Public Labor Premiums

Thaddeus J. Woodman

Nickerson & Associates, Summer 2009

We conducted a preliminary study on how compensation (e.g., income, health benefits, paid vacation, retirement plants, etc…) for public employees compares to compensation for private employees and how this differential has changed over time. Our general hypothesis is that public employees receive a premium for their labor and that this premium has increased over the last 30 years. Though this hypothesis can be tested using a straightforward linear regression, several problems with the available data make an adequate empirical analysis difficult (if not impossible). Despite this, we find a few interesting trends in specific occupations, which conform to our hypothesis, indicating that this study is worth pursuing. These preliminary findings along with a number of recommendations provide a good starting point for future research on this topic.

1. Introduction and Hypothesis

With the tightening of government budgets, policymakers and pundits alike have taken interest in public employee compensation. Today, a general news search for "public employees" will return a host of stories about governors and state legislators fighting to reduce benefits for public workers. What underlies these recent political efforts appears to be the notion that public sector workers receive too much or are unaffordable. There appears to be what economists refer to as a *premium* on public labor: an extra profit received by public workers simply because they are employed by the public sector. The general objective of this report is to test whether such a premium exists and, if it does exist, how it has changed over time.

Our general hypothesis is that there is a premium on public labor which has increased over the last thirty years. We believe this premium exists even after controlling for the major characteristics of the public sector: e.g. that it is highly unionized; tends to employ more minorities than the private sector; and that it employs older, more experienced workers than the private sector. However, we expect this general hypothesis to fail for specific occupations and for certain periods of time. For instance, we expect that for many high-skilled occupations, especially high-tech occupations, private employees began to receive significantly more compensation than their public sector counterparts in the 1990s as a result of immense growth in the private sector during this period.[1] With this general caveat aside, there are several theoretical considerations support our general hypothesis.[2]

First, many of the market forces that determine compensation in the private sector do not play an active role in determining compensation for public employees. Compensation for public employees is set by politicians and bureaucrats whose objectives may differ markedly from employers in the private sector. Whereas private sector employers typically seek to maximize profits/productivity, the motivations of public sector employers may be diverse. For example, politicians and bureaucrats may set public sector compensation in such a way to correct or compensate for discrimination in the private labor market. Unlike firms in the private sector, the productivity of labor in the public sector bears a weaker connection (if any at all) to generating government revenue. For instance, a decline in productivity for laborers in the Department of Transportation will not necessarily imply a decrease in state revenue. As a result, political decisions to set levels of compensation and employment are less constrained by efficiency, giving politicians more freedom to make such decisions with other objectives in mind.

Second, it is plausible that bureaucrats are in a special position to inflate the budgets of their bureaus and compensation for public employees. Monitoring the performance of public sector employees is usually costly and politicians may not have

---

[1] For more on this point, see: Borjas, G. (2002). The Wage Structure and the Sorting of Workers into the Public Sector. NBER Working Paper Series. Cambridge, National Bureau of Economic Research**:** 45.

[2] Gregory and Borland provide thorough overview of the public sector labor theory as well as a nice survey of the empirical studies ranging from the early 80s to the mid 90s. See: Gregory, R. G. and J. Borland (1999). Recent Developments in Public Sector Labor Markets. Handbook of Labor Economics. O. Ashenfelter and D. Card. Amsterdam, North-Holland. 3C: 3573-3630.

direct incentives to increase the efficiency of public labor. Insofar as bureaucrats are interested in increasing the size of their bureaus and compensation for public employees there appear few constraints to limit their efforts.

Finally, public sector employees constitute a sizeable and reliable voting block (especially in mid-term elections). All things equal, the self-maximizing public employee has a direct incentive to support candidates who are most likely to maximize their bureau's budget and/or the employee's personal wealth. Blair (2003)[3] finds evidence that the percentage of bureaucrats in a representative constituency does influence the representative's fiscal liberal/conservativeness in the expected direction. Building off Blair's analysis, it is plausible that electoral influence of public employees has gradually added to a premium on public labor since the late 70s.

2. Models and Data

To test our hypothesis we developed a general model intended to capture the major determinants of employee compensation:

$$\mathbf{I}_i + \mathbf{NIC}_i = f\,(\,\mathbf{SECT}_i,\ \mathbf{OCC}_i,\ \mathbf{DEMO}_i,\ \mathbf{GEO}_i,\ \mathbf{UNION}_i,\ \mathbf{YEAR}_i,\ \mathbf{YEAR}_i \bullet \mathbf{SECT}_i\,)$$

Where:

I: Annual income of the i^{th} employee.
NIC: All non-income compensation of the i^{th} employee.
SECT: A dummy variable that = 1 when the i^{th} employee is employed by the public sector.
OCC: A dummy variable for the occupation of the of the i^{th} employee.
DEMO: Demographic variables for the i^{th} employee (e.g. A dummy the i^{th} employee's sex, race, age, educational attainment, etc…).

---

[3] Blair, W. (2003). Bureaucratic Influence in Congressional Roll-Call Voting. The Department of Political Science, Louisiana State University. Doctor of Philosophy: 94.

GEO: A series of dummy variables for the geographic region (for our analysis, the state) of the i^{th} employee.
UNION : Whether the i^{th} employee is a member of a union or receives union benefits.
YEAR: A series of dummy year variables ranging from 1978-2008 that correspond the year that I + NIC of the i^{th} employee. This variable would account for much of the variation between years due to inflation and other macroeconomic forces.
YEAR•SECT: An interaction variable between YEAR and SECT intended to capture the impact of the i^{th} employee's sector for a particular year.

For this model, the estimated coefficient for the SECT and YEAR•SECT variables are most pertinent to our study: the first estimated coefficient capturing the overall impact of the public/private sector on compensation across time; the second capturing the impact of the public/private sector for a given year. Charting the YEAR•SECT estimated coefficients for each year should yield at least a rough picture of how public sector premiums have increased over time (if they exist at all).

In addition to the first model we also considered an 'occupation-specific' model that would allow for analysis of a certain occupation across time.

$$\mathbf{I}_{i,o} + \mathbf{NIC}_{i,o} = f\,(\mathbf{SECT}_{i,o}, \mathbf{DEMO}_{i,o}, \mathbf{GEO}_{i,o}, \mathbf{UNION}_{i,o}, \mathbf{YEAR}_{i,o}, \mathbf{YEAR}_{i,o}\bullet\mathbf{SECT}_{i,o})$$

Where:

o: is the occupation of the i^{th} employee.

As mentioned earlier, there are empirical and theoretical grounds to assume that different occupations would experience different compensation differentials over time. Following a 2008 study by the Canadian Federation of Independent Business, we opted for a model that would capture how compensation has changed for different occupations over time.[4] We believe that disaggregating the data by occupation is essential for a preliminary

[4] This study can be downloaded at: http://www.cfib.ca/research/reports/rr3077.pdf. It is important to note that our methodology departs substantially from the CFIB study, which analyzed median wages of comparable occupations between the public and private sectors in Canada.

analysis as it yields a more fine-grained perspective and may allow researches to explain trends in a more aggregated model.

After considering a variety of options, we concluded the best way to test our hypothesis and approximate one of the above models was to use data gathered from the Census' CPS March Supplement.[5] While the March Supplement has some strengths, it also has several shortcomings that prevented tests using one of the above models. Most notably the survey lacks detailed information about non-income compensation that would allow for an approximation of the NIC variable.[6] This drawback meant that the dependent variable had to be reduced simply to income, I. The general form of the specification we tested was thus,

$$\mathbf{I}_{i,o} = f(\mathbf{SECT}_{i,o}, \mathbf{DEMO}_{i,o}, \mathbf{GEO}_{i,o}, \mathbf{UNION}_{i,o}, \mathbf{YEAR}_{i,o}, \mathbf{YEAR}_{i,o} \bullet \mathbf{SECT}_{i,o})$$

There were other serious limitations of the March Supplement survey that we will address after briefly summarizing our results.

3. Methodology, Results, and Analysis

[5] The other surveys considered were National Compensation Survey by the Bureau of Labor Statistics, and the Survey of Consumer Finances by the Federal Reserve. The brevity of this study prevented us from taking more than a cursory look at both studies. While we believe that it would be worth while to critically examine both of these datasets we also suspect that they will have the same sorts of pitfalls that we encountered using the March Supplement.

[6] The March Supplement does contain several variables that measure whether the respondent received health benefits and pension, but we considered this indicator to crude to allow for an estimation of what these benefits are worth. Moreover, there are many other forms of non-income compensation that the survey lacks: e.g., paid vacation, dental insurance, and life insurance.

On the whole, the results were mixed. There were fewer steady trends over time than we expected. Nevertheless, there do appear to be steady and believable trends for a few of the occupations surveyed.

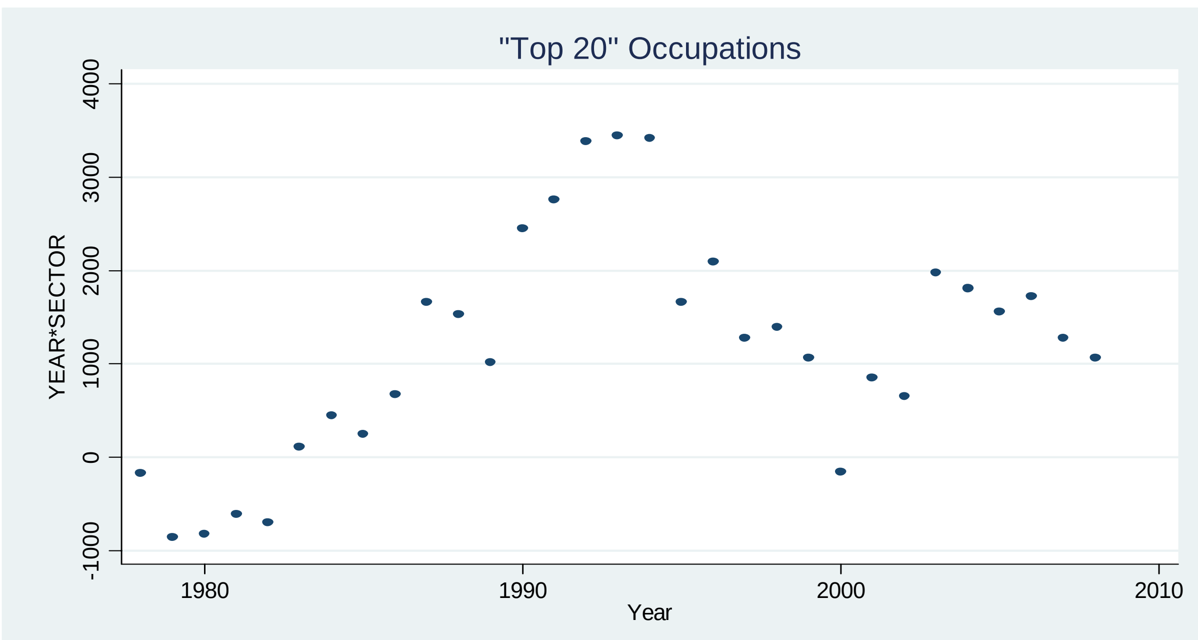
Since the purpose of the study is to gain a general idea of compensation premiums and how they have changed over time, we reduced the size of the dataset to facilitate our analysis. We decided to only keep respondents between the ages of 18 and 65, who worked a fulltime and full-year job and whose occupation code was found in observations for both public and private employees. Finally we filtered out observations where respondent's income was too low—well below the minimum wage for that year---or too high---twice as high as the average income for occupation *o* for that year *y*. We then sorted occupations by frequency of observations and kept only the 20 most commonly observed occupations listed below.

1. Managers and administrators
2. Secretaries
3. Truck, delivery, and tractor drivers
4. Supervisors and proprietors of sales jobs
5. Salespersons
6. Primary school teachers
7. Production supervisors or foremen
8. Janitors
9. Registered nurses
10. Accountants and auditors
11. Cooks, variously defined
12. Bookkeepers and accounting and auditing clerks
13. Nursing aides, orderlies, and attendants
14. Machine operators
15. Secondary school teachers
16. Laborers outside construction
17. Cashiers
18. Computer systems analysts and computer scientists
19. Managers and specialists in marketing, advertising, and public relations
20. Administrative support jobs

We then ran the following regression on this aggregate dataset using the following specification to get a rough idea if there were any trends.

$$\mathbf{I}_i = \mathbf{SECT}_i + \mathbf{DEMO}_i + \mathbf{UNION}_i + \mathbf{YEAR}_i + \mathbf{YEAR}_i \bullet \mathbf{SECT}_i$$

The results from the regression are listed below:



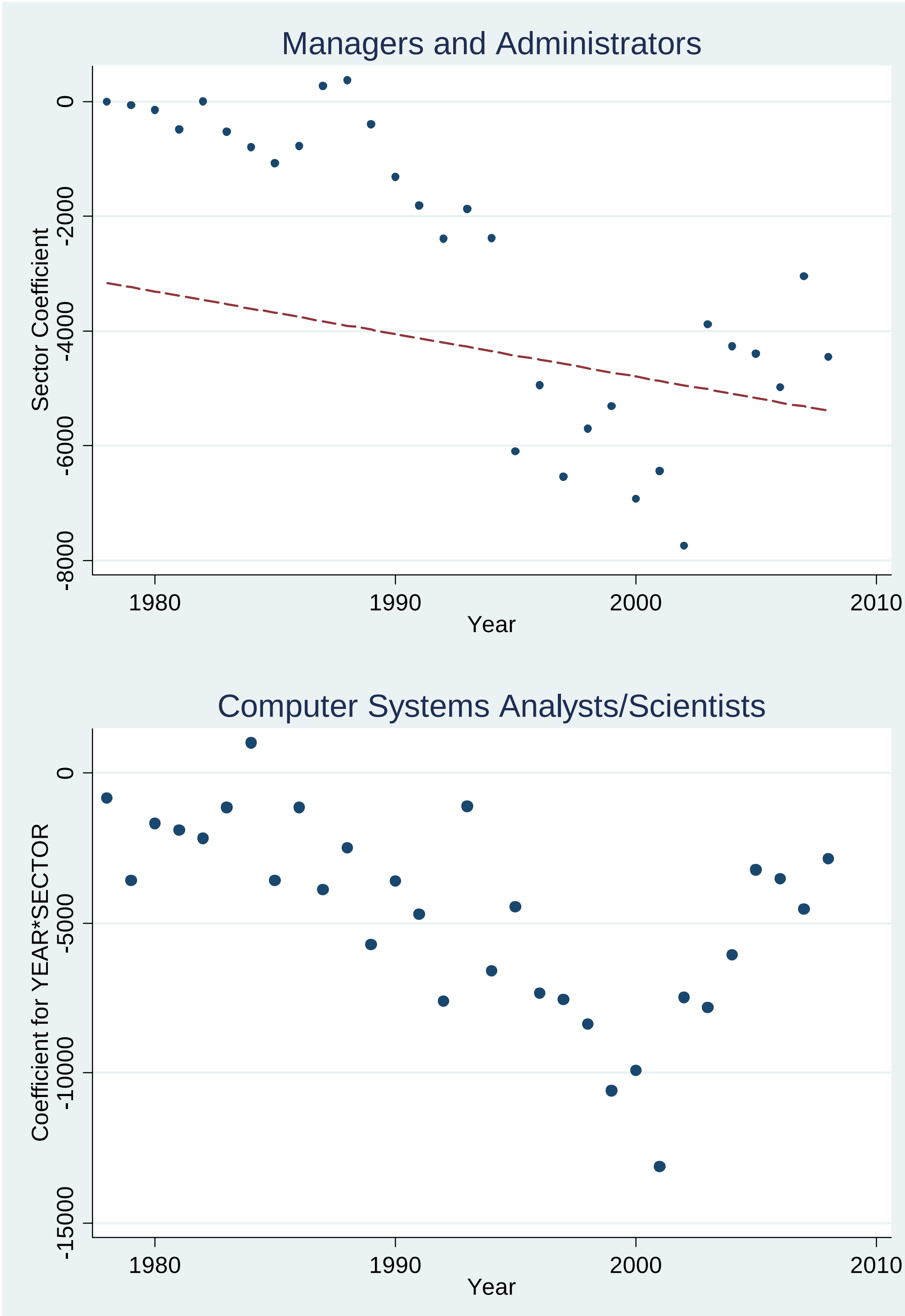

```
-------------------------------------------------------------------------------
      adjinc |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+-----------------------------------------------------------------
      sector |  -5936.414   373.3209   -15.90   0.000    -6668.112   -5204.717
 sector*1978 |  -170.1192   525.1943    -0.32   0.746    -1199.484    859.2453
     .  1979 |  -859.2861    525.301    -1.64   0.102     -1888.86    170.2875
     .  1980 |  -819.5963   502.5883    -1.63   0.103    -1804.654    165.4612
     .  1981 |  -614.8632   502.4177    -1.22   0.221    -1599.586    369.8599
     .  1982 |  -699.7941   516.5955    -1.35   0.176    -1712.305    312.7171
     .  1983 |   110.9663   516.6979     0.21   0.830    -901.7455    1123.678
     .  1984 |   445.3643   516.1704     0.86   0.388    -566.3137    1457.042
     .  1985 |   246.0825   520.6969     0.47   0.636    -774.4673    1266.632
     .  1986 |   669.8294   520.9391     1.29   0.199    -351.1951    1690.854
     .  1987 |   1658.894   521.9022     3.18   0.001     635.9823    2681.807
     .  1988 |   1528.867   522.0571     2.93   0.003     505.6508    2552.082
     .  1989 |   1015.193    531.772     1.91   0.056    -27.06378    2057.449
     .  1990 |   2453.822   522.1629     4.70   0.000     1430.399    3477.245
     .  1991 |   2759.788   526.3816     5.24   0.000     1728.096     3791.48
     .  1992 |   3386.243   525.4765     6.44   0.000     2356.326    4416.161
```

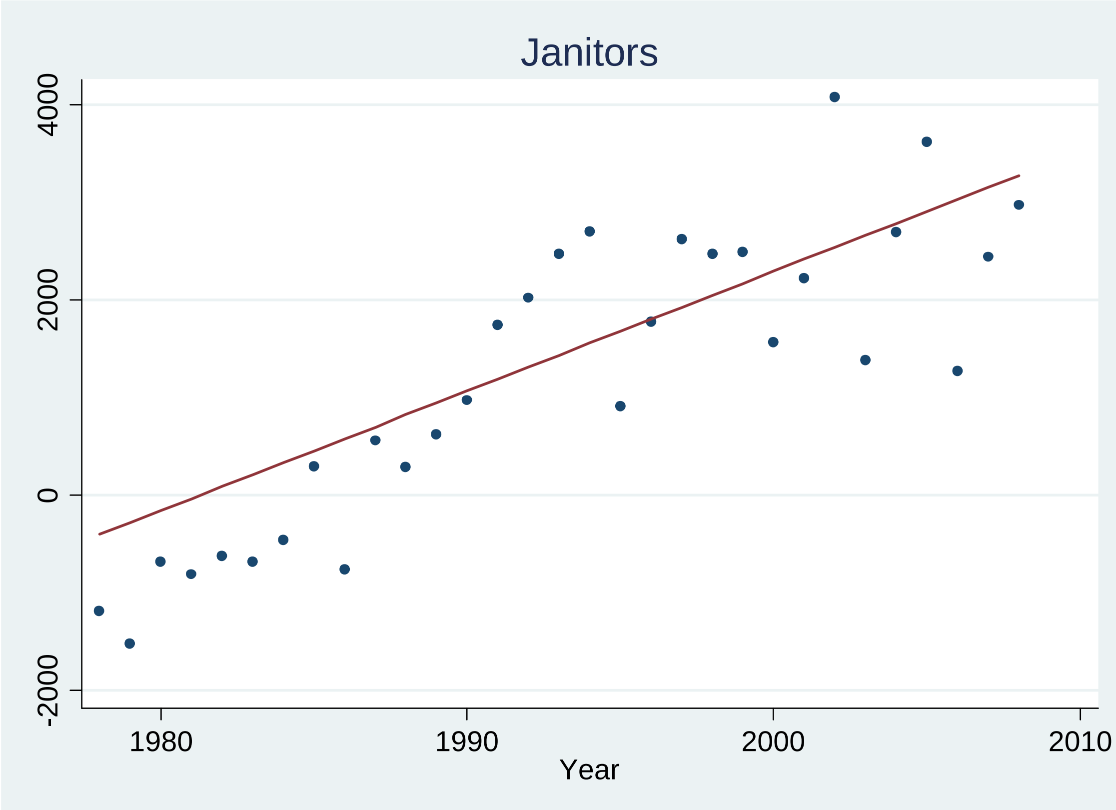
. 1993	3447.833	519.0811	6.64	0.000	2430.451	4465.216
. 1994	3416.495	530.1695	6.44	0.000	2377.38	4455.611
. 1995	1659.614	534.3792	3.11	0.002	612.2474	2706.981
. 1996	2088.063	560.0934	3.73	0.000	990.2977	3185.829
. 1997	1275.369	563.9751	2.26	0.024	169.9952	2380.743
. 1998	1393.086	561.6969	2.48	0.013	292.1777	2493.995
. 1999	1061.575	554.0342	1.92	0.055	-24.31458	2147.465
. 2000	-162.5871	554.8969	-0.29	0.770	-1250.168	924.9936
. 2001	852.5819	556.1758	1.53	0.125	-237.5054	1942.669
. 2002	649.3499	489.0554	1.33	0.184	-309.1835	1607.883
. 2003	1975.503	491.9146	4.02	0.000	1011.365	2939.64
. 2004	1807.965	495.3303	3.65	0.000	837.1327	2778.797
. 2005	1552.798	497.1475	3.12	0.002	578.4042	2527.191
. 2006	1722.738	495.9444	3.47	0.001	750.7029	2694.774
. 2007	1277.827	495.2363	2.58	0.010	307.1793	2248.475
. 2008	1057.319	492.1918	2.15	0.032	92.63858	2022
age	369.6337	2.345136	157.62	0.000	365.0373	374.2301
male	11192.42	52.1842	214.48	0.000	11090.14	11294.7
white	3912.692	79.90724	48.97	0.000	3756.076	4069.307
hischool	9699.464	89.39997	108.50	0.000	9524.243	9874.685
college	16550.77	59.55948	277.89	0.000	16434.03	16667.5

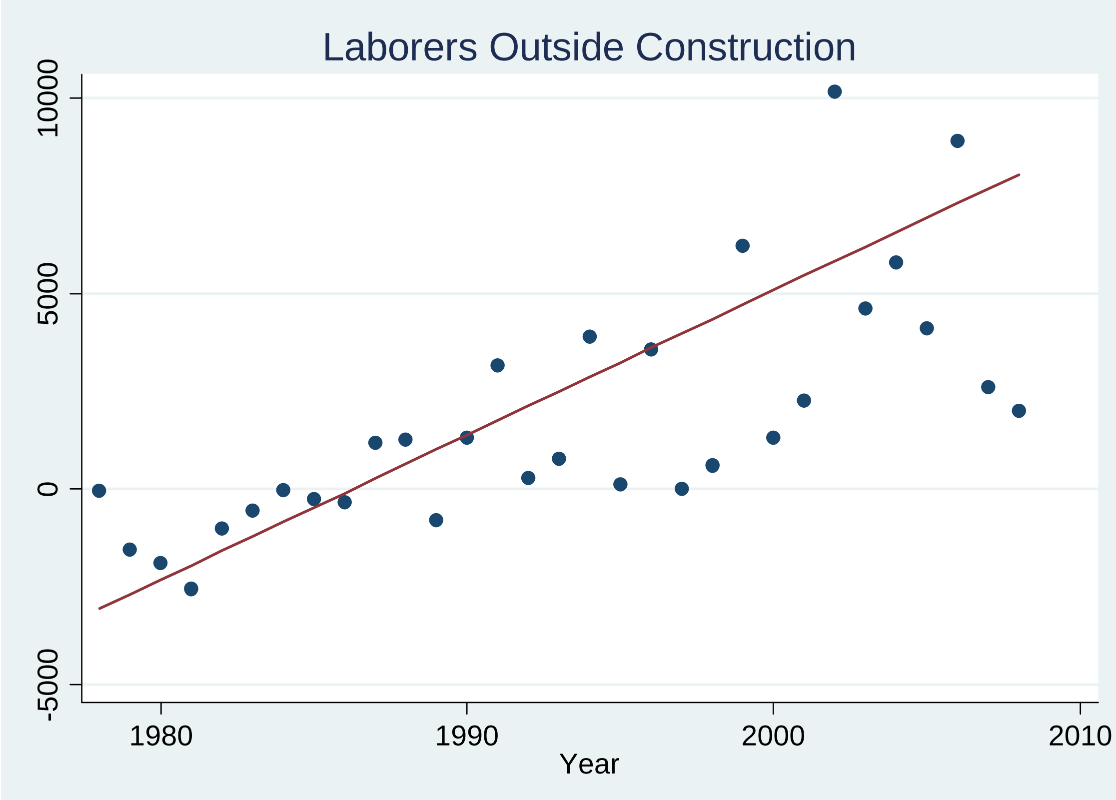
There are several notable features about these results. The first is that many of the estimated coefficients for the SECTOR*YEAR variable appear statistically significant. The second is that there do appear to be trends for certain periods. The most recognizable is a steady trend in public sector premiums from 1978 to 1996. Similarly, there appears to be a steady decline in the public premium from the mid-90s until about 2002. Both of these trends agree with our hypotheses stated at the outset: the premium for public labor increased until the mid-90s and then fell considerably from the mid-90s until around 2002. Finally, the estimated coefficient for SECTOR is large (almost \$6,000), negative and appears highly significant (t-value ≈ 16). While this does not necessarily conflict with our general hypothesis that premiums have increased over time, it does seem odd; it is not clear how this should be interpreted where nearly all the estimated coefficients for SECTOR*YEAR are positive.

The volatility of the estimated coefficients after 1994 raises some concern, especially the jumps from 1999 to 2000, 2000 to 2001, and 2002 to 2003. It is not clear what accounts for this high degree of year to year variation. On one hand the variation in the magnitudes of the estimated coefficients for these years is not incredible (the largest year-to-year differences are around $1,200), but they are relatively large compared with other year-to-year differences in the data and seem out of place. Moreover, it is not clear what explains what appears to be a rapid rise in public labor premiums from 2000-2003 and a steady decline thereafter.

To see if trends within certain occupations might explain some of these unexpected results, we ran the same specification on each of the 20 occupations. The results on the whole were not helpful although they tend did agree with hypotheses. For many occupations, there was a large degree of year to year variation in the estimated coefficients for YEAR*SECTOR and no noticeable long-term trend (although there tended to be a consistent premium or lack there of across time.) However, the results that did exhibit a long-term trend seemed consistent with a major thesis of Borjas (2002): public compensation for *lower-skilled* labor has increased (since the 80s) relative to private compensation, while public compensation for *higher-skilled* labor has decreased relative to private compensation during the same period. Below are several data plots that reflect this finding; two higher-skilled occupations and two lower-skilled occupations.


Managers and Administrators
0
-2000
-4000
-6000
-8000
Sector Coefficient
1980
1990
2000
2010
Year
Computer Systems Analysts/Scientists
0
-5000
-10000
-15000
Coefficient for YEAR*SECTOR
1980
1990
2000
2010
Year


Janitors
4000
2000
0
-2000
1980
1990
2000
2010
Year


Laborers Outside Construction
10000
5000
0
-5000
1980
1990
2000
2010
Year

We believe several problems with the March Supplement data account for much of the year-to-year variation in the estimated coefficients for the YEAR*SECTOR variable. First, the occupational codes in the March Supplement are problematic. Many occupational categories are highly general, if not vague. Some of the most frequently observed occupation categories were among the worst offenders (e.g., categories such as "administrative support jobs (not elsewhere classified)"). In almost all categories, there is a large degree of variation among incomes for a particular year. For example, for many lower-skilled occupations, fulltime full-year employees made well below what they would have earning minimum wage, while many others in the same category made nearly 300% the average wage for that year. This variation in income suggests that the recorded wages are inaccurate or that the categories are not fine-grained enough to allow for a meaningful comparison between workers with the same occupational codes.

Second the occupation coding scheme of the March Supplement changed three times between 1978 and 2008. While there is a recoded occupation variable designed to amend these code changes, it seems likely that problems remain. For example, for many of the occupations we observed there appears to be an explained jump between 2002 and 2003, a year when the coding scheme changed. The details of the recode schema are no doubt complex, and frankly we did not closely examine the recoded variable. However, we suspect that some of the variation in our results may result similar groups of laborers being classified differently across different coding schemes.

Finally, the variable for occupation in the March Supplement is not synchronized with the variable for income: the former variable measures the respondent's *current* job, the latter measures the respondent's income for the *previous calendar year*. Thus if the respondent has changed jobs within the last year, her recorded income will not correspond with her current occupation. Though we made adjustments to the dataset in an attempt to correct for this difficulty[7], it is possible that this discrepancy is responsible for some of the noise we observed in our results.

Moreover, our inability to estimate employees' non-income compensation using the March Supplement is almost certainly biasing the results of our regression. A simple analysis of the dataset shows that public employees receive both health benefits and retirement plans at a much higher rate (10-25%) than private employees. Since our analysis only focused on income, our results may tell us very little about how total compensation differs between sectors. It seems that for lower-paying jobs where non-income compensation may comprise a larger share of total compensation this variable is essential for a cross-sector comparison.

As mentioned in an earlier footnote, these problems with the March Supplement data probably occur the other commonly used surveys as well. In general, it seems that for many surveys non-income compensation measures are too crude (if they exist at all) and occupational codes are too general to test the ideal model for this study. There is one promising data source that we did not pursue: the IRS. We suspect that the IRS has

[7] Specifically, we compared the variable OCCLY which indicates the respondent's occupation in the previous calendar year to OCC (the respondent's current occupation) and rejected observations where the two did not match.

detailed information on non-income compensation and occupation for taxpayers. Those who are interested in taking up this study should see what they can get from America's favorite bureau.

We also suspect that a new, relatively inexpensive survey could go a long way to answering some of the main questions of this topic. Such a survey could have a fairly narrow scope: find 10 or 20 specific occupations (ideally spanning across the income spectrum) that are comparable across sectors, and each year survey about a 100 employees for each sector for each occupation. Although it would take some time before such a survey could reach any conclusions about how public labor premiums change over time, the cross-sectional should be valuable in itself. That we currently lack anything more than educated guesses about public labor premiums should surprise economists and policy makers alike, and should motivate both to pursue further research.