Analysis of Regression and Surveys in Ohio LSC Report on S.B. 102 on Claimed Cost Savings from Exempting School Construction from Prevailing Wage Requirements

Herbert F. Weisberg

Professor, Ohio State University

July 8, 2002

1. Summary and Overview

On May 20, 2002 the Ohio Legislative Service Commission (LSC) issued Staff Research Report #149 claiming $489 million of cost savings since S.B. 102 took effect in August 1997 exempting school construction from the state's prevailing wage requirements. They used data from F.W. Dodge, a company that collects information on construction projects, including their bid prices. This cost-savings estimate is based on a statistical procedure known as regression analysis, but examination of the statistical analysis in the LSC Report shows that estimate is not valid.

The main statistical problem is that the LSC Report's regression equations explain a miniscule portion of the differences in costs between projects. Regression estimates would be meaningful only if the equations account for 70+% of the differences between projects, but these account for only a trivial 1% to 3% of the cost differences. The regression equations do not fit the data, so the cost-savings estimates are statistical fiction.

Additionally, the LSC equations find prevailing wage to be statistically insignificant, meaning that there is no statistical reason to believe that prevailing wage affects costs. A small cost savings might not be found significant, but it is not reasonable to claim that an effect leading to nearly $500 million in cost savings would not be found significant if it were real. In fact, every preceding analysis of Dodge construction data for Ohio and other states has found that prevailing wage does not significantly increase costs, and the LSC Report actually confirms that finding.

Finally, the Dodge data that the LSC analyzed show only the construction costs at the start of projects. They do not show the final construction costs, which can be considerably higher if the company lacks the expertise to keep the costs within the level of the bid. Therefore, the estimated cost savings are not relevant to actual project costs.

All in all, LSC's claimed cost savings obtained by exempting school construction projects from Ohio's prevailing wage law are based on flawed interpretations of statistical analysis.

This analysis will focus on the regression of the Dodge construction data. Regression analysis is briefly explained in Section 2. Section 3 examines the analysis of new construction and additions in Appendix 2. The alterations analysis in Appendix 2 is described separately in Section 4, because it builds on the regression analysis of new construction and additions. Section 5 briefly considers the discussion of omitted variables in Appendix 5. The several surveys in the LSC Report are examined in Section 6. Finally, some conclusions are presented in Section 7.

2. Regression Analysis

Regression analysis is a completely standard statistical technique for ascertaining whether possible explanatory variables account for a variable of interest. It has been used to look at the effect of Prevailing Wage legislation since an analysis by Fraundorf (1983) of the Davis-Bacon Act. It has been used on the Dodge construction data in several past studies, including Prus (1996, 1999), Philips (2001), and Wial (1999).

To understand regression, imagine a graph with a lot of points in it and think about trying to draw in the straight line that best fits those points. That's what regression does. Say hypothetically that construction costs were $150 a square foot one year, $155 the next year, $160 the next year, and $165 the next year. If we graphed that, we'd find a very simple straight line relationship: costs = $150 + $5 more per year. That's a regression equation. The $150 is what the regression tables in the LSC Report call an "intercept" and the $5 would appear as the "coefficient" for year.

Real data, of course, are always more messy. Say instead that construction costs were $150 a square foot one year, $154 the next year, $161 the next, and $165 the next year. The best-fitting line to these values would still probably be the one found above: costs = $150 + $5 more per year. That line would still provide a good fit to the data, but there would be some "error" because that line no longer fits the data perfectly. The R-Square values in the regression tables show how good the fit is: 1.00 means that the fit is perfect, as in the previous paragraph, and the example in this paragraph would still give a very high R-Square, but it could go down as low as zero if the data do not fit a straight line at all.

Regression analysis is called "linear" because it is looking for a straight-line pattern, as in the above paragraphs. If the construction costs first climbed regularly but then declined regularly, it would not find any particular relationship because there is no longer a straight line pattern to the data.

Next, imagine adding some other considerations to the equation. Say we are explaining costs pretty well with a year variable, but maybe adding in unemployment rates as an additional explanatory variable would help account for the part of costs that the year wasn't accounting for. We can keep adding in explanatory variables to try to account for the differences in costs between projects better and better. And the R-Square value tracks for us how well we are doing in the explanation. (The adjusted R-Square is actually the best statistic to use since it adjusts for the statistical effect that adding more explanatory variables should improve the explanation.)

Going back to our equation above, we could plug the year number into the equation to get an estimated cost value for that year. For example, the estimated cost for the second year would be $155, and the R-square shows how good that estimate is.

When the R-square is high, near 1.00, that will be a good estimate. However, the estimated cost would be a poor estimate if the R-square were only .01, since that means the data do not fit the regression line. The cost-savings estimates in the LSC Report are based on plugging in values into regression equations in this manner. However, the adjusted R-squares are only in the order of .01 to .03, so the estimates of cost savings will not be accurate.

3. Dodge Construction Data for New Construction and Additions

Dodge construction data for new school construction and additions are analyzed in Appendix 2 of the LSC Report using regression analysis. Prevailing wage legislation is found to be statistically insignificant in this analysis, meaning that it did not increase project costs. Also, the analysis does not succeed in accounting for the differences between projects in their costs.

3.1 Data. The F.W. Dodge construction data seems to be the gold standard in the field, being used in several past regression analyses on prevailing wage effects (e.g. Prus 1996, 1999; Philips 2001; Wial 1999). The Ohio data used here are for 1992-2001, covering years before and after the passage of S.B. 102.

The LSC Report uses regression analysis to look at the factors that affect "inflation-adjusted cost per square foot" ($SQFT), which is the "dependent variable" in this analysis. In regression analysis, the analyst checks which of several potential explanatory factors ("independent variables") affect that dependent variable. In the 3 regression analyses in Tables 20, 21, and 22, the LSC Report uses 4 sets of explanatory variables: 1) whether the project was undertaken before September 1997 when school construction was subject to prevailing wage requirements (PW), 2) whether the school was in a rural county (Rural), 3) the type of school (Primary School, Junior High, Secondary High, or Vocational), and 4) a time counter for when the construction occurred (Trend). Additionally, it uses an "interaction term" (PW-Rural) to test whether the prevailing wage law had a different effect in urban and rural counties. (The Appendix refers to some of these as "dummy variables" -- all that means is that they are two category variables, such as the county either being rural or not being rural.) (The tables have rows only for Junior High, Senior High, and Vocational School, not for Primary School -- that is appropriate; it just means that Primary School is being used as the baseline to find out how much more (or less) expensive it is to build the other types of schools.) (The time counter is a variable that is set at 1 for the first month in the data, January 1992, 2 for the next month, February 1992, and so on up to 120 for December 2001. It does not take into consideration the possibility of seasonal differences.)

It is important to emphasize that the Dodge data refer only to accepted bids for projects. They do not show the actual cost of the construction. Obviously the actual cost can be much higher than the original bid, especially if the low bidder is inexperienced in keeping costs within the level of the bid. Therefore, the Dodge data can never be used to show actual cost saving.

3.2 Explanatory Strength of the Regression. Tables 20-22 contain an important summary statistic showing the statistical quality of the regressions: the R-Square value (or, better yet, the Adjusted R-Square). This shows how well the regression accounts statistically for the differences in costs between different projects (known as the "proportion of variance that is explained"). This statistic can range from 0.00 (for a regression that is so useless that it does not account for any of the differences) to 1.00 (for a regression that totally accounts for the differences). (The adjusted R-square just takes into account that adding more explanatory variables inevitably increases the chances for accounting for differences.)

These statistics show that the regressions in Tables 20-22 account for only 1-3% of the differences in project costs. This is an exceedingly low level of statistical explanation. When the regressions are this useless in accounting for the observed differences, it is appropriate to wonder what went wrong -- if the wrong model was used or if relevant explanatory variables were omitted.

The low level of the R-square shows that there is little tendency for this straight-line model to fit the data. Instead, it is likely that there remains considerable differences in cost-per-square-foot within each combination of categories on the explanatory variables. The regression equation does not satisfactorily fit the data.

These R-square statistics show the regression analyses for new construction and additions were useless. Prevailing wage legislation does not account for the differences found in costs for new construction or additions. The LSC Report argument on page 60 against using statistical significance tests does not address this problem at all. The regression equations show that 97-99% of the differences in costs for new construction and additions remain unexplained.

The LSC Report claims cost savings of $487.9 million in the post-exemption period, of which $408.0 million (84% of the claimed cost savings) are from additions projects. However, the adjusted R-square for the additions regression is only .01, meaning that 99% of the differences between additions projects are not being accounted for by the regression. As a result, there is no statistically valid basis for estimating cost savings.

3.3 Statistical Significance. The three right-hand columns of Tables 20-22 include material designed to show whether each of the explanatory variables has a "significant" effect. The usual convention is to require the right-hand column value (the "P-value" for "probability value") to be less than .05 for the result to be considered meaningful. A value less than .05 would mean that there is less than a 5% chance of getting the obtained regression effect by chance alone, and most scientific fields consider that an appropriate standard. (Some fields would instead require a more stringent .01 level, while exploratory work sometimes allows a more lenient .10 level.) Prevailing wage does not have a statistically significant effect in Table 20, 21, and 22, nor is its interaction with rural significant in any of those tables.

Recall Gertrude Stein's wonderful line about Oakland: "there's no there there." Statistical significance tests are designed to test whether there is any there there -- and there isn't any there here! The statistical test shows that concluding that prevailing wage has an effect on cost is incorrect.

Significance tests are standard in the scientific literature. As one example, they are used in the medical field to decide whether a claimed effect of a new disease treatment is greater than would have been expected by chance. A new treatment would not be accepted if its effect were found not to be statistically significant. As another example, I would not be able to publish a result in a journal in my field if the result was not statistically significant. The regression results do not show that prevailing wage increased cost for new construction or additions.

Significance tests are designed to test whether small effects could have occurred by chance alone. Some of the claimed effects shown here for prevailing wage are large, particularly the effects on additions. It would be unreasonable to claim that a large effect leading to a $408 million would not be found significant.

The LSC Report adds two more tables, Tables 24 and 25, to Appendix 2 to summarize the probabilities of the results for each explanatory variable in Tables 20, 21, and 22. The last column of Table 24 then shows the minimum of the probabilities. The implication is that it is appropriate to look at the best statistical result across several separate regression equations. This is totally invalid. I have never seen such a use of a minimum of probabilities. And it is totally misguided. If one were using the .05 level, the chance of at least one of three regression equation finding an effect of a variable at the .05 level is .1526 (using a standard "binomial" logic). Thus, there is a .15 chance that at least one of those three regressions would have found a significant prevailing wage effect, so it is even more telling that it was not significant in any of the 3 regressions.

The LSC Report gets around the significance issue with an argument on page 60 that the data should be considered full data for a population instead of data based on a sample of school construction projects. That is one possible argument that is sometimes given for not using significance tests. However, if the LSC Report did not consider significance calculations appropriate, there would have been no reason to include the right-most three columns in Tables 20-23. Additionally, the standard argument in the statistical literature for using significance tests even in this situation is that the construction projects that were undertaken can be considered to be a sample of those that could have been conducted. To repeat, regardless of the implication of the LSC Report, prevailing wage did not have a statistically discernable effect on school construction costs for new construction or additions.

That the effects of the Prevailing Wage are not significant in regression analysis should not be surprising since other regression analysis of Dodge data by Prus (1996), Prus (1999), Philips (1999), Bilginsoy and Philips (2002), and Philips (2001), reviewed on pages 14-15 of the LSC Report, all find the same thing -- prevailing wage does not have a statistically significant effect on school construction costs.

Finding an insignificant effect with a small sample is sometimes excused when the probability is .06 rather than .05, with the effect being described as "marginally significant." Here, however, the probability levels are all .52, .23, and .34, far above .05. Furthermore, the number of projects for Tables 20-22 are not small -- an effect that is estimated to be in the order of many millions of dollars should be detectable when dealing with regressions of 256, 194, and 676 projects. Finding insignificant results here shows that Prevailing Wage simply does not matter.

3.4 Regression Coefficients. Regression analysis permits the writing of an equation estimating the dependent variable on the basis of the explanatory variables. The regression "coefficients" are used to construct the regression equation. The coefficients for Tables 20-23 are summarized in the table below, with significant effects marked by asterisks.

	New Construction -- Large Projects	New Construction -- Small Projects	Additions
Trend	.14	-.14	1.54*
Rural	.98	-14.49	10.42
Junior High	6.78	.96	80.37*
Senior High	1.52	-2.00	10.06
Vocational School	15.17	9.18	-43.18
Prevailing Wage	3.99	-11.45	46.47
Prevailing Wage for Rural Counties	-5.54	5.50	8.73

First, there are few significant explanatory variables. Only 2 of the 21 numbers in the table are statistically significant. At the .05 significance level, one would expect 1 of 20 values to be significant by chance alone, so the results are basically at chance levels. Second, the coefficients in most rows bounce around considerably. Trend has a large effect for additions, a trivial positive effect for large new constructions and a trivial negative effect for small new constructions. Rural has a large negative effect for small new construction, a large positive effect for additions, and a small positive effect for large new construction; senior high has the same pattern. Junior high has a very large effect for additions, but small effects for new construction; prevailing wage has the same pattern. Vocational schools has a large negative effect for additions, but small positive effects for new construction. It is possible that these differences reflect differences between the different types of projects, but no rationale is given in the LSC Report for viewing the instability of these coefficients as plausible. However, inconsistent patterns like those in this table are usually an indication of the effects not being real. Random effects would be expected to bounce around in the same manner that these do. The effects of the variables appear to be random, so basing cost-savings estimates on them is risky at best.

3.5 Cost-Savings Estimates. The estimated savings from S.B. 102 claimed in the LSC Report are based on the regression analysis. The equations in the coefficient columns of Tables 20 and 21 are used for new construction and the coefficient column of

Table 22 for additions. The equations are estimated under Prevailing Wage and without Prevailing Wage, and the difference is taken as the estimated savings.

Using regression analysis to estimate values is standard if the regression equation has a good fit to the data. However, the R-Squares show that these equations do not fit the data, so estimated values based on them are worthless. The actual estimates are based on regressions accounting for only 1-3% of the differences in observed cost-per-square-foot, which is not a reasonable level for valid estimates of the cost savings.

As an example of how this works, the equation for large new construction projects from Table 20 (using the "coefficients" column) estimates that the cost-per-square-foot for additions (in inflation-corrected dollars) as:

86.64 + (.14*time indicator) + (.98 if the county was rural) + (6.78 for a junior high, 1.52 for a senior high, and 15.17 for a vocational school) + (3.99 when prevailing wage was in effect) + (-5.54 in rural counties when prevailing wage legislation was in effect).

Evaluating this equation for large new construction projects, the estimated cost-per-square foot (inflated to Dec 2001 dollars) for primary schools in Sept 1997 (the month when S.B. 102 took effect) would be:

	With Prevailing Wage	Without Prevailing Wage	Claimed Effect of Prevailing Wage
Urban	$100.29	$96.73	$3.99
Rural	$95.73	$97.28	-$1.55

Each of those cost figures would be $6.78 higher for a junior high, $1.52 higher for a senior high, and $15.17 higher for a vocational school. Each of those estimates would have been 14 cents lower the previous month and 14 cents higher the next month. To repeat what has been said earlier, the prevailing wage effect in this equation is not statistically significant, and the regression on which these estimates are based accounts for only 3 percent of the differences in costs between these projects, which shows that these estimates are not statistically valid. And, of course, these are costs according to accepted bids, not the final project costs, which could run higher.

The LSC analysis would have plugged in the characteristics of each of the 256 large new construction projects into this equation, and estimated the cost with and without the prevailing wage, and then summed that over all the projects to get a cost-savings estimate.

Similarly, from Table 21, the estimated cost-per-square foot (in Dec. 2001 dollars) for primary schools in September 1997 for small new construction projects would be:

	With Prevailing Wage	Without Prevailing Wage	Claimed Effect of Prevailing Wage
Urban	$85.39	$96.84	$11.45
Rural	$76.40	$82.35	$5.95

These figures would be $.96 higher for junior high schools, $2.00 lower for senior highs, and $9.18 more for vocational schools. These estimates would be 14 cents higher the previous month and 14 cents lower the next month. The prevailing wage actually reduces project costs according to this equation, which the LSC Report does not point out. Again, this regression accounts for only 1 percent of the differences in costs between these projects, so these estimates are not statistically valid.

The total of $24.6 million for new construction cost savings is based on the figures above: $3.99 saving per square foot for urban large projects, -$1.55 (negative) for rural large projects, $11.45 for urban small projects, and $5.95 for rural small projects, each multiplied by the total number of square feet of projects of those types.

From Table 22, the estimated cost-per-square-foot (stated in Dec. 2001 dollars) for additions projects for September 1997 would be:

	With Prevailing Wage	Without Prevailing Wage	Claimed Effect of Prevailing Wage
Urban	$181.61	$135.14	$46.47
Rural	$200.76	$145.56	$55.20

These figures would be $80.37 higher for junior highs, $10.06 more for senior highs, and $43.18 less for vocational schools. The estimated costs would be $1.54 less each for August 1997 and $1.54 more each for October 1997. Again, the Prevailing Wage term is not statistically significant in this equation, and the equation accounts for only 1 percent of the differences in project bids, so that these estimates are not statistically valid.

The total of $408 million in cost savings claimed in the LSC Report for the 676 additions are based on these figures: multiplying the total square footage of urban projects by $46.47 and the total square footage of rural projects by $55.20 and then summing those values. However, this estimate is totally based on an invalid equation with an adjusted R-square of only .01. Again, these are not actual costs, as they are based on the accepted bids rather than the final project costs.

4. Dodge Construction Data for Alterations

Table 23 in the LSC Report is used to generate the cost-savings figures for alterations projects. However, it does not actually analyze alterations projects! This analysis differs from the rest because, as clearly stated in Appendix 2 to the LSC Report, the nonavailability of square-footage for alterations projects made it impossible to analyze the cost-per-square-foot as in the analysis for new construction and additions. Instead, regardless of how it is labelled, Table 23 is not a regression analysis of alterations as it is labeled. Instead, Table 23 reanalyzes the new construction and additions project to obtain an equation that the LSC Report uses to estimate cost savings for alterations.

The regression analysis of alterations in Table 23 is based on just combining the new construction and additions data. Page 58 explains this: "the alteration subset was analyzed using the estimated percentage saving by project *for the new and additions data subsets. The two subsets were combined, and a regression was run with estimated percentage savings as the dependent variable"* (emphasis added). [This can be substantiated by noticing that the number of projects in Table 23 (1126 observations) exactly equals the sum of the numbers for Tables 20 (256), 21 (194), and 22 (676).] Thus, Table 23 is a regression analyzing new construction and additions that is used to estimate the savings for alterations.

The regression analysis in Table 23 is intended to examine differences between projects in their "estimated percentage savings due to the absence of a prevailing wage," with that estimated percentage being based on the previous regressions. The cost per square foot for each project without the prevailing wage is estimated from the applicable regression in Tables 20, 21, or 22, based on when the project occurred ("Trend"), if it was in a Rural county, and the type of school (primary, junior high, senior high, or vocational). Next, an estimate is obtained for the cost with the prevailing wage by adding the PW coefficient in the corresponding table (for example, 3.99 for new construction-large projects in urban areas). Then these two estimated cost figures are compared to determine estimated cost savings. Hypothetically, if the cost per square foot without the prevailing wage for a project were estimated to be $135 and the cost with the prevailing wage estimated at $150, the "estimated percentage savings due to the absence of a prevailing wage" examined as the dependent variable in Table 23 for that project should be ($135-$150)/$150 = 10% savings. And if instead the cost without the prevailing wage were estimated to be $165 and the cost with the prevailing wage were estimated to be $150, the value for that project that is used in Table 23 should be ($165-$150)/$150 = 10% decreased savings.

This dependent variable for Table 23 is very shaky. It is based on regression equations in Tables 20-22 that account for only 1-3 of the differences in costs-per-square-foot and in which the prevailing wage is always statistically insignificant. The analysis in Table 23 is only as good as the regression estimates on which the dependent variable is based, and they are terribly poor estimates.

The explanatory variables used in the regression in Table 23 are 1) whether the school was in a rural county (Rural), 2) the type of school (Primary School, Junior High, Secondary High, or Vocational), 3) a time counter for when the construction occurred (Trend), and 4) the inflation-adjusted General Contract Value using Dodge data on General Contract Value and *Engineering News Record* data on inflation for construction cost and building cost ("ENR Value").

Table 23 reports that this regression accounts for 13% of the differences in the dependent variable (the adjusted R-Square value). How could it do even this well when it is based on regressions that are trivial? The use of General Contract Value as a predictor could be partly responsible. Also, because it combines new construction large projects, new construction small projects, and additions, this analysis is getting its explanatory power from the differences between those three different types of projects. Prevailing Wage has more of an effect on cost for additions in Table 22 than for new construction in Tables 20-21, so the regression pattern in Table 23 could be obtained if there were more additions for senior highs and vocational schools and fewer for junior highs as well as fewer additions for rural counties than urban counties. (The data presentation in Appendix 2 does not permit a check as to whether that supposition is correct.) Table 23 reflects the differences between different types of schools in additions versus new construction, not any observed differences in actual costs.

Since the square-footage of additions projects is not in the Dodge data that the LSC Report analyzed, they could not conduct a regression analysis of alterations in the same manner that Table 23 was generated. Specifically, Tables 20-22 could not be used to estimate project savings with Prevailing Wage as was done for the dependent variable in Table 23. Instead, the LSC used the equation in Table 23, based on new construction and additions, to estimate the project savings for alterations. The coefficient column in Table 23 gives an equation:

Estimated Project Savings = -.251916 + (.000004 * General contract value, inflation adjusted) + (.001496 * time indicator) + (.005441 if the project is in a rural county) + (.026332 if it is a junior high school) - (.067186 if it is a senior high school) - (.089969 if it is a vocational school).

For each alterations project, the general contract value, the month of the bidding, whether it was in a rural county, and the type of school are plugged into this equation, to obtain an estimated projected savings. This means that the estimated project savings for alterations projects are estimated on the basis of the savings found in the preceding analysis for new construction and additions, where the effects of prevailing wages were always not statistically significant. Thus, the analysis in Table 23 is as sturdy and reliable as would be constructing a school building out of balsa wood on quicksand! Since the regressions in Tables 20-22 accounted for only a trivial 1-3% of the differences in project values and since prevailing wage was never statistically significant in those equations, the dependent variable in Table 23 is invalid and using that regression to estimate effects for alterations is doubly invalid.

To understand this equation in terms comparable to those used in section 3 above, for an urban primary school alterations project in September 1997, it estimates a cost savings in percentage terms of: -14.8692% + .0004*the General Contract Value (ENR in thousands of dollars) of the project. Table 38 shows that the General Contract Value for all urban alterations projects for 1997 was $38.9 million, for an average of $437,079 per project, so let's say a typical project is about $400,000. For a $400,000 project, this equation would estimate a savings of 14.71%. The savings would be .54% less for a rural primary project: 14.17%. The savings would be 2.63% less for an urban junior high project: 12.07%. The savings would be 6.72% more for an urban senior high project: 21.43%. And the highest savings would be an urban vocational school project: 23.71%. But, the trend variable means that these savings would go away over time! For example, by December 2001, the estimated savings for the urban primary school would be down to 7.08%. If this trend line were projected forward, by December 2008, for example, the urban primary school would be 1.90% more expensive because of the removal of prevailing wage legislation. Indeed, urban and rural primary and junior high $400,000 alterations projects would all be expected to cost more because of the removal of prevailing wage legislation well before December 2008.

The illustrations in the above paragraph are meant to demonstrate how the formula for alterations projects works. However, the basic point is still that the formula is not statistically valid. All the alterations estimates are based on regression equations for new construction and additions from Tables 20-22 in which prevailing wage was not statistically significant and in which only a trivial 1-3% of differences in project costs were being accounted for statistically.

5. The Discussion of Omitted Variables

Appendix 5 in the LSC Report provides a brief example of how an omitted variable can alter the results of a regression analysis. This seems to imply that the cost savings from Prevailing Wage are higher than estimated in Appendix 2, since taking into account whether the project received funding from the School Facilities Commission (SFC) would increase the effect found for the Prevailing Wage on cost-per-square foot of new construction-large projects. (The School Facilities Commission funding variable was not included in the main analysis in Appendix 2 because it is not coded with full accuracy -- they tried to match the SFC funding to the projects in the Dodge data, but a perfect match could not be made.)

However, the technical details of the new regression in Table 45 are not reported in an appropriate manner that permits an assessment of whether this regression provides statistically valid information. Even so, Table 45 does not report the new regression fully enough to tell if the inclusion of the formerly omitted variable makes a difference. The statistical significance of the explanatory variables is not reported in Table 45. It is not clear whether SFC is significant, nor is it clear whether the higher value found for Prevailing Wage is significant. And the R-squared is not reported, so the overall explanatory power of the new regression cannot be assessed.

6. Surveys Analyzed in the LSC Report

While the main focus of my report is on the regression analysis in the LSC Report, the surveys included in the Report also raise serious concerns.

6.1 <u>Surveys of School Districts</u>. On page 27, the Report measures quality of construction by asking districts about the quality of school construction before and after the exemption. However, several of the responses reprinted in the Report clearly tell more about the respondents' preconceived opinions on the prevailing wage than about the actual quality of the work. The LSC Report justifies this by referring to a Building Research Board report that indicated that quality depends on "one's point of view" and emphasized the importance of "conformance to adequately developed requirements" and "satisfaction of user's needs." However, measuring quality validly, even under this definition, would therefore require separate questions that directly ask about more specific parts of the process in a manner that would obtain objective replies. Additionally, quality cannot be fully assessed over a short period of time, since the goal for the construction is to be of high enough quality to be useful for a long term.

I am also concerned about the low response rates to the surveys. The January 1999 survey of district superintendents received replies from only 31 percent of the 611 districts and the August 2000 survey received answers from only 58 percent of the districts. The lack of response can bias the survey results. I am surprised that the LSC Report does not indicate the distribution of the responding districts around the state or even what proportion of the districts with funding from the School Facilities Commission responded to the survey.

6.2 Surveys of Contractors. The contractor surveys ask contractors to state what their bid prices would have been under prevailing wages. Not only is it difficult for people to answer hypothetical questions, but, as the LSC Report admits on page 18, non-union contractors have an incentive to overstate the prevailing wage price.

Additionally, there is no indication given of the response rate to this survey. Having a total of 774 responses to 3 waves of this survey (Table 2, page 19) strikes me as very low given the large number of districts and the likelihood of several contractors per district. The LSC Report acknowledges on page 18 that "many school districts and companies instead chose to not participate in our exploratory survey" but justifies the analysis of them because these surveys were intended "to narrow the range of the possible savings that may result from the exemption." However, it is not clear to me how reports from an unrepresentative set of respondents can narrow the range of savings.

6.3 Surveys on Construction Wages. Appendix 4 of the LSC Report analyzes Current Population survey data on wages. As the LSC Report admits on page 68, this survey is meant to represent the national population, so "the data obtained is not a representative sample of Ohio construction workers." The data from large-scale surveys of this type are usually not broken down into as small categories as here. Table 44 on page 74, for example, shows that several of the wage statistics in the preceding tables were based on just 1 or 2 respondents. The largest claimed gain in hourly pay rates in Table 40 is for glaziers, increasing 156.5%, but Table 44 shows that is based on comparing the wages for one non-union glazier before the exemption to one union glazier after the exemption. Sample surveys are designed to permit generalizations to larger populations, but certainly not from data for a single person. Averages are generally not computed in statistical reports when they are based on less than 10 or 20 instances, because they are susceptible to being thrown off by atypical cases. Furthermore, statistical significance tests are usually performed to make sure that observed differences between categories are greater than would be expected given the differences observed within categories, but no significance tests are reported here.

7. Summary and Conclusions

The estimated savings from S.B. 102 are based on the regression analysis. The equations in the coefficient columns of Tables 20 and 21 are used for new construction and the coefficient column of Table 22 for additions. The equations are estimated under Prevailing Wage and without Prevailing Wage, and the difference is taken as the estimated savings. Table 23 is used to estimate savings for alterations.

Using regression analysis in to estimate values is standard if the regression equation has a good fit to the data. However, the R-Square shows that these equations do not fit the data, so estimated values based on them are worthless. The actual estimates are based on regressions accounting for only 1-3% of the differences in observed cost-per-square-foot, which is not a reasonable level for valid estimates of the cost savings. Furthermore, the Prevailing Wage effects are not statistically significant in any of the reported tables. All estimates of "cost savings" are thus based on faulty use of statistical procedures. The alterations analysis is even more shaky, as it assumes the same process underlies cost savings on those projects as on other types of construction projects.

References

Bilginsoy, Cihan, and Peter Philips. 2000. "Prevailing Wage Regulations and School Construction Costs: Evidence from British Columbia," *Journal of Education Finance*, 24: 415-32.

Fraundorf, Martha Norby, John Farrell, and Robert Mason. 1983. "The Effects of the Davis-Bacon Act on Construction Costs in Rural Areas," *The Review of Economics and Statistics*, 66:142-46.

Philips, Peter, Garth Mangum, Norm Waitzman, and Anne Yeagle. 1995. "Losing Ground: Lessons from the Repeal of Nine 'Little Davis-Bacon' Acts." University of Utah.

Philips, Peter. 1999. "Kentucky's Prevailing Wage Law: Its History, Purpose, and Effect."

Philips, Peter. 2001. "A Comparison on Public School Construction Costs: In Three Midwestern States that Have Changed Their Prevailing Wage Laws in the 1990s." University of Utah.

Prus, Mark. 1996. "The Effect of State Prevailing Wage Laws on Total Construction Costs."

Prus, Mark. 1999. "Prevailing Wage Laws and School Construction Costs: An Analysis of Public School Construction in Maryland and the Mid Atlantic States."

Wial, Howard. 1999. "Do Lower Prevailing Wages Reduce Public Construction Costs?" Harrisburg, PA: Keystone Research Center.