the current approach, these areas would be included in the "Rest of the US" and hence would be assigned RUS's pay gap. In Table 6, therefore, we compare the estimated non-Federal salaries and pay gaps to those from RUS to determine whether things would change much for these localities if the gaps were determined for them specifically.

Conclusions

The above analysis yields four main findings. First, it does seem to be feasible to use the NCS and OES in combination for the purposes of pay comparability. Second, the regression method does appear to be capable of estimating reasonable-looking pay gaps with greater precision than does the current approach. Third, it seems that one can use the OES to buffer any future cuts in the NCS sample, as the regression method is fairly robust to a large sample cut. Fourth, the regression method can be used to extend the estimation of pay gaps to areas that are not present in the NCS sample.

While we are confident in drawing these conclusions from our analysis, there are a number of additional issues we are planning on examining in the near future or have identified as possible areas of future research. Most importantly, we are planning on repeating our analysis using data from 2006 and 2008 in order to determine whether the comparisons between the regression method and the current approach are fairly robust across time. We anticipate this work will begin in late spring when the 2008 data becomes available.

In addition, while unlikely to substantially change the comparisons of the precision of the two methods, it must be noted that we assumed that the OES data were fixed when calculating the precision of the regression method estimates. While published OES estimates generally suggest rather precise estimates of mean wages, the precision of the regression method estimates should nevertheless account for variation in OES mean wages. At this point, we have no definitive plans for incorporating this factor.

Finally, further research might also yield sharper insights into the reasons why the regression method estimates somewhat higher pay gaps than does the current approach.

[12] The OPM model could not be used in this exercise as it requires area-specific data.

In analysis not shown here, we decomposed the gap between estimates from the regression method and those from the OPM model into a portion attributable to the difference between the NCS SOC code-area means and their OES-based counterparts and a portion due to differences in how pay estimates rise with grade, termed level effects. The first factor seems to be more important in explaining the difference, as the level effects are fairly similar. Understanding why the two surveys generate different SOC code-area means would be valuable, though it is not obvious that further investigation will definitely be fruitful. The NCS-OES Wage Comparisons team spent considerable time making comparisons across the two surveys, but found it difficult to come up with systematic explanations for the differences noted.

Table 1. Estimated Non-Federal Salaries for Current OPM Areas

Area	Federal Salary	Current Method	NCS with No Grade Information	OES only	OPM Model Mean	OPM Model Standard Error	OPM Model RSE	Regression Method Mean	Regression Method Standard Error	Regression Method RSE
Atlanta	$63,444	$94,031	$58,827	$59,966	$95,057	$1,471	1.5	$92,229	$983	1.1
Boston	$59,238	$92,703	$59,667	$65,669	$91,429	$1,793	2.0	$96,091	$723	0.8
Buffalo	$53,669	$75,783	$49,276	$53,663	$75,379	$1,542	2.0	$75,175	$387	0.5
Chicago	$60,985	$92,029	$57,511	$64,233	$91,067	$2,655	2.9	$96,392	$883	0.9
Cincinnati	$52,254	$69,477	$44,295	$51,066	$68,745	$1,299	1.9	$72,366	$497	0.7
Cleveland	$58,802	$82,793	$52,864	$55,978	$81,917	$1,455	1.8	$82,635	$677	0.8
Columbus	$51,760	$72,314	$49,544	$52,248	$72,495	$2,132	2.9	$71,755	$296	0.4
Dallas	$60,354	$89,963	$56,171	$59,924	$88,098	$1,082	1.2	$89,567	$791	0.9
Dayton	$56,528	$74,960	$45,708	$53,526	$74,556	$596	0.8	$78,096	$557	0.7
Denver	$63,680	$91,563	$59,358	$66,553	$91,727	$1,678	1.8	$101,173	$1,000	1.0
Detroit	$57,518	$84,390	$57,283	$64,041	$84,937	$1,069	1.3	$92,410	$657	0.7
Hartford	$55,337	$85,791	$57,914	$60,928	$86,337	$2,079	2.4	$86,028	$507	0.6
Houston	$64,558	$95,829	$58,605	$65,261	$93,711	$1,756	1.9	$99,231	$1,160	1.2
Huntsville	$70,566	$98,335	$58,448	$65,388	$98,215	$2,343	2.4	$102,254	$1,229	1.2
Indianapolis	$50,690	$68,165	$46,065	$49,370	$68,676	$1,884	2.7	$67,748	$345	0.5
Los Angeles	$55,156	$84,731	$60,492	$63,454	$85,428	$1,829	2.1	$90,335	$530	0.6
Miami	$55,490	$80,461	$51,866	$56,333	$78,376	$2,223	2.8	$80,670	$565	0.7
Milwaukee	$53,259	$73,568	$49,218	$54,481	$73,487	$1,805	2.5	$76,251	$457	0.6
Minneapolis	$53,008	$77,152	$54,483	$56,502	$77,205	$1,200	1.6	$79,052	$491	0.6
New York	$57,666	$91,638	$63,011	$66,599	$94,322	$1,726	1.8	$96,310	$721	0.7
Philadelphia	$55,923	$80,894	$54,699	$60,229	$80,574	$1,720	2.1	$85,459	$547	0.6
Phoenix	$53,989	$77,600	$52,951	$53,460	$77,420	$1,143	1.5	$75,080	$435	0.6
Pittsburgh	$55,390	$76,473	$47,700	$52,109	$76,033	$1,156	1.5	$74,414	$501	0.7
Portland	$61,544	$90,046	$56,173	$58,743	$88,905	$1,347	1.5	$87,764	$764	0.9
Rest of U.S.	$51,746	$70,025	$45,103	$48,595	$65,502	$773	1.2	$66,989	$299	0.4
Richmond	$53,542	$71,655	$49,916	$56,030	$71,210	$1,214	1.7	$77,644	$367	0.5
Sacramento	$58,197	$86,905	$59,710	$60,164	$87,156	$1,163	1.3	$87,152	$581	0.7
San Diego	$52,815	$81,216	$58,328	$59,606	$80,142	$2,078	2.6	$82,816	$374	0.5
San Francisco	$59,990	$100,160	$69,936	$73,074	$99,941	$1,472	1.5	$108,118	$926	0.9
Seattle	$56,343	$85,316	$60,148	$61,972	$87,353	$1,282	1.5	$89,432	$505	0.6
Washington, DC	$72,883	$120,578	$66,813	$76,716	$119,469	$2,740	2.3	$126,579	$1,990	1.6

Table 2. Estimated Pay Gaps for Current OPM Areas

Area	Current Method	NCS with No Grade Information	OES only	OPM Model			Regression Method		
				Mean	Standard Error	RSE	Mean	Standard Error	RSE
Atlanta	48.21	-7.28	-5.48	49.83	2.32	4.7	45.37	1.55	3.4
Boston	56.49	0.72	10.86	54.34	3.03	5.6	62.21	1.22	2.0
Buffalo	41.20	-8.19	-0.01	40.45	2.87	7.1	40.07	0.72	1.8
Chicago	50.90	-5.70	5.33	49.33	4.35	8.8	58.06	1.45	2.5
Cincinnati	32.96	-15.23	-2.27	31.56	2.49	7.9	38.49	0.95	2.5
Cleveland	40.80	-10.10	-4.80	39.31	2.47	6.3	40.53	1.15	2.8
Columbus	39.71	-4.28	0.94	40.06	4.12	10.3	38.63	0.57	1.5
Dallas	49.06	-6.93	-0.71	45.97	1.79	3.9	48.40	1.31	2.7
Dayton	32.61	-19.14	-5.31	31.89	1.05	3.3	38.16	0.99	2.6
Denver	43.78	-6.79	4.51	44.04	2.64	6.0	58.88	1.57	2.7
Detroit	46.72	-0.41	11.34	47.67	1.86	3.9	60.66	1.14	1.9
Hartford	55.03	4.66	10.10	56.02	3.76	6.7	55.46	0.92	1.7
Houston	48.44	-9.22	1.09	45.16	2.72	6.0	53.71	1.79	3.3
Huntsville	39.35	-17.17	-7.34	39.18	3.32	8.5	44.91	1.74	3.9
Indianapolis	34.47	-9.13	-2.60	35.48	3.72	10.5	33.65	0.68	2.0
Los Angeles	53.62	9.67	15.04	54.88	3.32	6.0	63.78	0.96	1.5
Miami	45.00	-6.53	1.52	41.24	4.01	9.7	45.38	1.02	2.2
Milwaukee	38.13	-7.59	2.29	37.98	3.39	8.9	43.17	0.86	2.0
Minneapolis	45.55	2.78	6.59	45.65	2.26	5.0	49.13	0.93	1.9
New York	58.90	9.26	15.49	63.56	2.99	4.7	67.01	1.25	1.9
Philadelphia	44.65	-2.19	7.70	44.08	3.07	7.0	52.81	0.98	1.9
Phoenix	43.73	-1.92	-0.98	43.40	2.12	4.9	39.07	0.81	2.1
Pittsburgh	38.06	-13.88	-5.92	37.27	2.09	5.6	34.35	0.91	2.6
Portland	46.30	-8.74	-4.55	44.45	2.19	4.9	42.60	1.24	2.9
Rest of U.S.	35.32	-12.84	-6.09	26.58	1.49	5.6	29.46	0.58	2.0
Richmond	33.83	-6.77	4.65	33.00	2.27	6.9	45.02	0.68	1.5
Sacramento	49.33	2.60	3.38	49.76	2.00	4.0	49.75	1.00	2.0
San Diego	53.78	10.44	12.86	51.74	3.93	7.6	56.80	0.71	1.2
San Francisco	66.96	16.58	21.81	66.60	2.45	3.7	80.23	1.54	1.9
Seattle	51.42	6.75	9.99	55.04	2.28	4.1	58.73	0.89	1.5
Washington, DC	65.44	-8.33	5.26	63.92	3.76	5.9	73.68	2.73	3.7

Table 3. The Effect of an NCS Sample Cut on Non-Federal Salary Estimates

	OPM Model					Regression Method				
	Full Sample	Index Sample			% Change (Index - Full)	Full Sample	Index Sample			% Change (Index - Full)
Area	Mean	Mean	Std Error	RSE		Mean	Mean	Std Error	RSE	
Atlanta	$95,057	$97,107	$3,434	3.5	2.16	$92,229	$92,869	$2,304	2.5	0.69
Boston	$91,429	$94,901	$3,259	3.4	3.80	$96,091	$96,210	$1,771	1.8	0.12
Buffalo	$75,379	$76,488	$2,497	3.3	1.47	$75,175	$74,986	$1,051	1.4	-0.25
Chicago	$91,067	$92,495	$2,279	2.5	1.57	$96,392	$96,699	$2,129	2.2	0.32
Cincinnati	$68,745	$69,615	$2,454	3.5	1.27	$72,366	$72,402	$1,230	1.7	0.05
Cleveland	$81,917	$80,210	$2,322	2.9	-2.08	$82,635	$82,662	$1,682	2.0	0.03
Columbus	$72,495	$70,012	$2,797	4.0	-3.42	$71,755	$71,338	$844	1.2	-0.58
Dallas	$88,098	$87,457	$3,788	4.3	-0.73	$89,567	$89,842	$1,875	2.1	0.31
Dayton	$74,556	$77,321	$1,976	2.6	3.71	$78,096	$77,978	$1,264	1.6	-0.15
Denver	$91,727	$96,706	$3,035	3.1	5.43	$101,173	$101,659	$2,347	2.3	0.48
Detroit	$84,937	$84,389	$1,331	1.6	-0.64	$92,410	$92,343	$1,635	1.8	-0.07
Hartford	$86,337	$85,867	$2,946	3.4	-0.54	$86,028	$85,880	$1,197	1.4	-0.17
Houston	$93,711	$92,661	$2,934	3.2	-1.12	$99,231	$99,493	$2,704	2.7	0.26
Huntsville	$98,215					$102,254	$102,961	$2,904	2.8	0.69
Indianapolis	$68,676	$69,826	$4,909	7.0	1.67	$67,748	$67,447	$831	1.2	-0.44
Los Angeles	$85,428	$85,093	$2,253	2.6	-0.39	$90,335	$90,143	$1,388	1.5	-0.21
Miami	$78,376	$84,026	$3,389	4.0	7.21	$80,670	$80,748	$1,434	1.8	0.10
Milwaukee	$73,487	$77,759	$3,304	4.2	5.81	$76,251	$76,144	$1,185	1.6	-0.14
Minneapolis	$77,205	$77,209	$2,247	2.9	0.00	$79,052	$78,935	$1,202	1.5	-0.15
New York	$94,322	$96,581	$2,065	2.1	2.40	$96,310	$96,460	$1,758	1.8	0.16
Philadelphia	$80,574	$82,235	$959	1.2	2.06	$85,459	$85,314	$1,332	1.6	-0.17
Phoenix	$77,420	$80,055	$2,301	2.9	3.40	$75,080	$74,901	$1,132	1.5	-0.24
Pittsburgh	$76,033	$79,231	$4,081	5.2	4.21	$74,414	$74,405	$1,212	1.6	-0.01
Portland	$88,905	$93,872	$3,534	3.8	5.59	$87,764	$88,144	$1,798	2.0	0.43
Rest of U.S.	$65,502	$65,579	$948	1.4	0.12	$66,989	$66,703	$815	1.2	-0.43
Richmond	$71,210	$78,758	$2,101	2.7	10.60	$77,644	$77,326	$970	1.3	-0.41
Sacramento	$87,156	$92,278	$2,263	2.5	5.88	$87,152	$87,128	$1,449	1.7	-0.03
San Diego	$80,142	$84,721	$1,854	2.2	5.71	$82,816	$82,454	$1,077	1.3	-0.44
San Francisco	$99,941	$102,023	$2,205	2.2	2.08	$108,118	$108,225	$2,285	2.1	0.10
Seattle	$87,353	$87,335	$2,761	3.2	-0.02	$89,432	$89,215	$1,333	1.5	-0.24
Washington, DC	$119,469	$122,602	$3,698	3.0	2.62	$126,579	$127,479	$4,485	3.5	0.71

Table 4. The Effect of an NCS Sample Cut on Federal Pay Gaps

Area	OPM Model					Regression Method				
	Full Sample	Index Sample			Change (Index - Full)	Full Sample	Index Sample			Change (Index - Full)
	Mean	Mean	Std Error	RSE		Mean	Mean	Std Error	RSE	
Atlanta	49.83	53.06	5.41	10.2	3.23	45.37	46.38	3.63	7.8	1.01
Boston	54.34	60.20	5.50	9.1	5.86	62.21	62.41	2.98	4.8	0.20
Buffalo	40.45	42.52	4.65	10.9	2.07	40.07	39.72	1.96	4.9	-0.35
Chicago	49.33	51.67	3.74	7.2	2.34	58.06	58.56	3.49	6.0	0.50
Cincinnati	31.56	33.22	4.70	14.1	1.66	38.49	38.56	2.36	6.1	0.07
Cleveland	39.31	36.41	3.95	10.8	-2.90	40.53	40.58	2.86	7.0	0.05
Columbus	40.06	35.26	5.40	15.3	-4.80	38.63	37.82	1.63	4.3	-0.81
Dallas	45.97	44.91	6.28	14.0	-1.06	48.40	48.86	3.11	6.4	0.46
Dayton	31.89	36.78	3.50	9.5	4.89	38.16	37.95	2.24	5.9	-0.21
Denver	44.04	51.86	4.77	9.2	7.82	58.88	59.64	3.69	6.2	0.76
Detroit	47.67	46.72	2.31	5.0	-0.95	60.66	60.55	2.84	4.7	-0.11
Hartford	56.02	55.17	5.32	9.7	-0.85	55.46	55.19	2.17	3.9	-0.27
Houston	45.16	43.53	4.55	10.4	-1.63	53.71	54.11	4.19	7.7	0.40
Huntsville	39.18					44.91	45.91	4.12	9.0	1.00
Indianapolis	35.48	37.75	9.68	25.7	2.27	33.65	33.06	1.64	5.0	-0.59
Los Angeles	54.88	54.28	4.08	7.5	-0.61	63.78	63.43	2.52	4.0	-0.35
Miami	41.24	51.42	6.11	11.9	10.18	45.38	45.52	2.58	5.7	0.14
Milwaukee	37.98	46.00	6.20	13.5	8.02	43.17	42.97	2.22	5.2	-0.20
Minneapolis	45.65	45.65	4.24	9.3	0.01	49.13	48.91	2.27	4.6	-0.22
New York	63.56	67.48	3.58	5.3	3.92	67.01	67.27	3.05	4.5	0.26
Philadelphia	44.08	47.05	1.72	3.6	2.97	52.81	52.56	2.39	4.5	-0.25
Phoenix	43.40	48.28	4.26	8.8	4.88	39.07	38.73	2.10	5.4	-0.34
Pittsburgh	37.27	43.04	7.37	17.1	5.77	34.35	34.33	2.19	6.4	-0.02
Portland	44.45	52.52	5.74	10.9	8.07	42.60	43.22	2.92	6.8	0.62
Rest of U.S.	26.58	26.73	1.83	6.9	0.15	29.46	28.91	1.57	5.4	-0.55
Richmond	33.00	47.10	3.92	8.3	14.10	45.02	44.42	1.81	4.1	-0.60
Sacramento	49.76	58.56	3.89	6.6	8.80	49.75	49.71	2.49	5.0	-0.04
San Diego	51.74	60.41	3.51	5.8	8.67	56.80	56.12	2.04	3.6	-0.68
San Francisco	66.60	70.07	3.68	5.2	3.47	80.23	80.40	3.81	4.7	0.17
Seattle	55.04	55.01	4.90	8.9	-0.03	58.73	58.34	2.36	4.1	-0.39
Washington DC	63.92	68.22	5.07	7.4	4.30	73.68	74.91	6.15	8.2	1.23

Table 5. The Effect of Excluding Area-Specific NCS Data on Non-Federal Salary and Pay Gap Estimates

Area	Regression Method					
	Non-Federal Salary			Pay Gap		
	Full Sample	With Area Excluded from NCS Sample	Percent Change	Full Sample	With Area Excluded from NCS Sample	Change
Atlanta	$92,229	$92,403	0.19	45.37	45.65	0.28
Boston	$96,091	$95,834	-0.27	62.21	61.78	-0.43
Buffalo	$75,175	$75,182	0.01	40.07	40.08	0.01
Chicago	$96,392	$96,227	-0.17	58.06	57.79	-0.27
Cincinnati	$72,366	$72,342	-0.03	38.49	38.44	-0.05
Cleveland	$82,635	$82,636	0.00	40.53	40.53	0.00
Columbus	$71,755	$71,722	-0.05	38.63	38.57	-0.06
Dallas	$89,567	$89,275	-0.33	48.40	47.92	-0.48
Dayton	$78,096	$78,094	0.00	38.16	38.15	-0.01
Denver	$101,173	$101,153	-0.02	58.88	58.85	-0.03
Detroit	$92,410	$92,616	0.22	60.66	61.02	0.36
Hartford	$86,028	$86,051	0.03	55.46	55.50	0.04
Houston	$99,231	$99,326	0.10	53.71	53.86	0.15
Huntsville	$102,254	$102,368	0.11	44.91	45.07	0.16
Indianapolis	$67,748	$67,697	-0.08	33.65	33.55	-0.10
Los Angeles	$90,335	$90,760	0.47	63.78	64.55	0.77
Miami	$80,670	$80,654	-0.02	45.38	45.35	-0.03
Milwaukee	$76,251	$76,255	0.00	43.17	43.18	0.01
Minneapolis	$79,052	$79,120	0.09	49.13	49.26	0.13
New York	$96,310	$96,565	0.26	67.01	67.46	0.45
Philadelphia	$85,459	$85,407	-0.06	52.81	52.72	-0.09
Phoenix	$75,080	$74,998	-0.11	39.07	38.91	-0.16
Pittsburgh	$74,414	$74,392	-0.03	34.35	34.31	-0.04
Portland	$87,764	$87,751	-0.01	42.60	42.58	-0.02
Richmond	$77,644	$77,580	-0.08	45.02	44.90	-0.12
Sacramento	$87,152	$87,145	-0.01	49.75	49.74	-0.01
San Diego	$82,816	$82,839	0.03	56.80	56.85	0.05
San Francisco	$108,118	$108,482	0.34	80.23	80.83	0.60
Seattle	$89,432	$89,548	0.13	58.73	58.93	0.20
Washington, DC	$126,579	$127,274	0.55	73.68	74.63	0.95

Table 6. Estimated Non-Federal Salaries and Pay Gaps for Potential OPM Areas

				Regression Method					
				Non-Federal Salary			Pay Gap		
Area	Federal Salary	Current RUS Non-Federal Salary Estimate	Current RUS Pay Gap Estimate	Current RUS Estimate	Area-Specific Estimate	% Change	Current RUS	Area-Specific Estimate	Change
Albany	$57,252	$70,025	35.32	$66,989	$80,331	19.92	29.46	40.31	10.85
Albuquerque	$53,977	$70,025	35.32	$66,989	$75,105	12.12	29.46	39.14	9.68
Bakersfield	$59,297	$70,025	35.32	$66,989	$94,993	41.80	29.46	60.20	30.74
Beaumont	$50,303	$70,025	35.32	$66,989	$59,427	-11.29	29.46	18.14	-11.32
Harrisburg	$54,883	$70,025	35.32	$66,989	$75,727	13.04	29.46	37.98	8.52
Lansing	$57,882	$70,025	35.32	$66,989	$83,793	25.08	29.46	44.77	15.31
New Orleans	$58,290	$70,025	35.32	$66,989	$76,563	14.29	29.46	31.35	1.89
Portland, ME	$60,362	$70,025	35.32	$66,989	$82,341	22.92	29.46	36.41	6.95
Wilmington	$51,394	$70,025	35.32	$66,989	$61,599	-8.05	29.46	19.86	-9.60

Figure 1. Ranks of Area Pay Gaps