Understanding the Gender Pay Gap in the Federal Workforce over the Past 20 Years

Ben Bolitzer, Senior Economist, bolitzerb@gao.gov
441 G. St., NW.
Washington, DC 20548
Phone: 202-512-5427
Erin Godtland, Senior Economist, godtlande@gao.gov
301 Howard St., Suite 1200
San Francisco, CA 94105
Phone: 415-904-2271
The authors are employed by the U.S. Government Accountability Office.*

This paper examines the extent to which the pay gap between men and women in the federal workforce changed over the past 20 years and the factors that accounted for the gap. Using the most recent data available for a representative sample of federal employees, we employ multivariate regression and decomposition methods to analyze the contributions of specific factors to the pay gap. We find that the raw gap—before controlling for differences between men and women in factors that affect pay—declined dramatically over the past 20 years. The unexplained pay gap remained relatively level over this period, but varied with different specifications of the model

.

* The authors are researchers at the U.S. Government Accountability Office (GAO). Any errors are the authors' and should not be attributed to the GAO. We are indebted to several colleagues at the GAO that provided helpful comments and suggestions on our analysis and/or previous drafts of this paper including Michele Grgich, Douglas Sloane, Monique Williams, Dan Concepcion, Rhiannon Patterson, Ron Fecso, Anne-Marie Lasowski, Andrew Sherrill, Shana Wallace, and Greg Wilmoth. The results of this paper were first published in the U.S. GAO (2009).

Introduction

Over the past few years and during the recent presidential campaign much attention has been devoted to the persistent pay gap between men and women in the U.S. workforce. The Lilly Ledbetter Fair Pay Act—a law intended to bolster protections against pay discrimination for women and other workers—was one of the first pieces of legislation enacted by the Obama Administration. Additional legislation addressing the gender pay gap, such as the Paycheck Fairness Act, is currently under debate in the Congress. To inform this debate, it is essential to have an understanding of the pay gap and its underlying sources.

This article examines the evolution of the gender pay gap and the factors that contribute to the gap among federal employees from 1988 to 2007. Using the most recent data on federal workers from the Central Personnel Data File (CPDF), a comprehensive administrative database of federal employees maintained by the Office of Personnel Management, we employ a decomposition approach to determine the extent to which the pay gap between men and women over this period has changed and what proportion can be explained by differences between men and women in factors that affect pay such as their occupations, education levels and years of federal work experience.

While the focus of the empirical methods in this article is the federal workforce, the analysis can inform the debate over the general workforce as well. The federal government is currently the largest single employer in the United States. Further, the federal government is arguably a more progressive employer than generally found in the private sector, such that any unexplained pay gap in the federal sector might be considered conservative relative to the private sector. In

addition, most recent analyses of the gender gap among the general workforce, such as Blau (2004), have relied on surveys like the current population survey. By using an administrative data source, this article may have a more precise measure of earnings and experience than can come from a survey.

The article is organized as follows: We first briefly review the literature on pay disparities in the federal workforce. We then describe our econometric approach and our data. Finally we present our empirical findings and some concluding remarks.

Prior Evidence of a Gender Pay Gap in the Federal Workforce

Over the past decades, there has been considerable research that has applied regression approaches to gender and racial pay disparities within the federal sector. For example, using a sample from the 1979 CPDF, Borjas (1983) investigated gender disparities in the federal workforce and found that white women earned 21 percent less and black women 27 percent less than white men, after controlling for differences in education and experience levels.[1] Similarly, using a sample from the 1986 CPDF, Lewis (1988) found that White non-Hispanic and black women earned 20 percent less and 28 percent less after controlling for differences in education and experience levels.[2] In a later article, Lewis (1998) presented two models with varying

[1] George J. Borjas. 1983. "The Measurement of Race and Gender Wage Differentials: Evidence from the Federal Sector" Industrial and Labor Relations Review, Vol. 37 (1): 79-91.

[2] Gregory Lewis. 1988. "Progress toward Racial and Sexual Equality in the Federal Civil Service?" Public Administration Review, Vol. 48 (3): 700-707.

numbers of controls to investigate unexplained differences in wages by race and gender using a sample from the CPDF from 1995. The first model, which included education, experience, and age, found that white women earned 15 percent less and black women earned 21 percent less than white men respectively. A second model, which also included agency, region of the country, and veterans and disability status, found that white women earned 14 percent less and black women 24 percent less than white men respectively.[3] More recently, Lewis (2009) and found in a sample of college graduates from the CPDF, that an additional 4 percent of the pay gap was explained by adding field of study. However, in a similar analysis, Lewis et al. (2009) found that women's migration into fields of study that were previously dominated by men did not play a large role in reducing the pay gap between men and women over time in the federal workforce because the average wage in fields experiencing an influx of women tended to go down.

Other research has explored the impact of occupational segregation on the gender pay gap in the federal workforce. For example, Lewis (1996) measures the effect of the percentage of male workers in a particular occupation on the difference between the salaries of men and women.

---

[3] Gregory Lewis. 1998 "Continuing Progress Toward Racial and Gender Pay Equality in the Federal Sector" Review of Public Personnel Administration 18: 23-40.

The results indicate that introduction of the variable reduced the disparity between men and women from 18 percent to less than 6 percent in 1992.[4]

Empirical Approach

We build on this literature by estimating several specifications of a wage model with data from 1988, 1998, and 2007, using two different techniques. Both techniques involve multivariate regression, and control for many factors that might affect pay, such as level of education or occupation.

The first technique conducts regression analysis on a data set which includes men and women. In this analysis, we use a variable for gender to measure the average difference between men and women's salaries. By adding additional variables to the regression, we control for other characteristics of men and women to determine the extent to which the difference is (or is not) explained by the addition of those variables.

The second technique, called a decomposition, analyzes men's and women's salaries in separate regressions. This method provides an additional tool for determining which attributes are the key explanations of the differences between men and women's salaries, and also what percentage of men and women's salary remains unexplained by the attributes characteristics measured in our data.

[4] Gregory Lewis. 1996 "Gender Integration of Occupations in the Federal Civil Service: Extent and Effects on Male Female Earnings" Industrial and Labor Relations Review 49: 472-483.

(a) Regression Approach:

In order to determine the extent to which gender differences persist when characteristics of men and women are taken into account, we performed a multivariate regression analysis for three years of data, 1988, 1998 and 2007. Specifically, we attempted to explain the differences by predicting the logarithm of annual adjusted pay on characteristics of federal workers. Because of concerns that some of the attributes of the individual could be endogenous, or due, to discrimination, we used two different types of characteristics—individual and position—as shown in the following equation:

(1) $$\text{Ln(annual pay)} = \alpha + \beta(\text{female}) + \delta^*(\text{set of characteristics of the individual}) + \gamma^*(\text{set of characteristics of the position})$$

The standard interpretation of β, the coefficient on female, is that it represents the average percent difference in earnings between men and women, after controlling for the other variables in the model.[5] Variables that control for characteristics of the individual included years of federal experience, age, race and ethnicity, educational degree attained, disability status, state and veteran status. Variables that control for characteristics of the position included occupation, agency, work schedule, and union status.

(b) Decomposition Approach

[5] The "extensions" section describes additional regressions were run with different sets of controls or on sub-samples of the population.

One possible explanation for the gap could be that women have different levels of important attributes, like years of experience, than men. Alternatively, women could have the same level of attributes, but women's attributes could be treated differently. For example, the return to an additional year of experience, or a master's degree, might be different for a woman than a man. In order to determine whether the difference between men's and women's pay is a function of men and women having different levels of characteristics, or different returns to those characteristics, we employed the "Oaxaca decomposition," a widely used tool for examining disparities between two groups.[6] To apply the "Oaxaca decomposition", we followed the following steps:

1. First, we estimated two versions of equation (1), one on the sample of women and one on the sample of men. This provided us with two sets of regression coefficients, one for men and one for women.
2. Then, we applied the regression coefficients for men to the average values of characteristics for men. This gave us the average wages of men. We repeated this analysis for women, producing the average wages for women.
3. We then applied the coefficients for men to the average values for the characteristics for women. This gave us a new predicted wage – the predicted wage for women if they had the same returns to characteristics as men.

With these three values, we were able to decompose the total difference between the average of male and female wages into two parts:

[6] For details on this technique see "Male-Female Wage Differentials in Urban Labor Markets," by Ronald Oaxaca, in International Economic Review, Volume 14, Issue 3 (Oct. 1973), 693-709.

Equation 2:

	(Average female wages) – (Male returns with female characteristics)	=		"Unexplained" or due to parameter difference between women and men
+	(Male Returns with female characteristics) – (Average male wages)	=	+	"Explained" or due to characteristics difference between men and women
	(Average female wages) – (Average male wages)	=		Total

Data Source and Descriptive Statistics

The data for the analysis comes from the status file of the Central Personnel Data File (CPDF). This dataset is produced by the Office of Personnel Management as a central source of information regarding the federal workforce. The CPDF contain information on most federal employees who were present in the federal workforce in September of 1988, 1998, and 2007.[7] We analyzed a random sample of 20 percent of the workers in the CPDF data for these three

[7] Specifically, CPDF coverage of the executive branch currently includes all agencies except the Board of Governors of the Federal Reserve, the Central Intelligence Agency, the Defense Intelligence Agency, Foreign Service personnel at the State Department, the National Geospatial-Intelligence Agency, the National Security Agency, the Office of the Director of National Intelligence, the Office of the Vice President, the Postal Rate Commission, the Tennessee Valley Authority, the U.S. Postal Service, and the White House Office. Also excluded are the Public Health Service's Commissioned Officer Corps, non-appropriated fund employees, and foreign nationals overseas. CPDF coverage of the legislative branch is limited to the Government Printing Office, the U.S. Tax Court, and selected commissions.

years.[8] The CPDF data contain information on the federal employee's adjusted basic pay, agency, age, education level, disability status, occupation, race or national origin, gender, veteran's preference and status, and work schedule.

Table 1 shows descriptive statistics for men and women for the three years we used in our analysis. As the table shows, there has been a significant narrowing in the pay gap and in the differences between men and women in many characteristics over this period.

The gap in annual salaries between men and women (as measured with the log of adjusted basic pay) narrowed markedly from 1988 to 2007.[9] The difference in the average log earnings of men and women was about 0.33 in 1988, 0.21 in 1998, and 0.12 in 2007.[10]

[8] In comparison compare Lewis (2009) used a 1 percent sample of the CPDF.

[9] Adjusted basic pay takes into account various differences in pay based on locality and special rates and existing pay caps. It reflects the amount an individual would have earned had he or she worked a complete year. It does not reflect their actual earnings, which are not available in the CPDF data. We deflated the salary using the consumer price index.

Differences between men and women in their years of federal work experience, education levels, and occupations also diminished over this period.[11] In 1988, men in the federal workforce had an average of three more years of work experience then their female counterparts. That difference narrowed to two years in 1998, and by 2007 there was no appreciable difference.[12] Similarly, in 1988, almost twice as many men than women in the federal workforce had bachelors, masters, professional, or doctoral degrees (40 percent versus 23 percent) compared to 2007, when the difference was less than 10 percentage points (46 versus 40 percent). Also, over the past two decades, male and female federal workers have also worked in increasingly similar

---

[10] The standard interpretation of the log difference is that it is equivalent to the percent difference; however at larger values this value will differ somewhat from the precise percent difference. To transform the coefficient to more exactly equal the percent difference (and to present it in dollar terms), we applied the following formula: exp(difference in logarithms)-1. The results of this computation yielded an estimated percent difference of negative 28 percent in 1988, negative 19 percent in 1998 and negative 11 percent in 2007.

[11] We defined occupation using occupational category in the CPDF, which groups occupations into six categories: Professional, Administrative, Technical, Clerical, Other White-collar and Blue-collar. For the purposes of our analysis, we called this categorical variable PATCOB.

[12] We measured federal experience by the months between the service computation date and the date the data were drawn (September of each year).

occupations. Much of this trend is due to the diminishing clerical sector in the federal workforce. In 1988, about 38 percent of women in the federal workforce were in a clerical occupation. By 2007, that number was 13 percent. A similar trend occurred with men in the "Blue-Collar" sector. In 1988 almost 28 percent of men in the federal workforce were "Blue-Collar" workers, but by 2007 that number was 17 percent.[13] Interestingly, differences between men's and women's work schedules did not change significantly over the study period, with the percentage of women working full time about 4 to 5 percentage points lower than men over the entire study period.[14]

From 1988 to 2007, the demographic composition of the federal workforce changed. First, the federal workforce became older. The average age for male and female federal workers in 1988 was 43 and 40 respectively; whereas, in 2007, the average ages were 47 and 46, respectively. There was also a decline in the proportion of white workers and an increase in the proportion of

[13]An index of dissimilarity is an alternate way to demonstrate the convergence of the occupational structure. The index of dissimilarity is defined as the fraction of either men or women that would have to switch occupations to make the distributions identical. The range of values are 1 (meaning that the 100 % of men or women would have to switch) to 0 (meaning that the distributions are identical). Using PATCOB, the dissimilarity index fell from 40% in 1988 to 30% in 1998 to almost 20% in 2007, indicating that the distributions are much closer today.

[14] In the CPDF, employees are classified by whether they worked full-time, part-time or held a flexible schedule (such as seasonal, intermittent, on-call, etc.)

Hispanic and Asian/Pacific Islanders.[15] Finally, the proportion of women in the federal workforce increased from 42 percent to 44 percent.

Although the CPDF does not contain variables for marital status and number of children, it contains a proxy. Specifically, the CPDF has a variable to measure whether an individual registered for health insurance for their family or themselves or declined health insurance coverage. Declined coverage may imply that the employee receives coverage through a spouse. In each year of our analysis, men are much more likely than women to participate in a family plan. In 1988, women were more than twice as likely to have declined coverage, although this gap narrowed substantially by 2007.

Over the period, there was little change in the rates of disability among male and female federal workers. Specifically in 1988, 93 and 95 percent of men and women were classified as having no disability respectively, while in 2007, 94 and 95 percent of men and women had no disability.[16]

[15] We measured race and ethnicity using the CPDF definitions. These definitions do not allow for multiple races. Unlike many data sets, they do not record Hispanic status distinctly from race.

[16] We defined disability by whether the employee did or did not have a CPDF code for a disability condition and whether that condition indicated a targeted disability as defined by EEOC's Management Directive 715.

In addition to the variable listed in the table, we included two other sets of variables in the econometric analysis. Specifically, we controlled for whether an employee (1) was a veteran, and (2) qualified for a veteran's preference. We also controlled for an employee's geographic location by including dummy variables representing the state in which the individual was employed.

Table 1: Descriptive Statistics for selected CPDF variables used in our cross-sectional analysis

	1988			1998			2007		
	Men	**Women**	**Difference between men and women**	**Men**	**Women**	**Difference between men and women**	**Men**	**Women**	**Difference between men and women**
Log of Annual Adjusted Salary	10.847	10.52	-0.33	10.957	10.745	-0.21	11.059	10.938	-0.12
Federal Experience	14	10	-4	16	14	-2	15	15	0
Education									
Less than High School	4%	3%	-1%	2%	2%	0%	1%	1%	0%
High School Diploma	27%	35%	8%	25%	32%	7%	28%	28%	0%
Trade Degree	5%	8%	3%	3%	5%	2%	2%	4%	2%
Some College	24%	30%	6%	23%	29%	6%	19%	24%	5%
Bachelor Degree	26%	17%	-9%	28%	21%	-7%	27%	24%	-3%
Masters Degree	8%	4%	-4%	10%	7%	-3%	12%	11%	-1%
Professional Degree	4%	2%	-2%	5%	3%	-2%	3%	3%	0%
Doctorate Degree	2%	1%	-1%	3%	1%	-2%	3%	2%	-1%
Other Education	1%	1%	0%	1%	1%	0%	3%	3%	0%
Occupation									

	1988			1998			2007		
	Men	**Women**	**Difference between men and women**	**Men**	**Women**	**Difference between men and women**	**Men**	**Women**	**Difference between men and women**
Administrative	26%	21%	-5%	31%	28%	-3%	35%	35%	0%
Blue-collar	28%	5%	-23%	22%	3%	-19%	17%	3%	-14%
Clerical	5%	38%	33%	4%	20%	16%	5%	13%	8%
Other White Collar	3%	0%	-3%	4%	0%	-4%	5%	1%	-4%
Professional	23%	14%	-9%	27%	21%	-6%	24%	24%	0%
Technical	15%	22%	7%	14%	26%	12%	14%	25%	11%
Work Schedule									
Full Time	94%	89%	-5%	93%	89%	-4%	94%	90%	-4%
Part Time	2%	6%	4%	2%	5%	3%	2%	5%	3%
Another Type	4%	6%	2%	5%	7%	2%	4%	5%	1%
Age (years)	43	40	-3	46	44	-2	47	46	-1
Race/Ethnicity									
African American	12%	23%	11%	11%	23%	12%	12%	24%	12%
Asian Pacific Islander	4%	3%	-1%	5%	4%	-1%	5%	6%	1%
Hispanic	6%	5%	-1%	7%	6%	-1%	8%	7%	-1%
Native American	2%	2%	0%	2%	3%	1%	2%	3%	1%
White	78%	67%	-11%	76%	64%	-12%	73%	61%	-12%

	1988			1998			2007		
	Men	**Women**	**Difference between men and women**	**Men**	**Women**	**Difference between men and women**	**Men**	**Women**	**Difference between men and women**
Percentage Female	42%			44%			44%		
Health Plan									
Family Plan	60%	32%	-28%	60%	36%	-24%	52%	36%	-16%
Self Plan	19%	33%	14%	22%	35%	13%	24%	37%	13%
Declined Coverage	11%	23%	12%	10%	20%	10%	16%	19%	3%
Pending	3%	5%	2%	2%	2%	0%	3%	3%	0%
Not eligible	7%	8%	1%	6%	6%	0%	5%	5%	0%
Disability Status									
None	93%	95%	2%	93%	95%	2%	94%	95%	1%
Disabled not Targeted	6%	4%	-2%	6%	4%	-2%	5%	4%	-1%
Disabled	1%	1%	0%	1%	1%	0%	1%	1%	0%
Number of Observations	241611	175776		199153	158460		205767	162822	

Source: GAO Analysis of CPDF data

Empirical results

(a) Regression approach results

Table 2 presents the coefficients and standard errors for the regression results from estimating equation 1. As described above, the coefficient on female can be interpreted as the percent difference between women's and men's annual salary, after accounting for all of the measurable characteristics of men and women that we controlled for in the model. Additionally, table 2 presents values and standard errors of the coefficients associated with all of the other characteristics in the model.

As the table shows, the percent difference between women's and men's salary, controlling for the factors listed in the table, has fallen over the past 20 years. A negative value indicates that women's salary was less than men's. Specifically, the coefficient on female changed from approximately negative 10.9 percent in 1988, to negative 8.8 in 1998 and negative 8.3 in 2007.

Many of the other parameters associated with the control variables are in the expected direction. Higher education levels are associated with higher levels of salary. For example, after controlling for the other factors in the model, in 1988 a federal worker with a BA had a salary that was 18 percent higher than the salary of a person who did not complete high school. A person with an MA, in 1998, had a salary that was 25 percent higher than the salary of a person that did not complete high school. Salary increases at higher levels of federal experience and age, but the marginal effect of an additional year decreases as the years increase (as indicated by the negative sign of the estimate for the squared terms for age and experience). Clerical workers tend to be paid less, even after the other controls are introduced. For example, in 1988, the salary of a clerical worker was 15.5 percent lower than the salary of a technical worker, after controlling for the other factors in the model.

Similar to gender, there are disparities by racial and ethnic groups, as well as by disability status. For example, in 2007, the salary for an African American employee was 7.4 percent lower than the salary of a white person, after controlling for the other factors in the model.

Table 2: Regression Results

	1988		1998		2007	
	Estimate	Standard Error	Estimate	Standard Error	Estimate	Standard Error
Female	-.109	.001	-.088	.001	-.083	.001
Experience and Age						
Age	.018	.001	.041	.001	.050	.001
Age Squared	-.0002	.00002	-.0007	.00002	-.0008	.00002
Age Cubed	6.46E-7	1.53E-7	3.64E-7	1.73E-7	4.31E-6	1.59E-7
Federal Experience	.035	.0002	.031	.0003	.029	.0003
Federal Exp. Squared	-.001	.00002	-.001	.00002	.001	1.6E-6
Federal Exp. Cubed	.00001	2.60E-7	.00001	2.82E-7	.00001	2.62E-7
Race/Ethnicity (white is omitted)						
African American	-.079	.001	-.074	.001	-.074	.001
Asian Pacific Islander	-.015	.002	-.022	.002	-.005	.002
Hispanic	-.045	.001	-.042	.001	-.028	.001
Native American	-.033	.002	-.042	.002	-.055	.003
Other	-.043	.014	-.057	.016	-.037	.007
Education (less than high school is omitted)						
High school	.078	.002	.074	.003	.076	.003
Trade Degree	.112	.002	.112	.003	.112	.004
Some College	.110	.002	.112	.003	.114	.004
Bachelor Degree	.182	.002	.193	.003	.182	.004
Masters Degree	.258	.002	.272	.003	.247	.004
Professional Degree	.456	.003	.442	.003	.561	.004
Doctorate Degree	.411	.003	.418	.004	.398	.004
Other Education	.035	.004	.058	.004	.091	.004
Occupation (Technical is omitted)						
Administrative	.260	.001	.318	.001	.363	.001
Blue-collar	.095	.001	.053	.001	.036	.001
Clerical	-.156	.001	-.163	.001	-.204	.002
Other White Collar	-.124	.002	.006	.002	.097	.002
Professional	.370	.001	.397	.001	.432	.001
Work Schedule (part time is omitted)						
Full Time	.040	.002	.023	.002	.040	.002
Another Type	-.097	.002	-.171	.002	-.085	.003
Disability Status (targeted disability is omitted)						
None	.085	.003	.102	.003	.090	.004
Disabled not targeted	.062	.003	.076	.003	.061	.004
Observations	417,387		357,613		368,589	
R –Square	79%		78%		77%	

Source: GAO analysis of CPDF data. In addition to the variables listed above, the regression included a measure of state, larger agencies, and veteran status

Analysis of alternative specifications of the model

Because certain variables may be specified differently and our results could change with alternate variable specifications, we conducted additional cross-sectional analyses to better understand the degree to which different specifications of key variables might impact our results. For example, we tested several different specifications of the occupation variable. Specifically, we ran our model with "job family level"—a categorical variable that had about 50 different occupation categories—and then with "job series"—another categorical variable with more than 700 occupation categories.[17] We did so because the fact that men and women are hired into or remain in (albeit decreasingly) different occupations may itself reflect some level of discrimination associated with hiring, promotion, or other employer practices.[18] As such, using a more precise measure of occupation (as compared with the PATCOB categories) might hide the contribution of any such discrimination to the pay gap, and thereby understate the unexplained gap. To shed light on this, we estimated our model with no control for occupation, which would represent an upper-bound on the unexplained pay gap. We found that, with no control for

[17] Job Family Level" was constructed by combining PATCOB with the "occupational group" variable in the CPDF data, and collapsing blue-collar occupations into a single category. An occupational group is a set of occupations in a related field such as engineering or health care. In addition, those occupations that individually represented 0.35 percent of the population were combined into an "other" category. The number of categories included in a regression depended on whether that category had any individuals in a particular year.

[18] For discussions of sex discrimination in hiring, see Claudia Goldin and Cecilia Rouse, "Orchestrating Impartiality: The Impact of 'Blind' Auditions on Female Musicians," *American Economic Review*, vol. 90, no. 4 (2000); and David M. Neumark, "Sex Discrimination in Restaurant Hiring: An Audit Study," *Quarterly Journal of Economics*, vol. 111, no.3 (1996).

occupation, the unexplained pay gap was 19 percent in 1988, 13 percent in 1998, and 11 percent in 2007. In contrast, with the most disaggregated control for occupation, i.e., "job series," we found that the unexplained gap was 7 percent in 1988, 6 percent in 1998, and 5 percent in 2007. The results of the different specifications are described in table 3 below.

We also tested whether additional information on education and geography reduced the pay gap. Specifically, we included in the model a variable for an individual's educational major, which was only available for our 2007 cross-sectional analysis. We added separate dummy variables for 40 broad educational majors interacted with the level of education. For that year, we found that our educational major variable reduced the unexplained gap by less than one percent. We also included a more detailed measure of geography—the county in which an employee works. We found that the more specific control for geography had no impact on the pay gap.

Finally, we ran an alternative regression model with a rough proxy variable to reflect factors that may legitimately influence wages, but for which we lack data. As noted prior, the CPDF data do not contain information on marital status and number of children, variables that are commonly regarded as proxies for personal obligations and have been included in wage models in some literature.[19] However, due to the potential endogeneity of these variables, other authors

[19] See, for example, June O'Neill, "The Gender Gap in Wages, circa 2000," *The American Economic Review.* Vol. 93, No.2, Papers and Proceedings of the One Hundred Fifteenth Annual Meeting of the American Economic Association, Washington, DC, January 3-5, 2003. (May, 2003), pp.309-314, and Audrey Light and Manuelita Ureta, "Early-Career Work Experience and Gender Way Differentials," *Journal of Labor Economics,* Vol. 13, No. 1. (Jan., 1995), pp.121-154.

sometime do not include controls for marital status and family size in analyses of the pay gap.[20] To address this potential shortcoming, we analyzed a variable in the CPDF that indicates whether a federal employee is enrolled in a federal health benefit plan for single or family benefits. The health plan variable is a rough proxy of whether an individual has a family because individuals may receive family health benefits through a spouse. Including the health care variable in the model reduced the unexplained pay gap by less than 1 percent. In contrast to the above analyses, we did not have appropriate proxies for motivation and work performance—factors that may also influence wages—and therefore we could not test the potential effect of these factors on the pay gap.[21]

In addition, certain variables in our model reflect personal decisions that may be correlated with salary, such as whether an employee chooses to work part-time. Including such variables in the model has the potential to lead to biased estimates.

[20] See, for example, Francine D. Blau and Lawrence M. Kahn, "The U.S. Gender Pay Gap in the 1990's: Slowing Convergence, *Industrial and Labor Relations Review,* Vol. 60, No. 1 (October 2006).

[21]While the CPDF include data on performance ratings and grade information, which reflect promotions, these decisions feed directly into determining (and are therefore nearly synonymous with) salary. Therefore, it is more appropriate to evaluate these variables as dependent variables (in the same way that we are evaluating salary).

Table 3: Female coefficient under alternate specifications of the model

		Coefficient on Female (standard error)		
	Specification	1988	1998	2007
1	Main	-.109 (.001)	-.088 (.001)	-.083 (.001)
2	Job family level	-.097 (.001)	-.076 (.001)	-.072 (.001)
3	Disaggregated occupation, but with grouped blue-collar	-.084 (.001)	-.064 (.001)	-.055 (.001)
4	Job series (the most disaggregated occupation variable)	-.073 (.001)	-.056 (.001)	-.048 (.001)
5	Excluding occupation	-.190 (.001)	-.134 (.001)	-.113 (.001)
6	In addition to PATCOB, we included the proportion of women in the occupation	-.070 (.001)	-.054 (.001)	-.049 (.001)
7	Geography measured by county	-.109 (.001)	-.088 (.001)	-.081 (.001)
8	The addition of educational major to the model			-.076 (.001)
9	The addition of educational major to the model, with Job family level			-.066 (.001)
10	The addition of educational major to the model, with grouped blue-collar			-.053 (.001)
11	Excluding agency and occupation, but major was added			-.105 (.001)
12	Only age, federal experience and degree	-.175 (.001)	-.116 (.001)	-.108 (.001)
13	Only federal experience, PATCOB, and degree	-.112 (.001)	-.085 (.001)	-.089 (.001)
14	The addition of health plan to the model	-.102 (.001)	-.082 (.001)	-.076 (.001)

Source: GAO analysis of CPDF data.

(b) Decomposition approach results

Table 4 show the results of applying the decomposition methodology, as outlined in equation 2, to the econometric model presented in Table 2. The first row details the total difference, the unexplained or parameter difference, and the explained or characteristic difference in each year. The other rows indicate the contribution of each of the factors. The overall conclusion drawn from the decomposition approach is similar to the regression approach, i.e. under both, differences remain between men and women's salaries, even after correcting for a wide range of characteristics.

As the table shows, using the decomposition methodology, the unexplained percentage has been remarkably constant over the past 20 years. Specifically, it was 7.8 percent in 1988, 8.1 percent in 1998 and 7.5 percent in 2007. Because the raw pay gap has been falling, the percentage of the gap explained by characteristics has been decreasing. For example, the percentage explained by characteristics was 76 percent (-.249/-.327) in 1988 and 37 percent (-.045/-.249) in 2007.

The contribution of occupation is the largest component of any of the explanatory variables, accounting for over half of the explained difference in the gender pay gap in each year. Specifically, the contribution of occupation was 14.5 percentage points in 1988, 7.1 percentage points in 1998 and 2.9 percentage points in 2007. Other variables that accounted for a relatively larger portion of the pay gap in 1988 were education-level and years of federal work experience. However, these factors explained less of the pay gap in 1998 and 2007 since the differences between men and women in these factors had declined. These results are depicted graphically in Figure 1.

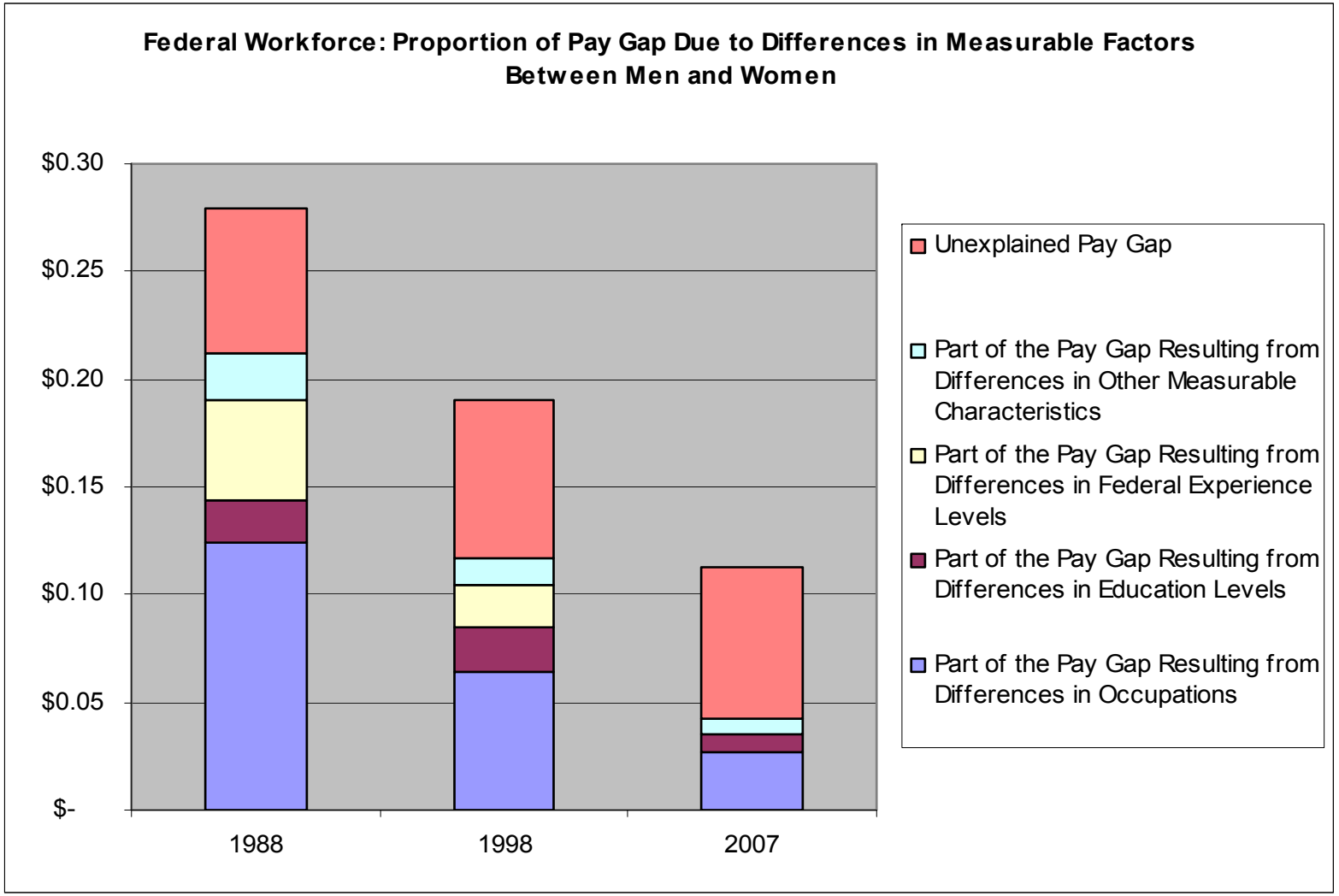
Table 4: Decomposition Results using Main Specification (with contributions of key factors)

	1988			1998			2007		
	Total Gap	Unexplained Gap	Explained Gap	Total Gap	Unexplained Gap	Explained Gap	Total Gap	Unexplained Gap	Explained Gap
Total	-.327	-.078	-.249	-.211	-.081	-.13	-.121	-.075	-.045
Detailed Factors									
Intercept	-.106	-.106	0	.0622	.0622	0	-.04	-.04	0
Age	-.162	-.143	-.019	-.236	-.227	-.009	-.096	-.096	4E-7*
Federal Experience	-.06	-.006	-.054	-.001	.024	-.022	.015	.014	.001
Race/Ethnicity	.004	.016	-.012	-3E-6	.011	-.011	-.002	.009	-.011
Education	-.034	-.01	-.024	-.039	-.016	-.023	-.013	-.005	-.008
Occupation	-.06	.085	-.145	-.01	.060	-.071	.0168	.0456	-.029
Work Schedule	.024	.027	-.003	.0258	.029	-.004	-.024	-.021	-.003
Disability Status	-.015	-.015	.001	-.041	-.042	.001	-.032	-.032	.0003
State	.0129	.0106	.002	-.017	-.019	.002	.0332	.030	.0031
Agency	.013	.023	-.011	.005	.018	-.013	.01	.023	-.013

Source: GAO analysis of CPDF data.

* E reflects multiplication by 10 to that power. For example, "-3E-6" refers to -3 multiplied by 10 to the negative 6th power.

Figure 1



* Note: Logarithmic estimates were converted to percent or dollar changes based on this formula: (exp(β)-1).

Alternate Specifications

As with the regression analysis, we also performed the decomposition using alternate models. The results—which are consistent with the regression approach—are shown in table 5 below. While the size of the unexplained gap varied between models and over time, in almost all of the specifications that included agency and occupation, the size of the unexplained gap remained constant over time. At the same time, for all models, the percentage explained by characteristics has decreased, such that the total gap (explained plus unexplained) has also been falling, albeit at a slower rate. Only in the models without agency and occupation did the unexplained gap decrease (from about 20 percent to about 11 percent) over the past 20 years. However, the percentage explained by characteristics has fallen at a faster rate (from almost 40 percent in 1988 to less than 10 percent in 2007) during that same time period.

Table 5: Decomposition results using alternate specifications

Specification		Total Gap	Unexplained Gap	Explained Gap	Percentage Explained
Main					
	1988	-.327	-.078	-.249	.76
	1998	-.211	-.081	-.130	.61
	2007	-.121	-.075	-.045	.37
Job family level					
	1988	-.327	-.064	-.263	.803
	1998	-.211	-.066	-.145	.688
	2007	-.121	-.067	-.053	.443
Disaggregated occupation, but with grouped blue-collar					
	1988	-.326	-.053	-.273	.836
	1998	-.211	-.054	-.157	.745
	2007	-.120	-.048	-.072	.601
Most disaggregated occupation – job series					
	1988	-.328	-.047	-.281	.857
	1998	-.211	-.050	-.162	.764
	2007	-.120	-.046	-.076	.622
In addition to PATCOB, included the percent female in the occupation					
	1988	-.327	-.022	-.305	.934
	1998	-.211	-.038	-.173	.818
	2007	-.120	-.035	-.085	.705
Geography measured by county					
	1988	-.327	-.080	-.247	.756
	1998	-.211	-.081	-.130	.616
	2007	-.121	-.074	-.047	.390
The addition of educational major to the model					
	2007	-.121	-.069	-.052	.428

Table 5: Decomposition results using alternate specifications

Specification		Total Gap	Unexplained Gap	Explained Gap	Percentage Explained
The addition of educational major to the model, with Job family level					
	2007	-.121	-.060	-.061	.504
The addition of educational major to the model, with grouped blue-collar					
	2007	-.121	-.046	-.074	.615
The addition of educational major to the model, with the most disaggregated occupation.					
	2007	-.122	-.045	-.077	.634
Excluding agency and occupation					
	1988	-.327	-.195	-.131	.403
	1998	-.211	-.141	-.070	.332
	2007	-.120	-.112	-.008	.070
Excluding agency and occupation, but major was added					
	2007	-.120	-.099	-.021	.175
Only age, federal experience and degree					
	1988	-.327	-.174	-.152	.466
	1998	-.211	-.118	-.093	.440
	2007	-.120	-.107	-.013	.112
Only federal experience, PATCOB, and degree					
	1988	-.327	-.065	-.262	.801
	1998	-.211	-.067	-.144	.681
	2007	-.120	-.084	-.036	.303
Health plan was added to the model					
	1988	-.328	-.077	-.251	.766
	1998	-.211	-.076	-.135	.638
	2007	-.121	-.069	-.051	.428

Source: GAO Analysis of CPDF data.

Conclusions

Ultimately, the gender pay gap for the entire federal workforce has declined primarily because the men and women in the federal workforce are more alike in characteristics related to pay than in past years. The existence of a persistent unexplained pay gap between men and women federal workers over a 20-year period, after we controlled for as many factors as our data allowed, means that we can not rule out the possibility that women are being treated unequally in the federal government. However, a few limitations, some of which are common to almost all multivariate analyses, prevent us from definitively determining whether unexplained differences in pay by sex are due to discrimination or to other factors. First, discrimination is not usually

overt, and as such direct measures of it generally do not exist. Second, we lack data on several factors that may legitimately influence wages, such as experience outside of the federal workforce and individual priorities. Third, certain variables included in our model—such as occupation, education level, and part-time status—may have been imprecisely measured or reported. In short, we cannot know whether the persistent unexplained pay gap in both our analyses are due to discrimination or our inability to account for certain factors that cannot effectively be measured or for which data are not available.

Bibliography

Altonji, Joseph G. and Rebecca M. Blank. "Race and Gender in the Labor Market." *Handbook of Labor Economics,* vol. 3 (1999): 3143-3259.

Black, Matthew, Robert Moffitt, John T. Warner. "The Dynamics of Job Separation: The Case of Federal Employees." *Journal of Applied Econometrics,* vol. 5, no. 3, (July – August, 1990): 245-262.

Blau, Francine D., and Lawrence M Kahn. "Gender Differences in Pay." *The Journal of Economic Perspectives*, vol. 14, no. 4, (Autumn 2000): 75-99.

Blau, Francine D., and Lawrence M Kahn. "The U.S. Gender Pay Gap in the 1990's: Slowing Convergence." *Industrial and Labor Relations Review,* vol. 60, no. 1 (October 2006): 45-66.

Borjas, George J. "The Measurement of Race and Gender Wage Differentials: Evidence from the Federal Sector." *Industrial and Labor Relations Review,* vol. 37, no. 1, (October 1983): 79-91.

Bridges, William P. "Rethinking Gender Segregation and Gender Inequality: Measures and Meanings." *Demography,* vol. 40, no.3, (August 2003): 543-568.

Congressional Budget Office. *Characteristics and Pay of Federal Civilian Employees*. Washington, D.C., March 2007.

GAO. *OPM's Central Personnel Data File: Data Appear Sufficiently Reliable to Meet Most Customer Needs*. GAO/GGD-98-199. Washington, D.C.: September 30, 1998.

GAO. *Women's Earnings: Work Patterns Partially Explain Difference between Men's and Women's Earnings.* GAO-04-35. Washington, D.C.: October 31, 2003.

GAO. *Women's Pay: Gender Pay Gap in the Federal Workforce Narrows as Differences in Occupation, Education, and Experience Diminish.* Washington, D.C.: March 17, 2009.

Goldin, Claudia and Cecilia Rouse. "Orchestrating Impartiality: The Impact of 'Blind' Auditions on Female Musicians." The American Economic Review, vol. 90, no. 4, (September 2000): 715-741.

Lewis, Gregory B. "Gender and Promotions: Promotion Chances of White Men and Women in Federal White-Collar Employment." *The Journal of Human Resources,* vol. 21, no. 3 (Summer 1986): 406-419.

Lewis, Gregory B. "Progress toward Racial and Sexual Equality in the Federal Civil Service?" *Public Administration Review,* vol. 48, no. 3, (May – June 1988): 700-707.

Lewis, Gregory B. "Gender Integration of Occupations in the Federal Civil Service: Extent and Effects on Male-Female Earnings." *Industrial and Labor Relations Review,* vol. 49, no. 3 (Apr., 1996): 472-483.

Lewis, Gregory B. "Continuing Progress Toward Racial and Gender Pay Equality in the Federal Service: An Update." *Review of Public Personnel Administration,* vol. 18, no.2 (Spring 1998): 23-40.

Lewis, Gregory B. and Seong Soo Oh. "A Major Difference? Fields of Study and Male-Female Pay Differences in Federal Employment." *The American Review of Public Administration,* vol. 39, no. 2 (March 2009): 107-124..

Light, Audrey and Manuelita Ureta, "Early-Career Work Experience and Gender Wage Differentials." *Journal of Labor Economics*, vol. 13, no. 1. (January 1995): 121-154.

Neumark, David, Roy J. Bank, and Kyle D. Van Nort. "Sex Discrimination in Restaurant Hiring: An Audit Study." The Quarterly Journal of Economics, vol. 111, no. 3, (August 1996): 915-941

Oaxaca, Ronald. "Male-Female Wage Differentials in Urban Labor Markets." *International Economic Review*, vol. 14, no. 3 (October 1973): 693-709.

Olson, Craig A., Donald P. Schwab, and Barbara L. Rau. "The Effects of Local Market Conditions on Two Pay-Setting Systems in the Federal Sector." *Industrial and Labor Relations Review,* vol. 53, no.2, (January 2000): 272-289.

O'Neill, June. "The Gender Gap in Wages, Circa 2000." *The American Economic Review.* vol. 93, no.2, Papers and Proceedings of the One Hundred Fifteenth Annual Meeting of the American Economic Association, Washington, DC, January 3-5, 2003 (May, 2003): 309-314.

Waldfogel, Jane. "The Family Gap for Young Women in the United States and Britain: Can Maternity Leave Make a Difference?" Journal of Labor Economics, vol. 16, no. 3, (July 1998): 505-545.