Appendix

Current Population Survey Data

The Current Population Survey is a monthly survey of about 60,000 households. In any given month, one adult household member reports employment and other information for each member of the household. A subset of households reports earnings and hours information. These are the outgoing rotation groups, and each year since 1979 these interviews are gathered together into a single Merged Outgoing Rotation Group (MORG) file. The CPS-MORG includes demographic information on schooling and age; and, information on the worker's main job held such as industry, occupation, and the employer's sector (State government, local government, or private sector).

Given that we are comparing CPS results to those from the ECEC, some of our decisions in processing the CPS differ from those common in the literature. For instance, the ECEC has no age or experience restrictions so we impose none on the CPS. The ECEC does, however, exclude private households, agriculture, the Federal government, and the U.S. postal service, so we exclude these in the CPS as well. Other decisions we made are more standard. We use only wage and salary workers who are currently employed. We make use of both a sample of hourly wages and of weekly, full-time earnings.

We define hourly wages as reported hourly wages for those paid by the hour and usual weekly earnings divided by usual hours worked per week for those who are not. Beginning in 1994, usual hours are missing for that portion of the sample that reports that usual hours are variable. We use a procedure developed by Anne Polivka at the Bureau of Labor Statistics to impute usual hours for this group, which was also used by Lemieux (2010). Weekly earnings are usual weekly earnings for those not paid by the hour and hourly wages times usual hours for those who are paid hourly. Weights used are the earnings weights appropriate for MORG calculations times hours per week for the hourly wage sample, and the earnings weight itself for the weekly earnings sample.

Following Autor, Katz and Kearney (2008), top-coded weekly earnings observations are multiplied by 1.5. Hourly earnings of below $1.675/hour in 1982 dollars are dropped, as are hourly wages exceeding 1/35th the top-coded value of weekly earnings. We deflate using the GDP Personal Consumption Expenditure deflator. Using modified programs made available on David Autor's website, we utilized the procedures of Jaeger (1997) to consistently assign education categories over time in light of the change in the education question in the CPS in 1992 from years of schooling completed to degrees attained, and we used figures from Park (1994) to assign years of potential of experience by race and gender under this revised education question.

We exclude allocated earnings from the calculations. There are four allocations that are potentially relevant in the calculations of hourly wages and weekly earnings: usual hours per week, whether the worker is paid hourly, earnings per hour for those paid hourly, and earnings per week for those paid weekly. If any of these is relevant for a given worker and sample and allocated, we exclude that worker from the sample. As Hirsch and Schumacher (2004) have shown, there is a potential for bias in estimates of wage differentials as a result of an imputation process that does not match on the variables of interest, in this case the sector of employment.

We did, however, redo the analysis including allocated workers and found it made only a small difference in absolute terms, tending to push estimated differentials somewhat closer to zero.

National Compensation Survey/ Employer Costs for Employee Compensation Data

We first describe some general aspects of the NCS and then turn to some specifics for the subsample of it that is used for the ECEC. Readers interested in further details might turn to Bureau of Labor Statistics (n.d., 2003). The NCS sample excludes the federal and agricultural sectors. For private industries in the survey, the establishment usually is at a single physical location. For State and local governments, an establishment can include more than one physical location, for example the sampled unit may be a school district or some other governmental unit within a defined geographic area or jurisdiction. Data collectors visit newly sampled establishments and obtain information on the establishment and a sample of jobs in the establishment. Establishment information includes establishment employment, location, and industry. Whether the establishment is in the private sector, in State government or in local government is known from the survey frame, which includes all employment covered by Unemployment Insurance.

Jobs are sampled with probability proportional to employment in the job, and therefore jobs sampled in an establishment represent the average worker (not the average job). The number of jobs selected generally ranges from 4 to 8 depending on establishment size, though up to 20 jobs may be sampled for State and local government. Other dimensions of the job include information on work level, and indicators for union coverage, full-time status, and whether the job involves commissions or piece rates. The occupational classification system makes use of 6-digit codes from the 2000 Standard Occupational Classification Manual. Where we indicate ECEC occupational controls are at the "detailed level", we mean at a 6-digit SOC level for teaching occupations and a 5-digit level outside of teaching. Overall, this represents a slightly finer classification system than the CPS (primarily due to additional detail within teaching).

To determine work level, data collectors evaluate each job using a "point-factor" system. Currently, four factors are used: Knowledge, Job controls and complexity; Contacts; and Physical environment.[1] Each factor consists of several degrees, each with an associated description and number of points. Generally, the greater the consequence, complexity, or difficulty of the factor, the higher is the number of points assigned and the higher the resulting work level. Factor level information is missing for about 10 percent of the sample. Our main results are similar if these observations are excluded.

For the full NCS sample, earnings are defined to include incentive pay but exclude premium pay for overtime, holiday, and weekend work; shift differentials; bonuses not directly tied to production; payments by third parties such as tips; and payment in kind such as room and board. In order to bring the wage concept closer in line to that of the CPS, we include premium pay and shift differentials in our measure of wages. We weight using ECEC sampling weights times annual hours worked for the hourly samples, and ECEC sampling weights times weeks worked for the weekly samples.

In addition to the information already noted for the wage sample, the ECEC measures the following types of benefits: paid leave—vacations, holidays, sick leave, and personal leave; supplemental pay—premium pay for work in addition to the regular work schedule (such as overtime, weekend, and holiday work) and for shift differentials, and nonproduction bonuses

[1] Up until 2004, a nine-factor system was used and some of the jobs in the dataset were leveled under that system.

(such as yearend, referral, and attendance bonuses); insurance benefits—life, health, short-term disability, and long-term disability insurance; retirement and savings benefits—defined benefit and defined contribution plans; and legally required benefits—Social Security, Medicare, Federal and State unemployment insurance, and workers' compensation. Note that the ECEC by design excludes retiree health plan costs. Also, the ECEC data do capture payments by State government to fund local government workers' benefits plans in local government plan costs.

Transformation of Log Wage Differentials

Using the notation of the text, we transform estimated private-public log differentials into percentage differentials defined as:

$$\Delta = 100 * \left(\frac{\exp\left(\bar{X}_g \hat{\beta}_p\right) E\left(\exp\left(\hat{\varepsilon}_{pk}\right)\right)}{\exp\left(\bar{X}_g \hat{\beta}_g\right) E\left(\exp\left(\hat{\varepsilon}_{gk}\right)\right)} - 1 \right)$$

Thus, not only do the log wage differentials need to be exponentiated, but one must make take account of the expectation of the exponentiated error term, because that will, in general, differ by sector. As the residual variance in the private sector is greater than that in the public sector, the adjustment increases estimates of private sector pay relative to that of the public sector. We calculated this in two ways, assuming normality for the error term and a nonparametric approach attributable to Duan (1983), where the residuals are exponentiated and the averages taken. These yield almost identical results, and we report estimates using the latter approach.

Measuring Teacher Pay

For teachers, the NCS survey captures daily "contract hours," which underestimates hours actually worked (Schumann, 2008; Allegretto, Corcoran and Mishel, 2004). Weeks worked are typically built up from contract school year length, stated in number of school days with allowances for teacher workdays, and we believe these figures to be accurate. Therefore in the ECEC we rely on weekly earnings, and when presenting results on hourly earnings we omit "teaching occupations" (which include certain administrative educators and even noneducation occupations like pilots or flight attendants). This occupational exclusion will not delete certain education sector employment such as office clerical staff or bus depot personnel, where we believe short weekday measures are accurate.

CPS data for teachers may also be subject to substantial measurement error. Podgursky and Tongrut (2006) hypothesize that teachers often generate a weekly earnings report equal to their salary divided by 52 (rather than actual weeks worked). They present some evidence from CPS-MORG and March CPS respondents consistent with this hypothesis. Such misreporting could noticeably affect public-private differentials, and we note that data source differences in weekly wages appear for both the State and local comparisons, but are starker for local government workers.

Benefit Differentials

The benefit cost data for categories like paid leave or health insurance have a large number of observations clustered at zero, with the rest of the observations positive and skewed.

While tobits can be used in this circumstance, we found that they did not fit the data well, and used two-part models instead (Manning, Duan and Rogers, 1987). For each of the three sectors and each benefit category, a dichotomous model was fit using dummy variables for work level as the sole regressors and then the log of the benefit level was regressed on these same dummy variables for those with positive benefit levels. We use the model parameters thereby obtained to predict the benefit costs for each public sector job, if the private sector returns applied. Calculating the percentage differential between an average of these predicted costs and the actual average for the public sector yields the unexplained differentials shown in Appendix Table 1.

Appendix Table 1. Private-Public Benefit Cost Differentials

	Private - State Government		Private - Local Government	
Benefit Category	Raw differential	Unexplained differential	Raw differential	Unexplained differential
Health Insurance	-44.5**	-32.7**	-48.5**	-39.8**
Retirement and Savings	-63.7**	-45.3**	-67.3**	-56.0**
Paid Leave	-43.4**	-15.3**	-27.0**	-3.7
Other Benefits	19.3**	47.7**	23.5**	41.9**

Source: 2009 NCS.

Notes: Estimates are percentage differentials between private and public sector weekly benefit costs, for the given benefit category. Samples are restricted to full-time jobs. Columns labeled "Unexplained differential" are differentials remaining after controlling for job levels. Asterisks ** indicate significance at 1 percent level, while * indicates significance at 5 percent level.