THE CENTURY FOUNDATION

ISSUE BRIEF

VIRTUAL LABOR ORGANIZING: APPENDIX ON METHODOLOGY

Mike Cassidy | June 10, 2015

Union Earnings Premium

Does unionizing lead to higher earnings? In this appendix, we use multivariable regression to answer this question. The benefit of multivariate regression is that it allows us to measure the relationship between unionization and earnings while holding constant other potentially confounding factors. Our model takes the following form:

$$\ln(Y_i)=\alpha U_i+X_i\beta+\varepsilon_i$$

The dependent variable, $\ln(Y_i)$ is the natural logarithm of hourly earnings for individual *i*, a choice guided by decades of research in economics. Logarithms represent rates of change in percentage terms, and, indeed, it is *proportional* changes in *rates* of pay that are most closely related to important economic variables, such as education.

U_i is our key variable. It is an indicator that equals 1 if individual *i* is a member of or covered by a union, and equal to 0 otherwise. Its coefficient, α, gives the percent change in hourly earnings associated with union membership.

X_i is our "all else." Mathematically, it is a vector (that is, a collection) of demographic and labor market characteristics that allow us to control for other factors that previous research suggests can influence earnings. In our case, these factors are: years of education, a quartic in potential years of work experience (earnings tend to increase with experience, but at a decreasing rate), the natural log of usual work hours, and indicators for sex, being married, having children under 18 years of age, six categories of race, three categories of citizenship, twenty-three occupational groupings, fifty states and Washington, D.C., and twelve survey months (to address seasonal effects). Including these factors means that the association we measure between earnings and union membership is net of these other factors.

This brief can be found online at: http://apps.tcf.org/virtual-labor-organizing.

Of course, it is still possible that our model omits other potentially relevant factors. One prime candidate is natural ability; another is work ethic. These and other unobserved factors (unobserved because we do not have data to measure them) are what are captured by ε_i, our equation's "error" term. ε_i can be thought of as what is left over after we have accounted for everything we can. So long as these omitted factors are not systematically related to both union membership and earnings, everything is fine. By contrast, if we have reason to believe, for example, that naturally talented people are more likely to join unions, then we must remember that our union variable may also be picking up the effect of ability. Keep this in mind as we discuss the results: in the absence of a randomized, controlled experiment, the relationship between unionization and earnings is best interpreted as associational, rather than causative.

Our data comes from the Current Population Survey (CPS), a monthly household survey conducted by the Census Bureau and used by the Bureau of Labor Statistics to estimate key labor force characteristics. The CPS provides a representative portrait of the U.S. adult civilian population age 16 years and older.

Because we are interested in earnings, we focus on a subset of the CPS sample, known as the Outgoing Rotation Group (ORG). The ORG is named for the CPS' sampling scheme, in which households are interviewed for four consecutive months, left out for the next eight months, and then interviewed again for four months before leaving the sample entirely. Households in their fourth or eighth months of interviews comprise the ORG—and it is to these households that Census asks questions about earnings. (The reason for confining earnings questions to just a quarter of the sample is to avoid over-burdening the respondents with long questionnaires.) Unlike other sources of earnings data, which tend to rely of retrospective reports, the ORG asks households about their earnings in near real-time—that is, about their recent usual earnings and usual hours—and as such is considered the most accurate comprehensive source of earnings data in the United States.

Our sample is the complete, twelve-month set of ORG data for 2014, formally known as the Merged Outgoing Rotation Group file, which is helpfully cleaned, standardized, and aggregated by the National Bureau of Economic Research. The full sample consists of 317,056 observations, weighted to be representative of the 16+ U.S. population in 2014. Of these, 164,778 report positive earnings.

We further refine our sample in two ways. First, we consider only private sector workers; that is, we exclude government workers, the self-employed, and those who work without pay (say, for family members). This leaves us with 137,399 observations.

Our second refinement is to exclude all respondents for whom the Census Bureau "allocates" earnings or hours. Allocation is necessary because the CPS has a fairly high rate of nonresponse, particularly for questions related to earnings. In these cases, Census researchers assign earnings to people by matching them with a randomly selected individual who shares similar characteristics in terms of sex, age, race, education, occupation, and hours.

Crucially, union status is not one of the parameters included in the allocation process. As a result, the procedure biases downward any union earnings premium that may exist, as non-union workers are allocated the wages of union members, and vice versa. (Allocation also occurs in cases of respondent or interviewer error.) Thus, if we want to accurately measure the impact of unions on wages, we must exclude workers whose earnings are allocated. This gives us 50,375 observations.

Finally, we omit those people for whom data on one of our regression variables is missing, leaving us with a final sample of 50,231 observations, which, after taking into account weighting, represents 39.2 million people.

We run the regression for the full sample, controlling for occupational group, and then for each occupational group separately. The CPS monthly outgoing rotation group weight, adjusted for annual coverage, weights the regressions, and standard errors are heteroskedasticity robust. To check the robustness of our model, we also ran separate specifications, variously controlling for industry, full-time/part-time status, and whether or not someone lives in a metropolitan area, but the results were not meaningfully affected. Overall, our (cross-industry) model explains 46 percent of the earnings variation in our sample, which is quite a good fit as far as earnings regressions go. Key results are presented in Table 1.

Each row represents a separate regression, one for each occupation. The columns are described as follows. The

TABLE 1
UNION WAGE PREMIUM REGRESSION: SUMMARY OF RESULTS

OCCUPATION	EARNINGS PREMIUM	STANDARD ERROR	TEST STATISTIC	HOURLY EARNINGS			# OBSERVATIONS
				UNION	NOT UNION	DIFFERENCE	
Sciences	0.482	0.138	3.50	$31.72	$19.60	$12.12	207
Construction	0.416	0.019	22.14	$25.56	$16.85	$8.70	3,189
Maintenance and repair	0.361	0.023	15.91	$25.69	$17.90	$7.79	2,264
Transportation	0.354	0.022	16.45	$18.52	$13.00	$5.53	4,106
Management	0.341	0.060	5.71	$25.04	$17.80	$7.24	1,659
Social services	0.329	0.119	2.76	$21.97	$15.81	$6.17	498
Groundskeeping	0.277	0.030	9.12	$14.54	$11.03	$3.51	2,435
Manufacturing	0.277	0.016	17.75	$19.23	$14.57	$4.65	5,116
Police and fire	0.248	0.071	3.47	$15.42	$12.04	$3.38	550
Teaching	0.247	0.074	3.32	$17.48	$13.66	$3.82	935
Arts, entertainment, and sports	0.225	0.108	2.09	$20.62	$16.46	$4.16	561
Administrative support	0.210	0.022	9.64	$16.65	$13.49	$3.16	7,985
Business and finance	0.193	0.062	3.12	$22.43	$18.49	$3.94	1,238
Food service	0.190	0.030	6.33	$12.18	$10.07	$2.11	5,004
Computer science	0.170	0.102	1.66	$29.26	$24.68	$4.58	609
Healthcare professional	0.146	0.025	5.95	$28.75	$24.85	$3.90	3,445
Sales	0.144	0.025	5.86	$12.61	$10.92	$1.69	5,855
Healthcare support	0.139	0.037	3.76	$14.72	$12.81	$1.91	2,057
Personal care	0.099	0.047	2.09	$12.27	$11.11	$1.15	1,717
Architecture and engineering	0.127	0.080	1.59	$29.62	$26.08	$3.54	532
Legal							No Obs
Farming, fishing, and forestry	0.010	0.231	0.04	$12.04	$11.92	$0.12	84
OVERALL	**0.276**	**0.007**	**38.35**	**$18.37**	**$13.94**	**$4.43**	**50,231**

union wage premium gives the change in log hourly wages associated with union membership, controlling for the other factors in our model (the coefficient α in our equation); more simply, it is the percent difference in hourly earnings between otherwise similar union members and non-members. Columns two and three give the corresponding standard errors and test statistics for these coefficients; t-statistics greater than 1.64 in absolute value indicate results that are statistically significant at the p = 0.10 level (and the cutoff for p = 0.05 significance is 1.96).

Columns four and five make the union wage premium more concrete. They represent, respectively, predicted average hourly earnings of union members and non-members, respectively, assuming they are otherwise average in all respects considered by the model. Column six is the difference between the two: the hourly earnings premium, in 2014 dollars, for the average worker. The final column lists the number of observations included in each regression.

The final step is to convert the hourly premia into lifetime earnings differences. To do so, we assume a thirty-five-year career of full-time work, which we define as forty hours a week, fifty-two weeks a year. We calculate the lifetime premium in three ways, each of which can be useful, depending on the context. Full results are shown in Table 2.

The most straightforward is the "simple difference," which, as the name suggests, is simply the annual earnings difference multiplied by thirty-five. This gives the cumulative career earnings foregone by not joining a union in today's dollars. Across all occupations, this averages $9,200 times thirty-five years, or approximately $323,000.

However, it is also true that a dollar today is worth less than a dollar tomorrow, because that dollar could be put to productive use. Our "net present value" calculation gets at this idea. We assume the nominal annual interest rate is 5 percent and annual inflation is 2 percent. The former indicates the typical rate of return we would expect on an investment (so investing $100 will give us $105 next year, or, equivalently, having $95.24 today is a good as $100 next year), while the latter reflects the fact that the value of a dollar decreases over time (so that $100 kept under your mattress will allow you to buy only $98.03 worth of goods next year). By putting the two together, we can arrive at the real interest rate—the factor we use to express the time-varying value of money in constant terms. We get the (approximate) real interest rate by subtracting the inflation rate from the nominal interest rate—in our case, it is 5 percent minus 2 percent, or 3 percent (the precise value, using the Fisher equation, is 2.94 percent). We then use the following equation to determine the net present value of the lifetime premium in today's dollars, by adding up the "discounted" premium each year:

$$\mathrm{NPV} = \sum_{i=1}^{35} \frac{\mathrm{A}}{(1+r)^i}$$

where *A* is the annual earnings premium in 2014 dollars ($9,200 across all occupations), *r* is the real interest rate, and *i* indexes years, from year 1 (today) through year thirty-five. Across all occupations, the net present value of the union premium is $200,000. One easy way to interpret this is as follows: $200,000 is the amount a nonunion member would need to receive in a lump sum today in order to make her as well off as a comparable union worker over her career.

There's also a third way to think of the lifetime premium, using similar logic to the NPV approach. Imagine that each year, a union worker invested her full earnings premium, saving for retirement. By the time she retired, how much more money would she have than a similar nonunion worker? Again assuming a real interest rate of 3 percent, we can calculate this amount by making a

TABLE 2

THE COSTS OF NOT UNIONIZING

OCCUPATION	LABOR FORCE SHARE	UNION DENSITY	UNION EARNINGS PREMIUM	HOURLY EARNINGS			ANNUAL EARNINGS DIFFERENCE	LIFETIME IMPACT		
				UNION	NOT UNION	DIFFERENCE		SIMPLE DIFFERENCE	NET PRESENT VALUE	INVESTMENT VALUE
Life, physical, and social sciences	0.7%	5.0%	48.2%	$31.72	$19.60	$12.12	$25,216	$883,000	$547,000	$1,507,000
Construction	5.4%	17.2%	41.6%	$25.56	$16.85	$8.70	$18,100	$634,000	$392,000	$1,082,000
Maintenance and repair	3.6%	13.8%	36.1%	$25.69	$17.90	$7.79	$16,195	$567,000	$351,000	$968,000
Management	9.4%	2.9%	34.1%	$25.04	$17.80	$7.24	$15,059	$527,000	$326,000	$900,000
Social services	1.4%	5.3%	32.9%	$21.97	$15.81	$6.17	$12,823	$449,000	$278,000	$767,000
Transportation	7.1%	15.3%	35.4%	$18.52	$13.00	$5.53	$11,492	$402,000	$249,000	$687,000
Manufacturing	7.1%	13.7%	27.7%	$19.23	$14.57	$4.65	$9,680	$339,000	$210,000	$579,000
Computer science	3.0%	2.8%	17.0%	$29.26	$24.68	$4.58	$9,535	$334,000	$207,000	$570,000
Arts, entertainment, and sports	1.8%	5.5%	22.5%	$20.62	$16.46	$4.16	$8,647	$303,000	$187,000	$517,000
Business and finance	4.6%	2.6%	19.3%	$22.43	$18.49	$3.94	$8,189	$287,000	$177,000	$490,000
Healthcare professional	5.9%	11.4%	14.6%	$28.75	$24.85	$3.90	$8,108	$284,000	$176,000	$485,000
Teaching	2.6%	14.4%	24.7%	$17.48	$13.66	$3.82	$7,952	$278,000	$172,000	$475,000
Groundskeeping	3.9%	7.1%	27.7%	$14.54	$11.03	$3.51	$7,309	$256,000	$158,000	$437,000
Police and fire	1.0%	8.3%	24.8%	$15.42	$12.04	$3.38	$7,035	$246,000	$152,000	$421,000
Administrative support	13.1%	4.9%	21.0%	$16.65	$13.49	$3.16	$6,571	$230,000	$142,000	$393,000
Food service	7.2%	3.7%	19.0%	$12.18	$10.07	$2.11	$4,383	$153,000	$95,000	$262,000
Healthcare support	2.7%	8.6%	13.9%	$14.72	$12.81	$1.91	$3,975	$139,000	$86,000	$238,000
Sales	12.2%	3.5%	14.4%	$12.61	$10.92	$1.69	$3,509	$123,000	$76,000	$210,000
Personal care	3.4%	4.1%	9.9%	$12.27	$11.11	$1.15	$2,398	$84,000	$52,000	$143,000
Architecture and engineering	2.0%	4.4%	Not statistically significant difference							
Legal	1.0%	3.3%	Not statistically significant difference							
Farming, fishing, and forestry	0.9%	1.9%	Not statistically significant difference							
Overall	**100.0%**	**7.4%**	**27.6%**	**$18.37**	**$13.94**	**$4.43**	**$9,214**	**$323,000**	**$200,000**	**$551,000**

Notes: (1) Author's calculation based on 2014 Current Population Survey, as obtained from National Bureau of Economic Research Merged Outgoing Rotation Group files. (2) Earnings are pre-tax, pre-transfer, but include regular overtime, tips, and commissions. Where applicable, earnings are expressed in 2014 dollars. (3) Union hourly earnings premium estimated by a multivariate regression of log hourly earnings (pre-tax, pre-transfer) on union membership/coverage, controlling for sex, a quartic in potential experience, years of education, race, marital status, presence of own children, citizenship, state, month, and log usual hours. (4) Includes private sector workers only. Excludes all workers for whom earnings or hours were imputed by the Census Bureau. (5) The regression was run separately for each occupation group. The CPS monthly outgoing rotation group weight, adjusted for annual coverage, was used to weight the regressions. (6) Hourly earnings represent average marginal effect of union status among workers included in the regression; log earnings are converted to level earnings to ease interpretation. (7) Annual earnings difference assumes 40 hours/week, 52 weeks/year of work. (8) Lifetime impact based on 35-year career; values rounded to nearest thousand. (8a) Simple difference is simply the annual earnings difference times 35. (8b) Net present value is what a hypothetical non-union worker would need to be given in a lump sum today to offset the career loss in earnings from not unionizing; it assumes a nominal annual discount rate of 5%, annual inflation of 2%, and therefore a real annual discount rate of approximately 3%. (8c) Investment value represents the wealth a hypothetical non-union worker could amass if she were to invest the annual union earnings difference each year during her career, again assuming a real annual discount rate of approximately 3%. (9) All results are statistically significant at the 10% level. (10) Caveats: Note that the model does not control for potentially influential factors, such as firm size and unobservable characteristics (e.g., ability), which could account for some of the earnings differences between union and non-union workers. In addition, the results are highly sensitive to the choice of discount rate. Does not include any cost of unionizing, such member fees. Does not include benefits, such as retirement, pensions, health insurance, or paid leave.

simple adjustment to the NPV formula, multiplying by the discount factor rather than dividing by it. In math terms, the investment value formula is:

$$IV = \sum_{i=1}^{35} A * (1+r)^{i}$$

By investing the earnings premium, the union worker will, through the magic of compound interest, see her retirement portfolio grow impressively; for the average worker, investing brings the lifetime premium to $551,000. And remember, we are assuming a fairly conservative rate of return; if the nominal interest rate was 7 percent (closer to average stock market performance) instead of 5 percent, the average union member would be $815,000 wealthier than the average non-union member upon retirement.

Note that these calculations do not include any of the costs of unionizing, such as union dues, which can amount of a few percent of salary each year. However, we also do not account for nonwage benefits, such as health insurance or paid leave. Previous research has shown the union advantage extends to these areas as well; union members typically receive more paid leave and better benefits. As a result, we may be understating the full advantage of unionizing.

Mike Cassidy is a policy associate at The Century Foundation. His research focuses on using economics to understand human behavior, especially as it relates to poverty, inequality, performance, and progress.