

Appendix C: Supplemental Tables and Figures

Table C-1: Regression Results from Fixed Effects Estimates of Teachers' Union Impacts on Resource Levels from the Census/Survey of Governments Using a Balanced Panel of All Districts That are Observed in Every Year of the Sample

Relative Time	Dependent Variable: Log of				
	Real Monthly Full Time Teacher Pay	Full Time Teacher Employment	Student-Teacher Ratio	Real COE Per Student	Total Enrollment
	-5 Years	0.017 (0.021)	0.029 (0.026)	-0.024 (0.022)	-0.153 (0.135)
-4 Years	0.038 (0.032)	-0.016 (0.027)	0.002 (0.016)	-0.006 (0.032)	-0.018 (0.026)
-3 Years	0.009 (0.019)	0.012 (0.023)	-0.010 (0.021)	-0.014 (0.029)	0.004 (0.020)
-2 Years	0.010 (0.017)	-0.022 (0.023)	-0.019 (0.028)	0.032 (0.028)	-0.042 (0.028)
0 Years	0.025 (0.022)	0.029 (0.022)	-0.018 (0.024)	0.076* (0.039)	0.004 (0.016)
1 Year	0.023 (0.023)	0.016 (0.027)	-0.005 (0.027)	0.023 (0.039)	0.005 (0.021)
2 Years	0.023 (0.024)	0.021 (0.033)	0.002 (0.033)	-0.081 (0.078)	0.036 (0.032)
3 Years	0.027 (0.025)	0.047 (0.036)	-0.012 (0.026)	-0.010 (0.031)	0.043 (0.039)
4 Years	0.032 (0.027)	0.045 (0.040)	-0.023 (0.029)	0.026 (0.026)	0.034 (0.043)
5 Years	0.021 (0.030)	0.050 (0.044)	-0.015 (0.030)	0.027 (0.027)	0.048 (0.050)
6 Years	0.012 (0.029)	0.067 (0.049)	-0.039 (0.034)	0.034 (0.029)	0.049 (0.052)
7 Years	0.026 (0.030)	0.021 (0.054)	-0.013 (0.039)	-0.003 (0.035)	0.042 (0.058)
8 Years	0.020 (0.033)	0.035 (0.056)	-0.007 (0.032)	0.012 (0.030)	0.046 (0.061)
9 Years	0.019 (0.036)	0.032 (0.059)	-0.002 (0.036)	0.012 (0.033)	0.031 (0.059)
10 Years	0.005 (0.036)	0.019 (0.062)	-0.044 (0.036)	0.0043 (0.036)	-0.013 (0.058)
Constant	8.225** (0.033)	5.650** (0.068)	2.565** (0.033)	8.697** (0.035)	8.226** (0.060)
N	3683	3684	3474	3835	4047
# Clusters	213	213	213	213	213
R^2	0.737	0.964	0.526	0.590	0.976

¹ Source: Parameter estimates from estimation of equation (2) in the text.

² Regressions include school district and state-specific year fixed effects. All standard errors are clustered at the school district level: * indicates significance at the 10 percent level and ** indicates significance at the 5 percent level.

³ The analysis sample contains only districts that are observed in every survey year and is comprised of never-unionized districts and observations with relative years to union election less than 11.

⁴ Relative year -1 is omitted to make all estimates relative to the year prior to certification.

Table C-2: Regression Results from Fixed Effects Estimates of Teachers' Union Impacts on Resource Levels from the Census/Survey of Governments – Rural Districts Only

Relative Time	Dependent Variable: Log of				
	Real Monthly Full Time Teacher Pay	Full Time Teacher Employment	Student-Teacher Ratio	Real COE Per Student	Total Enrollment
-5 Years	0.027 (0.018)	-0.004 (0.038)	0.009 (0.056)	-0.004 (0.012)	0.001 (0.014)
-4 Years	0.017 (0.012)	0.020 (0.021)	-0.022 (0.016)	-0.006 (0.013)	0.004 (0.013)
-3 Years	0.006 (0.011)	0.009 (0.019)	-0.009 (0.016)	-0.027** (0.014)	0.010 (0.014)
-2 Years	0.012 (0.011)	-0.021 (0.020)	-0.013 (0.018)	0.004 (0.014)	-0.031** (0.014)
0 Years	0.015 (0.012)	0.049* (0.026)	-0.040 (0.028)	0.039* (0.022)	0.004 (0.018)
1 Year	0.011 (0.012)	0.048** (0.023)	-0.032 (0.021)	0.009 (0.015)	-0.003 (0.019)
2 Years	-0.004 (0.014)	0.045* (0.023)	-0.017 (0.021)	-0.017 (0.028)	0.022 (0.022)
3 Years	0.013 (0.014)	0.062** (0.024)	-0.017 (0.021)	-0.011 (0.014)	0.035 (0.023)
4 Years	0.015 (0.014)	0.063** (0.026)	-0.026 (0.024)	0.003 (0.014)	0.042* (0.024)
5 Years	0.014 (0.014)	0.052** (0.026)	-0.019 (0.023)	0.002 (0.015)	0.042* (0.024)
6 Years	0.007 (0.014)	0.071** (0.027)	-0.036 (0.026)	-0.015 (0.016)	0.047** (0.024)
7 Years	0.007 (0.016)	0.058* (0.033)	-0.038 (0.031)	-0.004 (0.018)	0.042* (0.023)
8 Years	0.014 (0.016)	0.068** (0.034)	-0.033 (0.030)	0.004 (0.016)	0.043* (0.024)
9 Years	0.022 (0.018)	0.074** (0.033)	-0.034 (0.028)	0.002 (0.016)	0.045 (0.024)
10 Years	0.023 (0.017)	0.044 (0.030)	-0.032 (0.026)	0.010 (0.018)	0.031 (0.026)
Constant	8.077** (0.023)	4.266** (0.033)	2.540** (0.037)	8.198** (0.015)	6.803** (0.016)
N	6659	6659	6333	10408	10754
# Clusters	1005	1005	1005	1005	1005
R ²	0.695	0.965	0.622	0.933	0.979

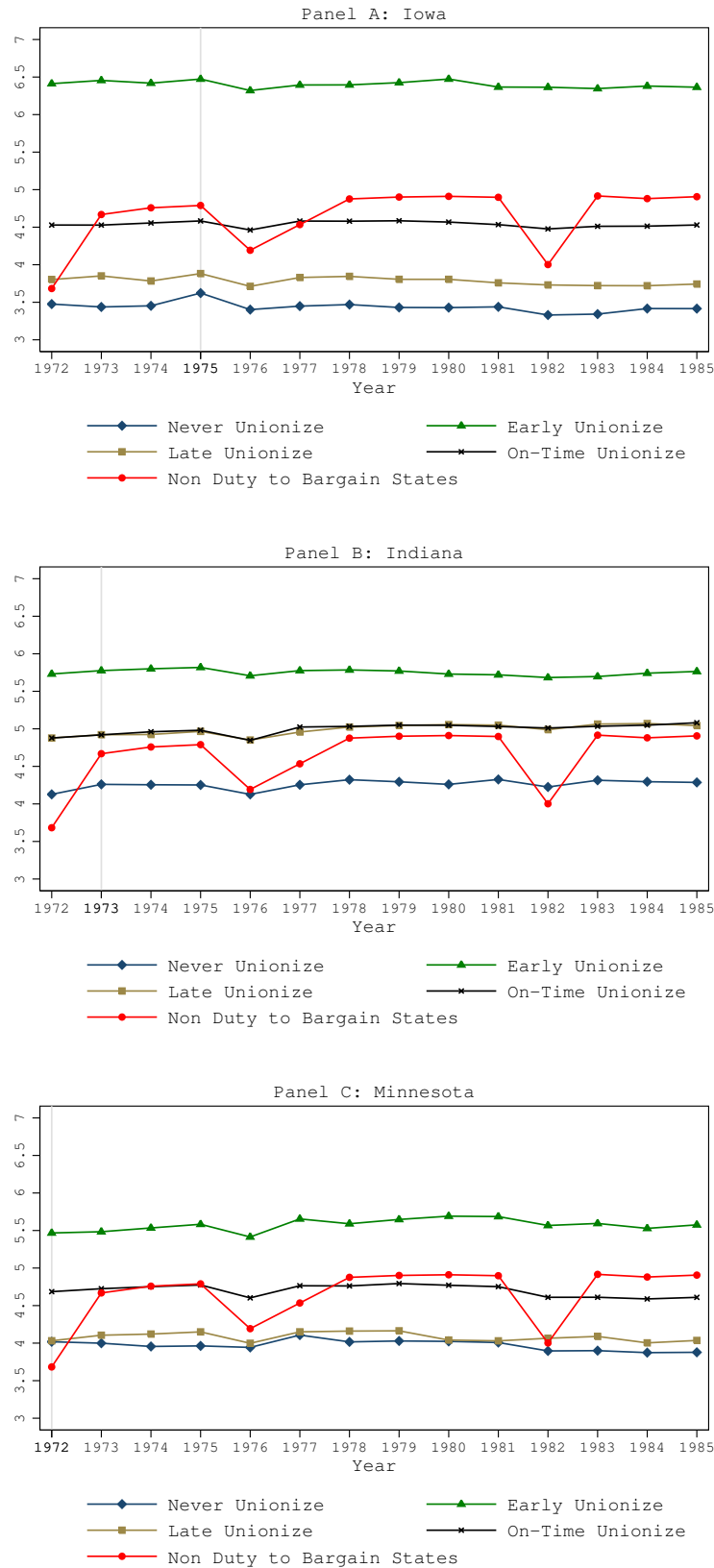
¹ Source: Parameter estimates from estimation of equation (2) in the text.

² Regressions include school district and state-specific year fixed effects. All standard errors are clustered at the school district level: * indicates significance at the 10 percent level and ** indicates significance at the 5 percent level.

³ The analysis sample contains only districts without a Census block in an urban area in any year and is comprised of never-unionized districts and observations with relative years to union election less than 11.

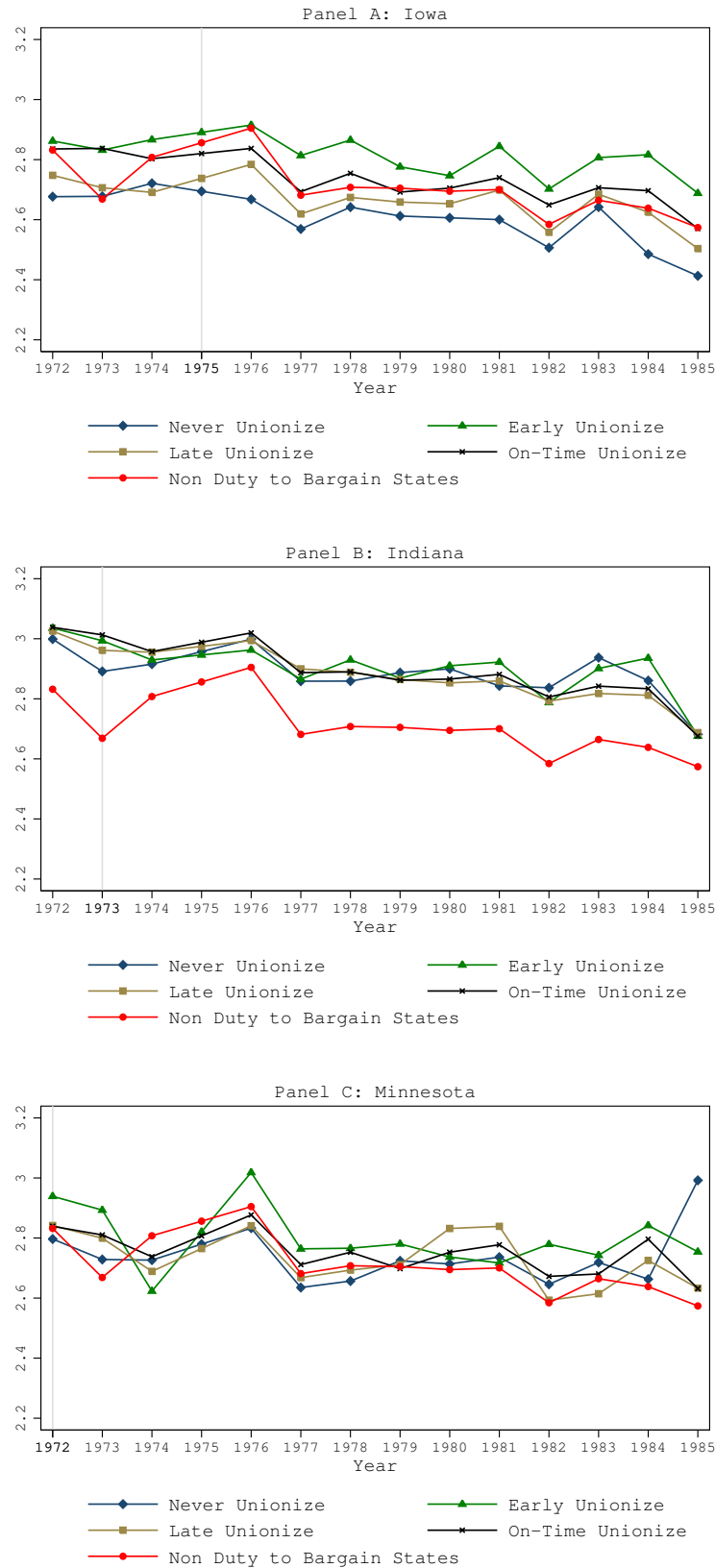
⁴ Relative year -1 is omitted to make all estimates relative to the year prior to certification.

Figure C-1: Trends in Log Full-Time Teacher Employment, by State



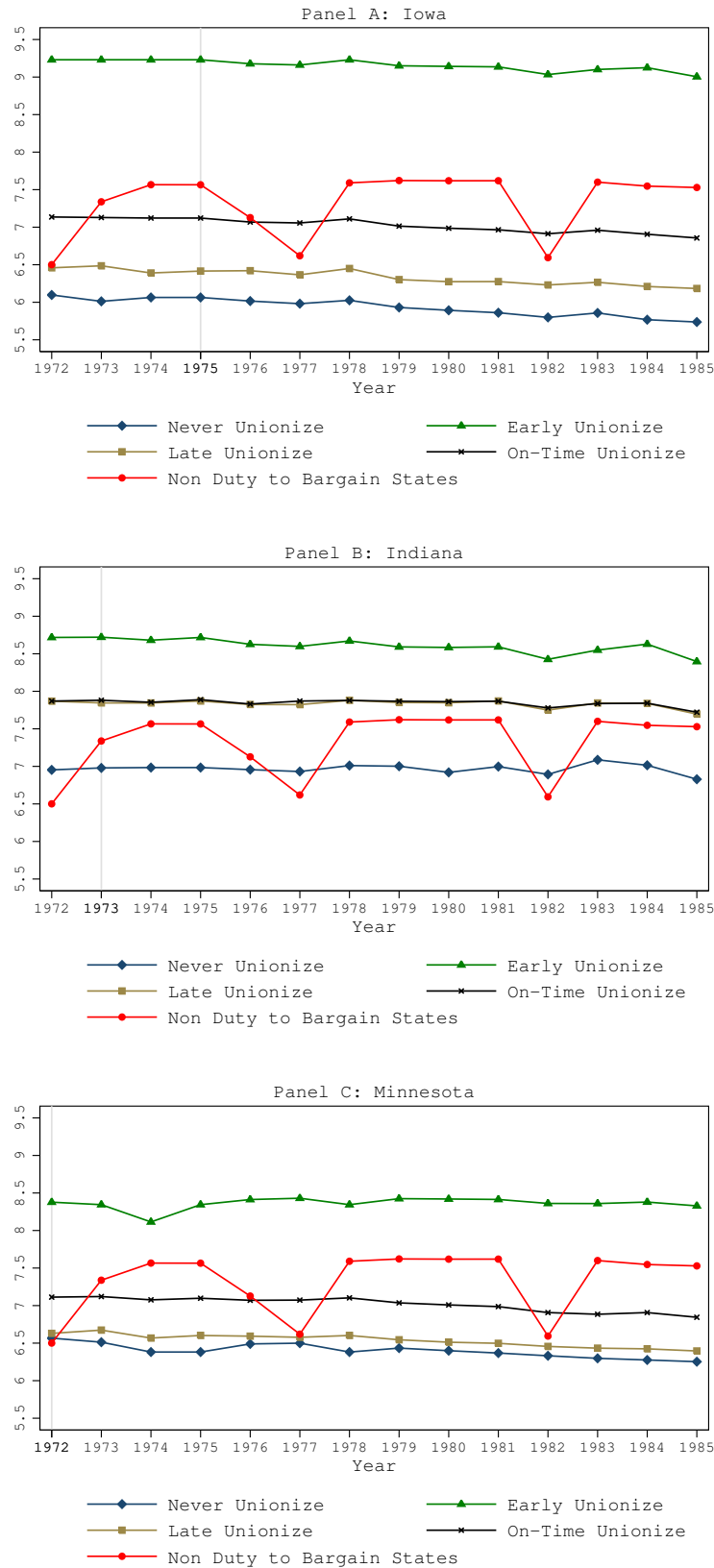
Source: Teachers' union election certification data and Census/Survey of Governments as described in the text. In each panel, the vertical line represents the year in which a duty-to-bargain law was passed in the state. "Never Unionize" districts are those that do not have a successful union election vote by 2004, "Early Unionize" districts are those that unionize prior to passage of their state's duty-to-bargain law, "Late Unionize" districts are those that unionize more than one year after passage of their state's duty-to-bargain law, and "On-Time Unionize" districts are those that unionize within 1 year of passage of their state's duty-to-bargain law. "Non Duty to Bargain States" are the 18 states that did not have a duty-to-bargain law as of 1982: Alabama, Arkansas, Arizona, Colorado, Georgia, Illinois, Kentucky, Louisiana, Missouri, Mississippi, Nebraska, New Mexico, Ohio, South Carolina, Texas, Utah, West Virginia, and Wyoming.

Figure C-2: Trends in Log Student-Teacher Ratios, by State



Source: Teachers' union election certification data and Census/Survey of Governments as described in the text. In each panel, the vertical line represents the year in which a duty-to-bargain law was passed in the state. "Never Unionize" districts are those that do not have a successful union election vote by 2004, "Early Unionize" districts are those that unionize prior to passage of their state's duty-to-bargain law, "Late Unionize" districts are those that unionize more than one year after passage of their state's duty-to-bargain law, and "On-Time Unionize" districts are those that unionize within 1 year of passage of their state's duty-to-bargain law. "Non Duty to Bargain States" are the 18 states that did not have a duty-to-bargain law as of 1982: Alabama, Arkansas, Arizona, Colorado, Georgia, Illinois, Kentucky, Louisiana, Missouri, Mississippi, Nebraska, New Mexico, Ohio, South Carolina, Texas, Utah, West Virginia, and Wyoming.

Figure C-3: Trends in Log Enrollment, by State



Source: Teachers' union election certification data and Census/Survey of Governments as described in the text. In each panel, the vertical line represents the year in which a duty-to-bargain law was passed in the state. "Never Unionize" districts are those that do not have a successful union election vote by 2004, "Early Unionize" districts are those that unionize prior to passage of their state's duty-to-bargain law, "Late Unionize" districts are those that unionize more than one year after passage of their state's duty-to-bargain law, and "On-Time Unionize" districts are those that unionize within 1 year of passage of their state's duty-to-bargain law. "Non Duty to Bargain States" are the 18 states that did not have a duty-to-bargain law as of 1982: Alabama, Arkansas, Arizona, Colorado, Georgia, Illinois, Kentucky, Louisiana, Missouri, Mississippi, Nebraska, New Mexico, Ohio, South Carolina, Texas, Utah, West Virginia, and Wyoming.

Appendix D: Classification Error in the Constructed Census of Governments Teachers' Union Measure

D-1 A Comparison of Alternative Unionization Measures

As discussed in Section 5 in the main text, my union impact estimates differ substantially from those in the established literature. In order to understand these differences more fully, it is instructive first to compare my union election certification data and the union measure constructed from the Census of Governments Labor Relations Survey used in Hoxby (1996), as this is the only other available district-level panel data union measure. The COG does not directly ask respondents about the existence of a teachers' union or a contract with that union. Instead, it contains three survey items related to labor relations that can be used to infer union status in a district:

1. Total number of full-time teachers who are members of an employee organization.
2. Does your agency engage in collective negotiations or meet and confer discussions with employee organizations for the purpose of reaching agreement on conditions of employment?
3. Total number of contractual agreements between your agency and employee organizations in effect as of October 15 of the survey year.

From these survey responses, one can construct a unionization measure using the following criteria: at least 50 percent of teachers are union members, the form of labor negotiations is collective bargaining, and the district has at least one contract or memorandum of understanding with *any* employee organization in effect as of October of the survey year. Note that this union measure is appropriately designed to identify teacher contracts that are collectively bargained with a school district rather than a contract with other employee unions.

While the above measure is the most sensible alternative in the COG, it has several drawbacks. The first is it effectively measures whether a district has a collectively bargained contract with the teachers' union, not whether a teachers' union exists. Given the short lag between certification and negotiation of a first contract, however, this discrepancy likely is small.

The second, more serious, problem is classification error in the COG union measure. Although the COG-based union measure is designed to reduce potential measurement error by making the definition of unionization relatively strict, there are significant differences between the COG and election certification measures of union status that suggest measurement error exists in the former data. Table D-1 contains a comparison of district-level unionization rates from the Census of Governments and the election certifications for each state in the sample. I constructed the Census of Governments measure by a straightforward implementation of the above definition of unionization. Note the COG is conducted every 5 years and labor relations information was only included in the 1972, 1977, 1982 and 1987 surveys.

The table illustrates the substantial differences between the two union measures. In the table, each four-cell square sums to one, and each diagonal within a cell represents the observations for which the union measures agree. For example, in Iowa in 1977, the COG and election certification measures agree 49.89% of school districts were unionized and 26.61% were not. However, 9.31% of the school districts are classified as unionized by the COG measure but had not successfully completed a teachers' union election by that date. Conversely, 14.19% of districts had completed an election but were measured as not unionized by the constructed Census of Governments union measure.

I interpret the disagreement between the two data sources as measurement error, with true union status measured by the election certifications. Given there was little voluntary recognition occurring in these states in this period and the validation study made every attempt to find such districts, measurement error in the Census of Governments is a natural explanation for why there are districts that had not completed a unionization election yet were measured as unionized by the COG. Further, since most districts achieve a contract within a year of certification, the lag between

certification and successfully negotiating a contract cannot explain why so many districts that had certified unions were not measured as unionized by the Census of Governments.

The accuracy of the COG unionization construct also is called into question by the differential time trends in union status within states across measures. Because there are no decertifications, unionization as measured by election certifications weakly increases over time. Thus, conditional on completing a successful election, a district always will be classified as unionized. In contrast, after 1977, unionization rates decline over time in the Census of Governments: while 788 districts were measured as unionized by the Census of Governments in 1977, this number fell to 771 districts in 1982 and to 742 districts in 1987. These declines are not consistent with the lack of any evidence of decertification or cessation of collective bargaining in this time period in these three states.

To investigate further the source of the discrepancy, I look at which of the three criteria used in the COG union measure “fail” when a district has completed a successful unionization election but is not classified as unionized in the COG. I find for such districts in all three states, the provision that the percentage of teachers who are union members must be greater than 50 fails at higher rates over time. It is likely this variable constitutes the main source of measurement error in the COG union measure. First, the Census of Governments is filled out by district administrators, who may not know how many teachers are union members.¹ Second, even if the union membership rate were accurately measured, Iowa, Indiana, and Minnesota are agency shop states, meaning employees are covered by the contract and must pay union dues even when they are not union members. That union membership is reducing over time in these states thus creates measurement error in the COG-based union measure, but this decline in membership, even if accurate, has not translated into changes in collective bargaining status. The problems surrounding the measurement of teacher union membership in the COG are the main source of measurement error in the COG union measure.

Another potential source of measurement error is an increasing number of districts report having no negotiated contracts over time, despite the fact that, conditional on obtaining a first contract, it is rare the teachers are ever without a negotiated contract with the district.² Some of this discrepancy could be due to the fact that even if a contract expires, teachers typically continue to work under that contract until a new one is negotiated with the district. One explanation for the decrease in unionization rates apparent in the COG is expired contracts are coded as “no contract,” despite the fact that negotiated work rules and wage schedules are in place in these districts. These results provide suggestive evidence that the measure of the existence of contracts in the COG contains measurement error.³

Taking the election certification data as the true measure of unionization status, Table D-2 reports the misclassification rates by state and year in the Census of Governments. Aside from 1972, the average misclassification rate remains relatively constant at between 31 and 36 percent in the sample. However, the misclassification rate is as high as 47.2% in Indiana in 1987. Saltzman (1985) provides some outside validation for these misclassification rates. He validates the 1977 Census of Governments union measure for 1000 districts in the U.S. and finds a misclassification rate of 30% for the U.S., which is similar to the 31% misclassification rate I report for my sample of 3 states in that year.

¹As correctly noted by a referee, many school districts will have information on union membership because they deduct union dues from teacher paychecks. However, there is enough year-to-year noise in the data to suggest that at least some school districts are not accurately reporting union membership.

²While there are no available credible aggregate statistics on this assertion, lawyers I have spoken to at both the AFT and NEA agree with this generalization.

³Given the errors in the Census of Governments labor relations data, one must be skeptical of the accuracy of the financial and employment information in these surveys as well. However, since the survey is filled out by the central administrative offices that have access to payroll records and budgets, it is reasonable to expect such data will be supplied with greater accuracy than the number of teachers belonging to the union.

D-2 The Effect of Different Union Measures on Union Impact Estimates

The high misclassification rates from the union measure constructed from the Census of Governments suggest this measure does not accurately characterize the history and state of collective bargaining in the school districts in the sample. In order to understand more fully the differences between my results and those from the existing literature, I undertake a comparison of union impact estimates using the two available district-level union measures. Specifically, I replicate estimates from Hoxby (1996) using both union measures because it is the most comprehensive and empirically sophisticated study of teachers' unions in the literature and because our studies use similar data and time periods.

The empirical specifications in Hoxby (1996) are of the form:

$$Y_{it} = \beta_0 + \beta_1 U_{it} + \delta X_{it} + \tau_i + \phi_t + \psi_i * t + \epsilon_{it}, \quad (\text{D-1})$$

where Y_{it} is an outcome variable of interest, U_{it} is an indicator variable equal to 1 if district i is unionized at time t , X_{it} is a vector of the demographic characteristics listed in Section 3.2 in the main text, τ_i are district fixed effects, ϕ_t are year fixed effects, $\psi_i * t$ are district-specific linear time trends, and ϵ_{it} is a normally distributed error term.

Tables D-3 through D-5 present the results of the regressions when Y_{it} is, alternatively, log real teacher pay, log real current operating expenditures per student, and student-teacher ratios, as these are the dependent variables analyzed by Hoxby. Each regression contains three years of data from the 1970, 1980 and 1990 U.S. Census school district files combined with the 1972, 1982 and 1987 Census of Governments data described in Section 3.2. Column (i) in Tables D-3 through D-5 presents the results taken directly from Hoxby (1996). Column (ii) contains estimates using Hoxby's methodology on my sample of three states and allows me to determine how much of the difference in our estimates is due to the fact I use only three states and she uses all districts in the U.S.. Column (iii) is identical to column (ii), except unionization is measured using the election certification data. This last column thus will yield insight into the effect of the measurement error on union impact estimates.

Hoxby identifies union effects by using only within-state variation through the use of district fixed effects and by an IV strategy that uses only cross-state variation over time in collective bargaining laws. As I only have data for three states, I am restricted to a within-state analysis. While this restriction is justified by my more accurate union data and the lack of evidence of union threat effects as discussed in Sections 4.1, I only will compare my results to the within-state estimates from Hoxby (1996).

Changing the unionization measure has a large impact on estimates of union effects in all three tables. In Table D-3, the union impact estimates on log real teacher pay using the COG union measure are similar in magnitude, sign and statistical significance for the national sample and the Midwest sample. However, when I employ the election certification definition of unionization, the coefficient on the union variable becomes negative, smaller in magnitude, and not statistically significant at even the 10 percent level. Note the standard error on the union coefficient increases by a factor of 3 between columns (ii) and (iii) in Table D-3. This increase occurs because there is variation in the Census of Governments measure that is due to measurement error and is correlated with the dependent variable (see Section D-3). Eliminating this variation increases the standard error estimate substantially. While the union estimate in column (iii) does not allow one to rule out the verity of the union estimate in column (i), it illustrates the fragility of the estimate to correcting for measurement error.

A similar pattern emerges in Table D-4, which presents results for current operating expenditures per student. Switching from the national to the midwest sample reduces the magnitude of the union coefficient, but the signs are the same across columns (i) and (ii). However, in column (iii), the union impact estimate becomes negative when I use the election certification measure and the standard error increases by a factor of 3.8. Table D-5 is more problematic because there is a marked difference between the estimates in the first two columns; the union impact on student-teacher ratios in the

3 midwestern states is of a different sign than for the nation as a whole. However, the difference in union coefficients and the increase in the size of the standard error of these coefficients between columns (ii) and (iii) in Table D-5 is consistent with the sensitivity of the results reported in Hoxby (1996) to measurement error.⁴

The central conclusion from Tables D-3 through D-5 is the classification error reported in Tables D-1 and D-2 in the COG union measure is not innocuous. My results using the Midwest sample are similar to those in Hoxby (1996) for two of the three comparisons, but switching the union measure illustrates those results may not be robust to correcting for measurement error. These comparisons underscore the importance of accurately measuring union status in an analysis of teachers' union impacts.

What is most interesting about the form of the measurement error bias is it is not attenuating, which is the form of bias one would expect from classical measurement error. Classical measurement error occurs when the error is uncorrelated with the dependent variable, the independent variables, the regression error, and the true value of the variable. Despite the fact the measurement error must be correlated with the true measure of union status as union status is a binary variable, Bound, Brown and Mathiowetz (2001) show as long as the misclassification is what Carroll, Ruppert and Stefanski (1995) term "non-differential," the bias in the coefficient will still be attenuating as long as the rest of the classical measurement error assumptions hold. I next turn to an examination of the apparent measurement error in the COG union measure in order to reconcile the results in Tables D-3 through D-5 with the error in variables suggested by Tables D-1 and D-2.

D-3 Statistical Properties of the Classification Error

Table D-1 presents the non-parametric identification of the measurement error in the Census of Governments union measure. I now investigate some properties of this classification error and perform a decomposition that breaks the bias due to the error into the part directly due to measurement error and the part due to the correlation of the measurement error with the regression error.

D-3.1 Non-Differential Classification Error

Let U be union status as measured by the Census of Governments variables and let U^* be true union status as indicated by the election certification data. If μ is the measurement error, then

$$U = U^* + \mu. \tag{D-2}$$

If one only can observe U instead of U^* , then instead of estimating the true model given by:

$$Y = \alpha + \beta U^* + \delta X + \epsilon, \tag{D-3}$$

one must estimate

$$Y = \tilde{\alpha} + \tilde{\beta} U + \tilde{\delta} X + \tilde{\epsilon}, \tag{D-4}$$

where ϵ is the regression error, X is a vector of demographic characteristics assumed to be measured without error, and Y is the outcome variable of interest that contains no measurement error. The standard result under the classical measurement error assumption in which μ is uncorrelated with

⁴Because the Census of Governments union construct measures whether a district has a contract with a teachers' union and the election certification data measure whether a teachers' union exists for the purpose of collective bargaining, one could argue the differences between the estimates in columns (ii) and (iii) of Tables D-3 through D-5 are due to the difference between having a union and having a negotiated contract. As discussed in Section 3.1 of the text, my validation study suggests most districts achieve a contract within one year of certification, and no district fails to achieve a contract conditional on certifying a union. While this difference may cause some attenuation in the results, it cannot account for the sign change in coefficient estimates and is likely to be small.

U^* , X , Y , or ϵ is $\tilde{\beta}$ will be less than β in absolute value. In other words, classical measurement error will cause an attenuation bias. Note this result holds regardless of the number of independent variables measured with error as long as the classical measurement error assumptions hold.

When the mismeasured variable is binary, such as union status, the measurement error (i.e., the classification error) cannot be classical. This result is due to the fact U^* and μ will have to be negatively correlated. For example, if $U^* = 1$, $\mu \in -1, 0$, but if $U^* = 0$, $\mu \in 0, 1$. Thus, the typical attenuation result does not necessarily hold.

Bound, Brown and Mathiowetz (2001) show as long as the misclassification is non-differential and none of the other classical measurement error assumptions are violated, the bias in the coefficient will still be attenuating. Non-differential classification error occurs when reporting errors are independent of the dependent variable. More formally, this can be written:

$$P(U=i|U^*=i, Y) = P(U=i|U^*=i), \tag{D-5}$$

where $i \in 0, 1$. I use a linear probability model to test for non-differential classification error for log real teacher pay, log real current operating expenditures per student, student-teacher ratios, and high school dropout rates. Specifically, I run models of the form:

$$U = \alpha_0 + \alpha_1 Y + \eta, \tag{D-6}$$

where U is an indicator variable that equals 1 if the school district is measured as unionized in the Census of Governments, Y are the dependent variables used in the analysis in the main text, and η is an error term. I perform this test separately for the probability of correctly classifying a district as unionized conditional on being unionized and for the probability of correctly classifying a district as non-unionized conditional on not being unionized. The estimates of α_1 test for the existence of differential classification error. These estimates are presented in Table D-6.

Assuming the election certification data accurately represent true union status, the data strongly reject that the measurement error from the Census of Governments is non-differential. In each row of Table D-6, the estimates of α_1 are statistically different from zero for at least one of the misclassification types. The implication of Table D-6 is the misclassification of union status in the Census of Governments is correlated with the dependent variables of interest; the classification error is differential. The bias due to the error in variables therefore is not guaranteed to be attenuating. This result is consistent with the positive biases in absolute value reported in Tables D-3 through D-5 of union effects when the imperfectly measured union measure is used.

D-3.2 Misclassification as a Function of X

Thus far, I have established the intuition about the effect of measurement error on parameter estimates when the error in variables is classical does not hold because μ is correlated with Y (as the error is differential) and with U^* (as the variable is binary). It is also instructive to determine whether the assumption holds that the measurement error is uncorrelated with the observable X s. To test the relationship between misclassification and the X s, I estimate the probability a district is reported as unionized in the COG when it had successfully completed a union election and the probability a district is reported as non-union in the COG when no union election certification was on file, conditional on observables. More formally, I estimate the following models using a linear probability model:

$$P(U=1|U^*=1, X) \tag{D-7}$$

$$P(U=0|U^*=0, X) \tag{D-8}$$

Table D-7 contains the results from these regressions from the pooled 1970, 1980 and 1990 U.S. Census and Census of Governments data described in the main text. Each cell in the table represents a separate regression. As Table D-7 illustrates, the probability of misclassifying a district's

union status is correlated with the observable demographic characteristics of the district. Some general trends do emerge from Table D-7: smaller, less urban districts with lower public school enrollment are less likely to be correctly classified as unionized, while those districts with lower average income, lower median rent, and a smaller proportion of BA recipients are more likely to be misclassified as unionized. School districts with a higher percentage of residents with 12 or more years of schooling are less likely to be classified as unionized regardless of true union status, and conversely, districts with a higher percentage of private enrollment have a higher probability of being classified as unionized regardless of true union status. Lastly, those districts with higher poverty and unemployment rates have a higher probability of being misclassified conditional on their true union status. The assumption necessary for classical measurement error that the error is independent of the correctly measured observables clearly does not hold in the data.

D-3.3 BBDR Decompositions

Since the misclassification error is correlated with both the dependent variables and the independent variables in the union impact regressions, it is interesting to determine the extent to which each of these correlations cause the observed differences in the estimated union effects. Bound, Brown, Duncan and Rodgers (1994) propose a decomposition of the difference between the biased coefficient and the unbiased coefficient into the difference directly due to measurement error and the difference due to the correlation of the measurement error with the regression error.⁵ More formally, let

$$Z = [U|X]' \tag{D-9}$$

be a matrix of all the data. Then

$$\begin{aligned} \tilde{\beta} &= (Z'Z)^{-1}Z'Y & \tag{D-10} \\ &= (Z'Z)^{-1}Z'[Z^*\beta + \epsilon] \\ &= (Z'Z)^{-1}Z'[(Z - \mu)\beta + \epsilon] \\ &= (Z'Z)^{-1}Z'Z\beta + (Z'Z)^{-1}Z'[-\mu\beta + \epsilon] \\ &= \beta - (Z'Z)^{-1}Z'\mu\beta + (Z'Z)^{-1}Z'\epsilon \end{aligned}$$

$$\begin{aligned} &\iff \\ \tilde{\beta} - \beta &= -(Z'Z)^{-1}Z'\mu\beta + (Z'Z)^{-1}Z'\epsilon & \tag{D-11} \\ &= -(E[\mu | U=1, X] - E[\mu | U=0, X])\beta + (E[\epsilon | U=1, X] \\ &\quad - E[\epsilon | U=0, X]), \end{aligned}$$

where the last line follows from the fact only union status is assumed to be measured with error in the data. The first term on the right-hand side of equation (D-11) gives the part of the total difference that is due to measurement error, while the second term shows the part of the total difference that is due to the correlation between the measurement error and the regression error. I perform this decomposition separately for each of the four dependent variables used above in a model that includes district fixed effects, year fixed effects and district-specific linear time trends. The coefficient estimates are thus identical to those reported in Tables D-3 through D-5.

Table D-8 presents the results of the BBDR decompositions. As is evident from the table, both forms of bias are present. These biases reinforce each other for log real teacher pay, log real current operating expenditures per student, and student-teacher ratios in this sample. The bias due to measurement error implies the direct effect of wrongly classifying a district as unionized is to increase the estimated union effect on teacher pay, expenditures per student, and student-teacher ratios. This result occurs because non-unionized districts have higher pay, expenditures, and class sizes than unionized districts, so mis-classifying non-unionized districts as unionized will

⁵See Black, Sanders and Taylor (2003) for an implementation of the BBDR decomposition similar to the one presented here.

bias upward the estimated impact of teachers' unions on all three measures. That the classification error is positively correlated with the regression error for the three inputs is due to the fact school districts incorrectly classified as unionized tend to have higher levels of teacher pay, expenditures, and student-teacher ratios than school districts for which union classification is correct. Thus, the misclassification of union status will serve to bias further upward the union impact estimates on these variables.

For the high school dropout rate decompositions, the biases offset each other somewhat, but the relatively large negative effect from measurement error dominates the positive correlation between the measurement error and the regression error. Non-union schools tend to have lower dropout rates than union schools, which is partially offset by the fact districts wrongly classified as unionized have higher dropout rates.

Table D-1: A Comparison of Union Status from the Census of Governments and the Union Election Certifications by State and Year

Census of Governments Union Measure						
1972						
Election Certification	Iowa		Indiana		Minnesota	
	Union	Non-Union	Union	Non-Union	Union	Non-Union
Union	0.00%	0.67%	3.63%	1.98%	53.58%	17.78%
Non-Union	5.99%	93.35%	10.23%	84.16%	19.40%	9.24%
1977						
Election Certification	Iowa		Indiana		Minnesota	
	Union	Non-Union	Union	Non-Union	Union	Non-Union
Union	49.89%	14.19%	57.43%	22.77%	55.89%	20.79%
Non-Union	9.31%	26.61%	11.55%	8.25%	16.17%	7.16%
1982						
Election Certification	Iowa		Indiana		Minnesota	
	Union	Non-Union	Union	Non-Union	Union	Non-Union
Union	51.22%	17.96%	55.12%	26.73%	58.20%	22.17%
Non-Union	8.43%	22.39%	8.58%	9.57%	13.16%	6.47%
1987						
Election Certification	Iowa		Indiana		Minnesota	
	Union	Non-Union	Union	Non-Union	Union	Non-Union
Union	46.78%	25.28%	42.24%	40.59%	64.67%	16.86%
Non-Union	9.76%	18.18%	6.60%	10.56%	13.63%	4.85%
Number of Districts	435		297		431	

Source: Author's calculations from the 1972, 1977, 1982, and 1987 Census of Governments and the teachers' union election certification data described in the text. The number of districts represents the total number of districts in the sample in 1987.

Table D-2: Misclassification Rates in the Census of Governments by State and Year, Treating the Election Certifications as the True Measure of Union Status

Year	Iowa	Indiana	Minnesota	Average
1972	6.65%	12.21%	37.18%	19.21%
1977	23.50%	34.32%	36.95%	31.17%
1982	26.39%	35.33%	35.33%	31.93%
1987	35.03%	47.19%	30.48%	36.48%

¹ Source: Author's calculations from the 1972, 1977, 1982, and 1987 Census of Governments and the teachers' union election certification data described in the text.

² The misclassification rate is the sum of the total number of times the Census of Governments and the election certification union measures disagree for each state and year. Each state-level misclassification rate is calculated by taking the sum of the off-diagonal entries from the appropriate four-cell square in Table D-1. The average misclassification rate is a weighted average of the state-level misclassification rates, where the weight is the number of school districts in each state and year.

Table D-3: Comparison of the Effect of Different Union Measures and Estimation Samples on Estimates of the Union Impact on Teacher Pay

Independent Variable	Dependent Variable: Ln(Real Monthly Average Teacher Pay)		
	COG/Hoxby (1996)		Election Certification
	(i)	(ii)	(iii)
	U.S. Estimation Sample	IA,IN,MN Estimation Sample	IA,IN,MN Estimation Sample
Union	0.051** (0.008)	0.054** (0.024)	-0.019 (0.072)
Log Population	-0.015** (0.004)	0.029 (0.063)	0.029 (0.057)
Percent Urban	0.0005** (0.0002)	0.0004 (0.0007)	0.0003 (0.0007)
Log Average Income	0.199** (0.022)	-0.066 (0.186)	-0.049 (0.187)
Log Median Rent	-0.021** (0.010)	0.064 (0.103)	0.078 (0.107)
Percent Below Poverty	-0.0001 (0.0006)	-0.009 (0.006)	-0.009 (0.007)
Percent Unemployed	-0.003** (0.001)	-0.009* (0.005)	-0.009* (0.005)
Percent Black	-0.004** (0.001)	0.0001 (0.010)	0.002 (0.009)
Percent Hispanic	-0.004** (0.001)	0.002 (0.010)	0.001 (0.011)
Percent 12–15 Years School	-0.002** (0.0003)	-0.002 (0.004)	-0.004 (0.004)
Percent 16+ Years School	0.004** (0.0004)	-0.005 (0.005)	-0.007 (0.005)
Percent Private Enrollment	0.001** (0.0002)	0.003 (0.004)	0.004 (0.004)
Log Public School Enrollment	0.041** (0.002)	-0.050 (0.057)	-0.053 (0.057)
R^2	NR	0.9366	0.9337

¹ Source: Estimates in column (i) come from Hoxby (1996) Table (IV) Column 6. Column (ii) contains estimates using the COG-based union measure on the IA, IN, and MN sample. Column (iii) presents estimates using the election certification union data on the IA, IN and MN sample. Estimates in columns (ii) and (iii) are calculated from the 1972, 1982, and 1987 Census of Governments, the 1970, 1980 and 1990 U.S. Census, and the teachers' union election certification data described in the text.

² Hoxby (1996) uses median household income, whereas I use mean household income because median household income is not included in the 1970 Census school district summary files.

³ All regressions include district and year fixed effects as well as district-specific linear time trends. Standard errors are clustered at the district level: ** indicates significance at the 5 percent level and * indicates significance at the 10 percent level.

Table D-4: Comparison of the Effect of Different Union Measures and Estimation Samples on Estimates of the Union Impact on Current Operating Expenditures Per Student

Independent Variable	Dependent Variable: Ln(Real Current Operating Expenditure Per Student)		
	COG/Hoxby (1996)		Election Certification
	(i)	(ii)	(iii)
	U.S. Estimation Sample	IA,IN,MN Estimation Sample	IA,IN,MN Estimation Sample
Union	0.029** (0.007)	0.017 (0.017)	-0.010 (0.064)
Log Population	0.029** (0.004)	0.013 (0.068)	0.013 (0.068)
Percent Urban	-0.001** (0.0001)	-0.0004 (0.0007)	-0.0004 (0.0007)
Log Average Income	0.116** (0.019)	0.146 (0.149)	0.151 (0.150)
Log Median Rent	0.232** (0.008)	-0.032 (0.105)	-0.027 (0.106)
Percent Below Poverty	-0.007 (0.001)	-0.009* (0.005)	-0.009* (0.005)
Percent Unemployed	-0.005** (0.001)	-0.006 (0.004)	-0.006 (0.005)
Percent Black	0.005** (0.001)	-0.005 (0.006)	-0.004 (0.005)
Percent Hispanic	0.003** (0.001)	-0.005 (0.008)	-0.005* (0.009)
Percent 12–15 Years School	0.005** (0.001)	-0.0001 (0.003)	-0.001 (0.003)
Percent 16+ Years School	0.004** (0.001)	-0.007 (0.005)	-0.008* (0.005)
Percent Private Enrollment	0.003** (0.001)	-0.001 (0.003)	-0.001 (0.003)
Log Public School Enrollment	-0.409** (0.011)	-0.024 (0.031)	-0.025 (0.030)
R^2	NR	0.9661	0.9660

¹ Source: Estimates in column (i) come from Hoxby (1996) Table (III) Column 6. Column (ii) contains estimates using the COG-based union measure on the IA, IN, and MN sample. Column (iii) presents estimates using the election certification union data on the IA, IN and MN sample. Estimates in columns (ii) and (iii) are calculated from the 1972, 1982, and 1987 Census of Governments, the 1970, 1980 and 1990 U.S. Census, and the teachers' union election certification data described in the text.

² Hoxby (1996) uses median household income, whereas I use mean household income because median household income is not included in the 1970 Census school district summary files.

³ All regressions include district and year fixed effects as well as district-specific linear time trends. Standard errors are clustered at the district level: ** indicates significance at the 5 percent level and * indicates significance at the 10 percent level.

Table D-5: Comparison of the Effect of Different Union Measures and Estimation Samples on Estimates of the Union Impact on Student-Teacher Ratios

Independent Variable	Dependent Variable: Student-Teacher Ratio		
	COG/Hoxby (1996)		Election Certification
	(i)	(ii)	(iii)
	U.S. Estimation Sample	IA,IN,MN Estimation Sample	IA,IN,MN Estimation Sample
Union	-1.112** (0.338)	0.117 (0.547)	-0.189 (0.836)
Log Population	-0.841** (0.071)	1.154 (2.618)	1.158 (2.604)
Percent Urban	0.029** (0.003)	0.020 (0.015)	0.020 (0.014)
Log Average Income	-1.170** (0.367)	-4.698 (4.100)	-4.660 (4.150)
Log Median Rent	-1.167** (0.161)	0.024 (2.519)	0.056 (2.483)
Percent Below Poverty	0.149 (0.012)	-0.038 (0.136)	-0.039 (0.137)
Percent Unemployed	0.123** (0.015)	-0.101 (0.103)	-0.102 (0.103)
Percent Black	-0.143** (0.012)	0.285 (0.478)	0.286 (0.455)
Percent Hispanic	-0.065** (0.014)	-0.162 (0.233)	-0.161 (0.225)
Percent 12–15 Years School	-0.129** (0.011)	0.071 (0.098)	0.067 (0.105)
Percent 16+ Years School	-0.082** (0.015)	0.165 (0.154)	0.162 (0.161)
Percent Private Enrollment	-0.098** (0.009)	-0.250 (0.344)	-0.249 (0.345)
Log Public School Enrollment	7.334** (0.217)	-2.990 (4.769)	-2.997 (4.745)
R^2	NR	0.9612	0.9612

¹ Source: Estimates in column (i) come from Hoxby (1996) Table (V) Column 6. Column (ii) contains estimates using the COG-based union measure on the IA, IN, and MN sample. Column (iii) presents estimates using the election certification union data on the IA, IN and MN sample. Estimates in columns (ii) and (iii) are calculated from the 1972, 1982, and 1987 Census of Governments, the 1970, 1980 and 1990 U.S. Census, and the teachers' union election certification data described in the text.

² Hoxby (1996) uses median household income, whereas I use mean household income because median household income is not included in the 1970 Census school district summary files.

³ All regressions include district and year fixed effects as well as district-specific linear time trends. Standard errors are clustered at the district level: ** indicates significance at the 5 percent level and * indicates significance at the 10 percent level.

Table D-6: Tests of Non-Differentiation in the COG Union Measure Classification Error

Independent Variable	P(U=1 U*=1,Y)	P(U=0 U*=0,Y)
Log Real Teacher Pay	0.356** (0.051)	-0.076 (0.064)
Log Real Expenditures per Student	-0.002 (0.015)	0.024** (0.004)
Student-Teacher Ratio	0.017** (0.004)	0.007** (0.002)
High School Dropout Rate	-0.0003 (0.001)	0.004** (0.001)

¹ Source: Author's calculations from the 1972, 1977, 1982, and 1987 Census of Governments and the teachers' union election certification data described in the text.

² Each cell above represents a separate pooled linear probability model regression. Standard errors are in parentheses: ** indicates significance at the 5 percent level.

Table D-7: Relationship Between Misclassification in the COG Union Measure and the Observables

Independent Variable	P(U=1 U*=1,X)	P(U=0 U*=0,X)
Log Population	0.035** (0.010)	0.009 (0.013)
Percent Urban	0.067** (0.036)	0.026 (0.041)
Log Average Income	-0.007 (0.022)	-0.042** (0.007)
Log Median Rent	-0.041 (0.036)	-0.126** (0.024)
Percent Poverty	-0.755** (0.268)	-0.570** (0.224)
Percent Unemployed	-0.002** (0.001)	-0.008** (0.001)
Percent Black	-0.001 (0.003)	0.022* (0.012)
Percent Hispanic	-0.07 (0.006)	0.026 (0.020)
Percent 12–15 Years School	-0.001 (0.001)	0.008** (0.002)
Percent 16+ Years School	0.001 (0.001)	-0.010** (0.002)
Percent Private Enrollment	0.004** (0.001)	-0.003** (0.001)
Log Public School Enrollment	0.039** (0.010)	0.010 (0.013)

¹ Source: Author’s calculations from the 1972, 1982, and 1987 Census of Governments, the 1970, 1980 and 1990 U.S. Census, and the teachers’ union election certification data described in the text.

² Each cell above represents a separate pooled linear probability model regression. Standard errors are in parentheses: ** indicates significance at the 5 percent level and * indicates significance at the 10 percent level.

Table D-8: BBDR Decompositions

Dependent Variable	COG Estimate	Union Election Estimate	Total Difference	Difference From Measurement Error	Difference From Correlation of Measurement Error and Regression Error
Log Real Teacher Pay	0.054	-0.019	0.073	0.018	0.054
Log Real Expenditures per Student	0.017	-0.010	0.027	0.010	0.017
Student-Teacher Ratio	0.117	-0.189	0.306	0.183	0.124
High School Dropout Rate	0.589	1.385	-0.796	-1.332	0.536

¹ Source: Author’s calculations from the 1972, 1982, and 1987 Census of Governments, the 1970, 1980 and 1990 U.S. Census, and the teachers’ union election certification data described in the text.

² Each regression includes district and year fixed effects as well as district-specific linear time trends.