NBER WORKING PAPER SERIES

THE EFFECT OF UNIONS ON PRODUCTIVITY IN THE PUBLIC SECTOR: THE CASE OF LIBRARIES

Ronald G. Ehrenberg

Joshua L. Schwarz

Working Paper No. 717

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge MA 02138

July 1981

The research is supported by National Science Foundation Grant No. DAR 80-14279. The research reported here is part of the NBER's research program in Labor Studies. Any opinions expressed are those of the authors and not those of the National Bureau of Economic Research.

NBER Working Paper #717
July 1981

The Effect of Unions on Productivity in the Public Sector: The Case of Libraries

ABSTRACT

This paper presents an analytical framework that can be used to analyze the effects of unions on productivity in the public sector. Our initial focus is on public libraries because considerable effort has been devoted to conceptualizing library productivity measures and because of the availability of data to implement the framework. Preliminary estimates are presented based upon data from 71 municipal libraries in Massachusetts. We conclude by indicating the direction that our future research on the subject will take.

Ronald G. Ehrenberg
Joshua L. Schwarz
New York State School of
Industrial and Labor Relations
Cornell University
Ithaca, New York 14853

(607) 256-3026

I. Introduction

Research by economists on the economic effects of unions in the private sector has tended to focus on unions' effects on their members' relative earnings positions. Following in the tradition of H.G. Lewis' pioneering work, a large number of studies have addressed this question.[1] The more sophisticated ones use micro-level data and seek to control both for quality differentials and the possibility that wages and union status may be simultaneously determined.[2] Most recently, other nonwage outcomes, such as job satisfaction and labor turnover, have been considered and analysts have attempted to ascertain if part of any observed union/nonunion wage differential merely compensates unionized employees for relatively unfavorable nonpecuniary conditions of employment.[3]

---

[1] H.G. Lewis, Unionism and Relative Wages in the United States (Chicago, University of Chicago Press, 1963).

[2] See Peter Schmidt and Robert Strauss, "The Effect of Unions on Earnings and Earnings on Unions: A Mixed Logit Approach," International Economic Review (February 1976), Lung-Fei Lee, "Unionism and Wage Rates: A Simultaneous Equations Model With Qualitative and Limited Dependent Variables," International Economic Review (June 1978) and Peter Schmidt, "Estimation of a Simultaneous Equations Model With Jointly Dependent Continuous and Qualitative Variables: The Union-Earnings Question Revisited," International Economic Review (June 1978).

[3] See Gregg Duncan and Frank Stafford, "Do Union Members Receive Compensating Wage Differentials," American Economic Review (June 1980); George Borjas, "Job Satisfaction, Wage and Unions," Journal of Human Resources (Winter 1979); Richard Freeman, "Individual Mobility and Collective Voice in the Labor Market," American Economic Review (May 1976); and James Medoff, "Layoffs and Alternatives Under Trade Unions in United States Manufacturing," American Economic Review (June 1979).

The traditional neoclassical view of unions asserts that although unions may benefit their members, by creating noncompensating wage differentials they cause allocative efficiency losses. Hence, their net impact on the economy as a whole is thought to be negative.[4] Recently, however, this view has been challenged by Richard Freeman, James Medoff, and their associates at Harvard.[5] Drawing on hypotheses put forth long ago by institutional economists, they argue that unions may well increase productivity. Such increases may occur through a number of routes including union induced reductions in turnover, increases in morale and motivation, and increases in formal and informal on-the-job training. Indeed, several of their econometric studies suggest that union/nonunion productivity differentials in the private sector are often positive.[6] To the extent that these results are generalizable, one must become much more agnostic on the question of whether unions in the private sector have had net adverse efficiency effects.

Research on the effects of unions in the public sector has paralleled the private sector studies. Numerous studies have sought to ascertain the

---

[4] See, for example, Albert Rees, The Economics of Work and Pay, 2nd ed. (New York, Harper and Row, 1979), Chapter 10, for a summary of the neoclassical view, which does not necessarily represent his personal view.

[5] A good nontechnical treatment of their views is found in Richard Freeman and James Medoff, "The Two Faces of Unionism," Public Interest (Fall 1979).

[6] See, for example, Charles Brown and James Medoff, "Trade Unions in the Production Process," Journal of Political Economy (June 1978), for evidence for U.S. manufacturing; Kim Clark, "Unions and Productivity in the Cement Industry" (unpublished Harvard University Ph.D. dissertation, 1978) and his "The Impact of Unionization on Productivity: A Case Study," Industrial and Labor Relations Review (July 1980), for evidence for the cement industry, and Steven Allen, "Unionized Construction Workers Are More Productive?" (mimeograph, 1979), for evidence for the U.S. construction industry. Lest one conclude that unions always increase productivity, the evidence available for coal mining in the U.S. suggests this is not true. See Marie Connerton, Richard Freeman, and James Medoff, "Productivity and Industrial Relations: The Case of U.S. Bituminous Coal" (mimeo, 1979).

effect of public sector unions on the relative wages of teachers, police, firefighters and other categories of municipal employees.[7] Recent studies have moved beyond wage effects and analyzed the effects of public sector unions on nonwage employee benefits and on the trade-off between wages and retirement system characteristics.[8]

In contrast to the private sector research, however, no research has been directed towards ascertaining the effects of unions on productivity in the public sector. This fact is not completely surprising; the concepts of output and productivity in the public sector are often not well-defined and the difficulties inherent in trying to measure productivity are consequently large. Nevertheless, the growing financial problems of state and local governments suggest that this important problem cannot be ignored. Prior studies of public sector wage determination have indicated, on average, that unions have tended to have only modest effects on their members' compensation; studies of public employee unions' effects on productivity are required to complete our understanding of the effects these unions have had on municipal finances and service flows.

---

[7] See, for example, Orley Ashenfelter, "The Effect of Unionization on Wages in the Public Sector: The Case of Firefighters," Industrial and Labor Relations Review (January 1971), Ronald Ehrenberg and Gerald Goldstein, "A Model of Public Sector Wage Determination," Journal of Urban Economics, July 1975, and Hirschell Kasper, "The Effects of Collective Bargaining on Public School Teachers Salaries," Industrial and Labor Relations Review, October 1970).

[8] See, for example, David Rogers, "Municipal Government Structure, Unions, and Wage and Nonwage Compensation in the Public Sector" (unpublished Cornell University M.S. thesis, 1979), Ronald G. Ehrenberg, "Retirement System Characteristics and Compensating Differentials in the Public Sector," Industrial and Labor Relations Review , July 1980, and Linda M. Edwards and Franklin R. Edwards, "The Effect of Unionism on the Money and Fringe Compensation of Public Employees: The Case of Municipal Sanitation Workers" (mimeo, 1979).

This paper represents our initial efforts at analyzing the effects of unions on productivity in the public sector. We first sketch an analytical framework that can be used to estimate these effects, focusing for expository purposes on municipal public libraries. We initially focus on libraries because considerable effort has been devoted to conceptualizing productivity measures for them and because of the availability of data to implement the framework.[9] After discussing the analytical framework, we present preliminary estimates of the effects of unions on productivity in public libraries based upon analyses of data from 71 municipal libraries in Massachusetts. We conclude by indicating how these analyses will be extended and the direction that we hope our future research will take.

II. A Simply Analytic Framework

Municipal libraries produce a variety of outputs which include, but are not limited to, the circulation of books, periodicals, and other audio-visual materials, responding to information and inter-library loan requests, and providing reference facilities. These outputs can in theory be evaluated in both quantitative and qualitative ways. While one can simply count circulation figures or the number of information requests, more sophisticated valuations of library output would focus on questions like "What proportion of information requests were answered correctly?" or "How long did the typical borrower have to wait for a book that he or she wanted?"

For now we shall ignore the fact that libraries can be thought of as multiple product firms and also ignore the quality dimension of the services they provide. Instead, we assume that we can treat library output (Q) as being a

---

[9] For prior discussions of library productivity, see Malcolm Getz, Public Libraries: An Economic View (Johns Hopkins Press, 1980), Karen Feldstein, The Economics of Public Libraries (unpublished MIT Ph.D. dissertation, 1977), F.W. Lancaster, The Measurement and Evaluation of Library Services (Washington, D.C., 1977) and Ernst DeProspo, Ellen Altman, and Kenneth Beasley, Performance Measures for Public Libraries (Chicago, 1973).

single variable. One can then specify the community demand function for library services by

$$(1) \qquad D = D(P|V_1, V_2)$$

Here P is the "price" the community must pay for a unit of library services, other things equal the higher the price the less library services will be demanded. The position of the demand curve will depend upon community income or wealth, with higher income areas demanding more library services and it will also obviously depend upon the size of the community (V_1). Finally, the demand curve will depend upon the community's "taste" for library services (V_2). For example, more highly educated communities may demand more library services, as may communities with a large proportion of school-age children.

The second element of our model is a production function for library services

$$(2) \qquad Q = F(K, L|V_3, U)$$

Here we have treated output, capital (K) and labor (L) as single variables. The capital stock includes the library's entire stock of materials as of the current period. V_3 is a vector which represents those community variables that affect the position or shape of the production function. For example, one early study found that two-thirds to three-quarters of all library users lived within one mile of a library.[10] This suggests that

---

[10] Bernard Berelson, The Library is Public (New York: Columbia University Press, 1949).

increases in population density, which make it easier to locate branch libraries within a mile of all individuals, would increase the output of library services, ceteris paribus.

The production function of library services may also be a function of whether the library's employees are represented by a union. As noted by Freeman and Medoff, unionization may well increase productivity through routes including union induced reductions in turnover, increases in morale and motivation, and increases in formal and informal on-the-job training.[11] On the other hand, unionization of library employees may well reduce productivity if it places limits on library management's ability to substitute factors of production or if it requires library management to devote more resources to the contract negotiation process and to the resolution of grievances.[12] Of key concern to us is what the net effect of unions on the production function is.

The stock of capital that a library has depends upon its stock of capital in the preceeding period (K_{-1}), its investment in new capital (I), and the rate at which its previous stock of capital depreciates (δ). The latter depends upon the age distribution of the library's books (in the main books are used most heavily in the initial years following their purchase) and the resources that the library devotes to maintaining its collection and avoiding theft. We shall ignore the latter two considerations here and treat the depreciation rate as a constant.

---

[11] Freeman and Medoff, "The Two Faces of Unionism," Public Interest, Fall 1979.

[12] See, for example, Marilyn Oberg, Mary Blackburn, and Joan Dible, "Unionization Costs and Benefits to the Individual and to the Library," in Margaret Chaplan, ed., Employee Organizations and Collective Bargaining in Libraries, Library Trends, 25, October 1976, and Milton Byran, "Implications for Public Libraries" in Frederick Schlipf, ed., Collective Bargaining in Libraries (Urbana, Illinois, 1975).

$$(3) \qquad K = I + (1-\delta)K_{-1}$$

The costs incurred by a library are primarily for labor, for new acquisitions, and for maintaining the library's collection and buildings. Let W be the cost per unit of labor, m the per unit cost of maintaining the collection and C the user cost of new materials. Then the total costs the library incurs is given by

$$(4) \qquad C = WL + CI + mK_{-1}$$

For later reference, remember that a primary goal of unions is to increase their members' wages. To the extent that they are successful, W will be an increasing function of U.

The cost function for library services is obtained by minimizing (4), subject to (2) and (3). From this, one can obtain the average cost function for library services.

$$(5) \qquad AC = AC(Q|W,C,m,V_3,\delta,K_{-1}F,U)$$

If the underlying library production function exhibits constant (decreasing) (increasing) returns to scale, average cost will be constant (increase with output) (decrease with output).

We have plotted the demand curve and average cost curve for library services in Figure 1; the latter under the assumption of constant average costs. The average cost curve represents the price to the library of producing different levels of library services. Given the demand curve, D, Q* units of library services will be demanded and produced. At this level of library services, library revenues just cover its costs.

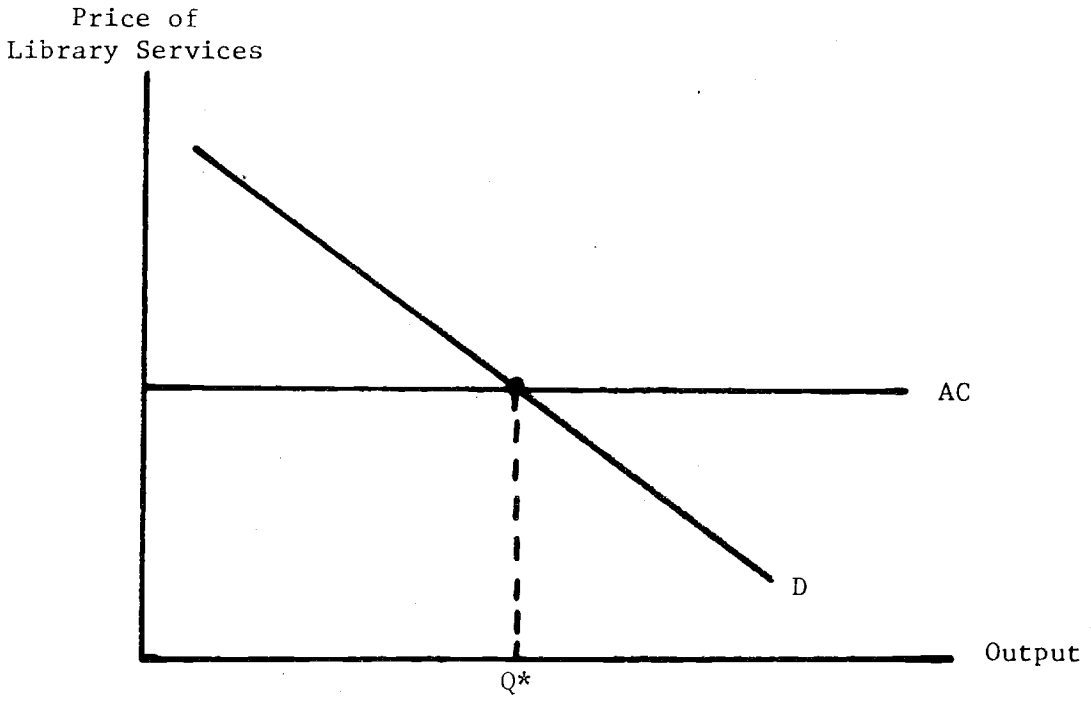

Figure 1

The Market for Library Services

The equilibrium level of library services can be expressed as the reduced form equation

$$(6) \qquad Q^* = G(V_1, V_2, V_3, W, C, m, \delta, K_{-1}, F, U)$$

Obviously, anything that shifts the demand curve up will increase output while anything that shifts the average cost curve up will reduce it. The key point to note is that observed output is determined by both demand and cost factors.

The effects of unions on service flows operate both via their effects on wages and their effects on the production function (2). If unions do increase the wages of library employees this would shift the average cost curve up and reduce output. If unions increase (decrease) the level of output associated with any given input levels (for the reasons discussed earlier) this will shift the average cost curve down (up), thereby increasing (decreasing) output.

Equation (6) provides a simple framework which can be used to estimate the effects of unions on productivity. If cross-section data on library services, the demand and cost variables, and unionization can be obtained, the model can be implemented. The coefficient of the unionization variable in this model would represent the net effect of library unions on productivity. If one were to estimate (6) omitting the wage variable, however, the coefficient of the unionization variable would capture both the net effect of unions on the production function for library services and the effect of union induced wage gains on average costs and hence output.

III. Extensions of the Framework

The simple framework sketched above may be inadequate for a number of reasons. First, the wage of library employees is endogenous, in the sense that it will be determined both by whether the library employees are unionized and the forces that affect the demand for library services, as well as other variables (V_4).[13] As such, if one were to specify a wage determination equation of the form

$$(7) \qquad W = W(V_1, V_2, V_4, U) \quad ,$$

it is likely that the error term in this equation would be correlated with the error term in (6). This might happen, for example, if data on some variable that affected the demand for library services was unavailable and omitted from both equations (6) and (7). If correlated error terms occurred, biased estimates of (6) would result if it was estimated by ordinary least squares.

Second, the extent of unionization, as measured by whether a library's employees are covered by a collective bargaining agreement is also likely to be endogenously determined. It is not unreasonable to expect that collective bargaining coverage will be a function both of state laws governing public employee unionization, and the proportions of public and private employees in a state that are union members. The size of the library is also likely to matter; large libraries may be more bureaucratic in nature and more conducive to unionization.[14] Finally, collective bargaining coverage is likely to be

---

[13] For a discussion of the variables that influence public sector wage determination, see Ehrenberg and Goldstein, op. cit.

[14] Theodore Guyton, Unionization: The Viewpoint of Librarians (Chicago, 1975).

related to both the estimated wage premium associated with collective bargaining and the estimated productivity differential associated with collective bargaining; the former because it influences both employees demand for collective bargaining coverage and library management's resistance to it, the latter because the productivity effects associated with collective bargaining also influence management's resistance to it.

This suggests the need for a full-blown "selectivity bias" corrected model.[15] Separate wage and library output equations can be estimated for libraries covered by and not covered by union contracts, in the context of a model in which the probability that a library is unionized is determined by the estimated union/nonunion wage and output differentials that exist for it, as well as other explanatory variables. The appendix traces out formally how this can be done.

Third, the analyses described above focuses on the effect of unions on observed output; the latter is a determined by both demand and cost considerations. One might prefer instead to focus directly on the underlying production process for library services and ask questions like "Does the existence of unions alter library output per employee?","Do unions affect the substitutability of capital for labor in the provision of library services?". Or shifting to the case of multiple types of library employees (e.g., librarians, other professional employees, library aides, other clerical employees), "Do unions alter the substitutability of different categories of library employees in the face of relative price changes?". Such generalizations would involve using a variant of the "production function" approach

---

[15] See Lung-Fei Lee, op. cit., for the genesis of this approach.

used by Brown and Medoff, and Clark, and/or estimating the parameters of a production function from cost share data.

For example, if the production function for library service in equation (2) can be written as

$$(8) \qquad Q = AK^{\alpha}(L(1+BU))^{1-\alpha}$$

where B represents the proportionate marginal productivity differential of union labor, and U equals one if the library is unionized and zero otherwise, then

$$(9) \qquad \log(\frac{Q}{L}) \approx \log A + \alpha \log(K/L) + (1-\alpha)BU$$

Hence, regressing the log of output per library employee on those demographic variables that affect library productivity (variables that underlie A, the log of the capital/labor ratio, and whether the library is organized would enable one to estimate the proportionate marginal productivity advantage of union labor.[16] The extension to allow for nonconstant returns to scale or more than one category of labor is straightforward in this model. To test for union effects on substitutability, however, obviously requires more flexible functional forms such as the CES or translog ones.

---

[16] See Brown-Medoff, op. cit.

IV. Preliminary Empirical Results: The Determinants of Productivity in Massachusetts Public Libraries in 1977

In 1977 the International City Management Association conducted a survey of municipal public libraries, obtaining data on library revenues and expenditures, employment and wage scales for different categories of library employees, the number of books in each library, and various measures of library usage including circulation, borrowers, and interlibrary loans. The latter three variables were published, by library, in the 1978 Municipal Yearbook, and when coupled with published data on socioeconomic characteristics of cities obtained from the 1977 City and County Databook and published data on whether any library employees were covered by a collective bargaining agreement in each Massachusetts municipality in 1977, permit us to estimate equations of the form

$$(10) \qquad Q_{Ki} = \sum_{j=1}^{10} B_{Kj} r_{ji} + \gamma_K U_i + \varepsilon_{Ki} \qquad K = 1,2,3,4,5$$

for a sample of 71 municipal libraries.

Equation (10) is a condensed version of the reduced form library output equation (6) derived in Section II. The output measures available are interlibrary loans per capita (Q_1), number of borrowers per capita (Q_2), circulation per capita (Q_3), interlibrary loans per borrower (Q_4), and circulation per borrower (Q_5); each measure expressed in natural logarithm form. The former three measures may be thought of as measures of total services provided, while the latter two may be regarded as measures of the quantity of services provided per library user.

The r_j are those variables that are expected to influence library output, either from the demand or cost sides of the model. The cost side is

captured here by population density (r_1); as noted earlier previous studies have suggested that increased population density reduces the cost of providing library services, and hence should increase library output. The demand side is represented by a set of variables expected to influence a community's preferences for library services; these include the percent of the population that is female (r_2), the percent nonwhite (r_3), the percents of the population that are older than age 18 (r_4) and age 65 (r_5), the median education level in the community (r_6), the female labor force participation rate (r_9), and the fraction of employees in the municipality employed in education (r_{10}). The demand side also is represented by the community's capacity to pay for library services, as measured by median family income (r_7), and per capita intergovernmental revenues (r_8). Finally, U_i is a dichotomous variable indicating whether any library employees in the municipality were covered by a collective bargaining agreement in 1977.

Several things should be noted about this specification. Data limitations in this initial study preclude a number of variables that appear in equation (6) from appearing in equation (10). The omitted variables include the wage rates of library employees, the lagged stock of library materials, and the rate at which library materials depreciate. The estimated coefficients which we report below should be considered very tentative then; they may well suffer from omitted variable bias. In particular, because library employees' wages were unavailable in this sample, the coefficient of the unionization variable will capture <u>both</u> the net effects of collective bargaining on the production function for library services and of union induced wage gains on average costs and hence output.

Our preliminary estimates of equation (10) appear in Table 1. The populations of the municipalities in our sample varied from 10,000 to over 500,000 and to control for heterscedasticity we have weighted each observation by the square root of its population. The employees were covered by a collective bargaining agreement in approximately one-quarter of the libraries in the sample.

Turning first to the vector of variables other than unionization, many of these variables affect library output in a manner consistent with our a priori predictions. An increase in the school age population (decrease in r_4) increases library usage as does an increase in the proportion of the population over age 65 (r_5). An increase in the median education level of the population (r_6) also leads to higher usage, as does an increase in the proportion of employees who are employed in the education industry (r_{10}).

In contrast, neither of the variables that reflect the communities' capacity to pay (r_7 and r_8) are significantly related to library output. While an increase in the proportion of the population that is female (r_2) leads to an increase in the number of borrowers, a result that might be expected if females tend not to be in the labor force, an increase in the labor force participation rate of females (r_9) is associated with higher circulation of library materials. Finally, an increase in population density (r_1) leads to a reduction in library output and an increase in the proportion of the population that is nonwhite (r_3) to higher output.[17]

---

[17] Malcolm Getz, Public Libraries: An Economic View, finds a similar negative relationship between density and library output in his study of branch libraries in New York City. He argues that population density may be a proxy for the rental cost of structures, with more dense areas having higher rental rates. If this occurs, the average cost of library services may well be higher in denser areas which would lead, from Figure 1, to a lower level of library services. Thus, a negative relationship between population density and library output may well be consistent with our model.

The primary variable of interest to us, of course, is the unionization variable. These data suggest that libraries covered by a collective bargaining agreement have, ceteris paribus, some 29.2 (exp(.256)-1) percent more borrowers per capita than do nonunion libraries. Both circulation and interlibrary loans per capita appear to be lower in unionized libraries, however these effects are not significantly different from zero. Because of these results, it is not surprising that on a per borrower basis, circulation and interlibrary loans are also lower in unionized libraries.[18]

How robust are these results to the estimation methods used? Row (1) of Table 2 summarizes compactly the estimated collective bargaining effects from Table 1. Rows (2) and (3) show that estimates obtained when the method of ordinary least squares is used (row (2)) and when an additional variable, median age, is included to more fully control for the age distribution of the population (row (3)) are virtually identical.

As noted in Section II, however, all of these estimates may well be subject to selectivity bias. Using the method described in the appendix, one can attempt to control for this problem.[19] First a reduced form probit equation is estimated that determines the probability that a library's employees are covered by a collective bargaining agreement. From these equations, one can compute estimates of variables which are then added to the productivity equations to control for the probability that a library's employees are covered by a collective bargaining agreement. These "augmented"

---

[18] Put another way, if an increase in a variable causes the log of X_1 to increase but does not affect the log of X_2, it is not surprising that the log of (X_2/X_1) falls.

[19] Actually, the method implemented here is simpler since it ignores library employees' wage rates.

productivity equations can then be estimated by ordinary least squares and consistent parameter estimates obtained; this is done separately for libraries that are covered by a collective bargaining agreement and those that are not. Finally, the estimated parameters and the mean values of the explanatory variables can be used to compute consistent estimates of the union/nonunion productivity differentials.

Estimates of the reduced form probit equation appear in Table 3. The variables that appear in the equation include those in the output equations as well as population size, population growth, and the share of employment in the city in a number of industries. These latter variables were included because they tend to be related to the extent of private sector unionization across SMSA's.[20] While the vector of coefficients is clearly jointly significantly at the .05 level, most of the individual coefficients are statistically insignificant. The few significant coefficients suggest that collective bargaining for library employees in Massachusetts tends to occur in cities with older populations (r_5), higher female labor force participation rates (r_9), and lower levels of service industry employment (r_{15}).

These estimates are then used, as described above, to obtain consistent estimates of the productivity equations and estimates of the union/nonunion productivity differentials.[21] These differentials are summarized in row (4) of Table 2; in the main their pattern is very similar to the previous results.

---

[20] See for example, Barry T. Hirsch, "The Determinants of Unionization: An Analysis of Interarea Differences," Industrial and Labor Relations Review, January 1980.

[21] For brevity, we do not report the regression coefficients for the "selectivity corrected" output equations here.

V. Extensions and Future Directions

This paper has laid out a methodological framework for estimating the effects of unions on productivity in the public sector and presented some preliminary estimates for a sample of 71 municipal public libraries in Massachusetts. The empirical estimates themselves should not be stressed, however, as the underlying data suffer from a number of weaknesses. First, there are numerous important variables omitted from the data set, including wages and employment levels of library employees, new acquisitions, and the stock of library materials; these omissions may seriously bias the estimated union effects. Second, we have not made a serious attempt to specify the determinants of whether a library is covered by a collective bargaining agreement, only a limited number of variables were entered into that equation and this could further bias our results. Finally, the libraries in our sample all are located in one state and span a wide range of city sizes (under 10,000 to over 500,000). Since libraries in different size cities perform different functions and the likelihood of collective bargaining coverage is positively related to city size, this will further distort our findings.

We hope to get around all of these problems in future work. The ICMA has made the data tape, upon which their published report on municipal libraries was based, available to us. This tape provides fairly comprehensive data on all of the "library variables" needed to implement the various approaches discussed in Sections II and III for approximately 250 cities of population size 25,000 or greater. Socioeconomic characteristic variables for these cities can be obtained from the 1977 City and County Databook. Finally, we have obtained data on the collective bargaining

coverage of library employees in these cities by a mail survey; our response rate to this survey has been well over 90 percent.

Because these cities do not all lie in one state, it will be possible for us to better model the forces that affect the probability that library employees in a city are covered by a collective bargaining agreement. These include the extent of public and private unionization in a state, variables for which published data exists, as well as the laws governing public sector collective bargaining in a state. Our colleague John Burton has expended considerable effort to collect data on, and to parameterize, these laws and has generously made these data available to us. A substantially better specified probability of collective bargaining coverage equation should reduce the likelihood that our estimates of union productivity effects are subject to selectivity bias.

In addition to our more comprehensive analysis of the effect of unions on productivity in public libraries, we also plan to pursue the question of the routes via which unions influence productivity in the public sector. Our initial focus in this aspect of the project will be on public education and we will make use of a unique set of <u>longitudinal</u> data on educational outcomes, school district background variables, and union contract provisions that our colleague Sam Bacharach has constructed for local school districts in New York State. Our analyses here will permit us to test for the effects of specific union contract <u>provisions</u> on educational outcomes, rather than for the effects of collective bargaining coverage per se. The longitudinal nature of the data will permit the application of econometric methods that allow one to control for omitted variables that otherwise might

bias the analyses.[22] That is, these methods substantially reduce the likelihood that the estimates that result will be subject to selectivity bias because omitted variables that affect the probability of collective bargaining coverage or contract provisions may also affect educational outcomes.

---

[22] For details, see Ronald G. Ehrenberg, "Unions and Productivity in the Public Sector" (proposal submitted to the National Science Foundation, March 1980).

Table 1

Determinants of Productivity in Massachusetts Public Libraries in 1977: Weighted Least Squares (absolute value t statistics)

Ind./Dep. Var./Var.	Q_1	Q_2	Q_3	Q_4	Q_5
r_1[a]	-.111 (3.0)	-.056 (4.3)	-.066 (7.1)	-.059 (1.6)	-.009 (0.8)
r_2	-.060 (0.6)	.090 (2.6)	-.004 (0.2)	-.151 (1.6)	-.094 (2.8)
r_3	.138 (2.1)	.041 (1.8)	.021 (1.3)	.097 (1.6)	-.020 (0.9)
r_4	-.138 (1.9)	-.021 (0.8)	-.067 (3.8)	-.117 (1.7)	-.045 (1.9)
r_5	.192 (1.6)	-.006 (0.1)	.127 (4.3)	.197 (1.7)	.132 (3.2)
r_6	.487 (2.0)	.173 (2.0)	.186 (3.1)	.313 (1.3)	.013 (0.2)
r_7[a]	-.038 (0.5)	-.014 (0.5)	.015 (0.8)	-.023 (0.3)	.028 (1.2)
r_8[a]	-.045 (0.0)	.094 (0.2)	.014 (0.0)	-.149 (0.1)	.080 (0.2)
r_9	4.873 (1.2)	-.879 (0.6)	2.096 (2.1)	5.752 (1.5)	2.975 (2.2)
r_{10}	3.566 (0.8)	2.139 (1.4)	4.668 (4.2)	1.427 (0.3)	2.529 (1.7)
U	-.540 (1.5)	.256 (2.1)	-.128 (1.5)	-.797 (2.3)	-.385 (3.2)
R^2	.334	.515	.731	.235	.417

n = 71 for all equations
a = coefficient has been multiplied by 1000

and

Q_1 = $\log_e$ (interlibrary loans per capita)
Q_2 = $\log_e$ (number of borrowers per capita)
Q_3 = $\log_e$ (circulation per capita)
Q_4 = $\log_e$ (interlibrary loans per borrower)
Q_5 = $\log_e$ (circulation per borrower)

} in 1977

r_1 = population density
r_2 = percent female
r_3 = percent nonwhite
r_4 = percent of the population age 18 and over
r_5 = percent of the population age 65 and over
r_6 = median education level
r_7 = median family income
r_8 = per capita intergovernmental revenue
r_9 = female labor force participation rate
r_{10} = fraction of employees in education

} in 1970

Table 1 (continued)

U = 1=any library employees have a collective bargaining agreement, 0=otherwise

Source:

(1) Q_1 to Q_5 - 1978 Municipal Yearbook

(2) r_1 to r_{10} - 1977 City and County Databook and 1970 Census of Population

(3) U - Collective Bargaining in Massachusetts Libraries: Guidance for Administrators and Staffs (Boston, 1978)

Table 2

Estimated Effects of Collective Bargaining Coverage on the Logarithms of Library Output Measures: Various Estimation Methods and Model Specifications (absolute value t statistic)

Specification/Dep. Var.	Q_1	Q_2	Q_3	Q_4	Q_5
(1)	-.540 (1.5)	.256 (2.1)	-.129 (1.5)	-.797 (2.3)	-.385 (3.2)
(2)	-.573 (1.7)	.252 (1.8)	-.140 (1.5)	-.825 (2.5)	-.393 (2.9)
(3)	-.535 (1.5)	.287 (2.1)	-.098 (1.0)	-.822 (2.3)	-.386 (2.8)
(4)	-.994	.186	-.277	-1.052	-.471

where (1) weighted least squares (union coefficients from Table 1)

(2) ordinary least squares estimates - same model as in Table 1

(3) ordinary least squares estimates - median age added as an additional explanatory variable

(4) selectivity bias corrected estimates, separate equations estimated for union and nonunion sectors (see Appendix A)

Table 3

Probit Estimates of Whether Any Library Employees Are Covered by a Collective Bargaining Agreement in 1977 (absolute value asymptotic t ratio)

Variable	Coefficient
C	17.499 (0.9)
r_1[a]	.119 (0.9)
r_2	-.189 (0.6)
r_3	-.543 (1.0)
r_4	-.496 (1.7)
r_5	.541 (1.9)
r_6	.586 (0.9)
r_7[a]	.116 (0.8)
r_8[a]	3.081 (1.2)
r_9	30.363 (1.9)
r_{10}	.011 (0.1)
r_{11}	-.045 (0.6)
r_{12}	-.052 (0.4)
r_{13}	-.470 (1.6)
r_{14}	.151 (0.6)
r_{15}[a]	.037 (1.4)
r_{16}[a]	.273 (0.1)
r_{17}	.042 (0.2)
Log Likelihood	-23.746
X^2	32.313
n	71

where r_{11} fraction of employees in manufacturing

r_{12} fraction of employees in transportation and public utilities

r_{13} fraction of employees in the service industry

r_{14} fraction of employees in construction

r_{15} 1975 population

r_{16} percentage population change 1960-1975

C intercept term

and

[a]Coefficient has been multiplied by 1000.

Appendix

Our goal is to estimate whether libraries whose employees are union members are more, or less, productive than otherwise identical libraries whose employees are not union members and to estimate the extent that unions increase library employees' wages. Suppose that the output of library i would be Q_{ui} if its employees were covered by a collective bargaining agreement and Q_{ni} if its employees were not covered by a collective bargaining agreement. Suppose also that the wages that the library's employees would receive in the two environments would be W_{ui} and W_{ni} respectively. Then we can define the relative output differential, d_{qi}, and the relative wage differential, d_{wi}, associated with collective bargaining for the i^{th} library, as

$$(A1) \qquad d_{qi} = (Q_{ui}-Q_{ni})/Q_{ni} \approx \log(Q_{ui}/Q_{ni})$$

$$d_{wi} = (W_{ui}-W_{ni})/W_{ni} \approx \log(W_{ui}/W_{ni})$$

In general, it is not possible to observe both Q_{ui} and Q_{ni}, or W_{ui} and W_{ni} with <u>cross-section</u> data, as at a point in time either a library's employees are covered by an agreement or they are not. A naive approach that circumvents this problem is to estimate wage and output equations separately for employees in cities with and without agreements, use the estimated coefficients from these regressions and the characteristics of a city to compute predicted values of the wage and library output that would be observed in both sectors, and then estimate the differentials by calculating the percentage difference in these predicted values.

More formally, suppose that we postulate that the wage rate library employees would receive in a city if they are unionized is a log linear function of a vector of variables, X, which represent all of the variables that would appear in equation (7) in the text, plus a random error term (ε_{1i})

$$(A2) \qquad \log W_{ui} = \sum_{j=1}^{K} \alpha_{ju} X_{ju} + \varepsilon_{1i}$$

and that a similar functional relationship exists that describes the wage that library employees would receive in a city if they were not unionized.

$$(A3) \qquad \log W_{ni} = \sum_{j=1}^{K} \alpha_{jn} X_{ji} + \varepsilon_{2i} \ .$$

Suppose also that similar output equations could be derived; these correspond to the reduced form output equation (6) in the text, where the Y_{ji} represent all of the variables in equation (6) save the extent of unionization, and ε_{3i} and ε_{4i} are random error terms

$$(A4) \qquad \log Q_{ui} = \sum_{j=1}^{M} B_{ju} Y_{ji} + \varepsilon_{3i}$$

$$(A5) \qquad \log Q_{ni} = \sum_{j=1}^{M} B_{jn} Y_{ji} + \varepsilon_{4i}$$

The naive approach would involve estimating the parameters of (A2) and (A4) by ordinary least squares from observations on libraries whose employees were unionized and the parameters of (A3) and (A5) by ordinary least squares

from observations on libraries whose employees were not organized. Given estimates of these parameters ($\hat{\alpha}_{ju}$, $\hat{\alpha}_{jn}$, $\hat{B}_{ju}$, $\hat{B}_{jn}$) and the relevant characteristics of a representative city (X_{ji}, Y_{ji}), one can then obtain estimates of the relative output and wage differentials from

$$(A6) \qquad \hat{d}_{qi} = \log(\hat{Q}_{ui}/\hat{Q}_{ni}) = \sum_{j=1}^{M}(\hat{B}_{ju}-\hat{B}_{jn})Y_{ji}$$

$$\hat{d}_{wi} = \log(\hat{W}_{ui}/\hat{W}_{ni}) = \sum_{j=1}^{K}(\hat{\alpha}_{ju}-\hat{\alpha}_{jn})X_{ji}$$

As is now well known, however, estimates of wage and output equations from truncated samples will not necessarily yield unbiased estimates of the parameters of the underlying wage and output equations (and hence $\hat{d}_{qi}$ and $\hat{d}_{ui}$) since the assumption that the error term in each equation is random and uncorrelated with the other explanatory variables is typically violated. This occurs because libraries are not randomly assigned to collective bargaining status, but rather employees and library management make explicit choices on the matter. Estimates of the wage and output equations that ignore the underlying choice model will be biased because they will confound the effect of an explanatory variable on wages and output with its effect on the probability that the library's employees are covered by a collective bargaining agreement. To correct for this sample selectivity problem requires us to model the underlying economic choice process that determines whether a library's employees are unionized. This problem is complicated by the fact that such an event is a product of both employee and employer decisions.

To keep our estimation problem manageable, we assume that the choice process that determines whether a library's employees are covered by a collective bargaining agreement can be approximated by

$$(A7) \qquad S_i^* = \delta_0 d_{qi} + \delta_1 d_{wi} + \sum_{r=3}^{R} \delta_r Z_{ri} + v_i$$

$$U_i = 1 \text{ if } S_i^* > 0$$

$$= 0 \text{ otherwise}$$

Here S_i^* is an unobserved variable that represents the likelihood that a municipal library will be unionized, v_i is a random error term, and the Z_{ri} are all of the variables expected to influence the probability of observing a collective bargaining agreement, other than d_{qi} and d_{wi}. The parameter δ_0 is assumed to be greater than zero, as positive output effects resulting from collective bargaining should reduce employers' opposition to collective bargaining. The sign of δ_1 is indeterminate, however, as positive union/nonunion wage differentials will increase library employees' demand for collective bargaining, but also increase municipal employers' attempts to resist unionization.

Although S_i^* is not observed, we can arbitrarily scale its cut-off value to be zero, so that if S_i^* is greater than zero, the library's employees will be covered by a collective bargaining agreement ($U_i = 1$). Similarly, if the index is less than or equal to zero, the employees would not be covered by an agreement ($U_i = 0$).

Consistent estimates of the model specified in (A1) through (A7) can be obtained using an iterative procedure originally suggested by Lung-fei Lee.* One can substitute the wage and output equations (A2) to (A5) into (A1) and (A7) to obtain a reduced form probit selection model

$$(A10) \qquad S_i^* = \sum_{t=1}^{T} B_t X_{ti}^* + n_i$$

---

*Lung-fei Lee, op. cit.

where the X^{*}_{ti} are all of the predetermined variables in the model (X, Y, and Z's) and n_i is a random error term. Now suppose that the error terms from this reduced form selection model and the wage and library output equations are jointly normally distributed with means zero and the following covariance matrix

$$(A11)\quad \begin{bmatrix} \varepsilon_{1i} \\ \varepsilon_{2i} \\ \varepsilon_{3i} \\ \varepsilon_{4i} \\ n_i \end{bmatrix} \sim N \left[\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} & \sigma_{14} & \sigma_{1n} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} & \sigma_{24} & \sigma_{2n} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} & \sigma_{34} & \sigma_{3n} \\ \sigma_{41} & \sigma_{42} & \sigma_{43} & \sigma_{44} & \sigma_{4n} \\ \sigma_{n1} & \sigma_{n2} & \sigma_{n3} & \sigma_{n4} & \sigma_{nn} \end{bmatrix} \right]$$

Under these assumptions one can show that

$$(A12)\quad E(\log W_{ui} \mid X_{ji}, U_i = 1) = \sum_{j=1}^{K} \alpha_{ju} X_{ji} + (\sigma_{1n}/\sigma_n)\lambda_{iu} + h_{1i}$$

$$(A13)\quad E(\log W_{ni} \mid X_{ji}, U_i = 0) = \sum_{j=1}^{K} \alpha_{jn} X_{ji} + (\sigma_{2n}/\sigma_n)\lambda_{in} + h_{2i}$$

and

$$(A14)\quad E(\log Q_{ui} \mid Y_{ji}, U_i = 1) = \sum_{j=1}^{M} B_{ju} Y_{ji} + (\sigma_{3n}/\sigma_n)\lambda_{iu} + h_{3i}$$

$$(A15)\quad E(\log Q_{ni} \mid Y_{ji}, U_i = 0) = \sum_{j=1}^{M} B_{jn} Y_{ji} + (\sigma_{4n}/\sigma_n)\lambda_{in} + h_{4i}$$

Here the h_i are normally distributed random variables with mean zero and the λ_i are given by

$$\text{(A16)} \qquad \lambda_{iu} = \phi(-\sum_{t=1}^{T} B_t X^*_{ti}/\sigma_n)/[1 - \Phi(-\sum_{t=1}^{T} B_t X^*_{ti}/\sigma_n)]$$

$$\lambda_{in} = -\phi(-\sum_{t=1}^{T} B_t X^*_{ti}/\sigma_n)/\Phi(-\sum_{t=1}^{n} B_t X^*_{ti}/\sigma_n)$$

where $\phi(\ )$ denotes the normal probability density function and Φ the corresponding distribution function.

Equations (A12) through (A16) make it clear why OLS estimates of the underlying wage and output equations (A2) through (A5) may lead to biased estimates. As long as the error terms in the wage or output equations are correlated with the error term in the reduced form selection rule ($\sigma_{1n} \neq 0$, $\sigma_{2n} \neq 0$, $\sigma_{3n} \neq 0$, $\sigma_{4n} \neq 0$) OLS estimates will be biased due to an omitted variable. While λ_{iu} and λ_{in} are not directly observed, estimates of them may be obtained by first estimating the reduced form probit selection model (A10) obtaining estimated coefficients $(\hat{B}_t/\sigma_n)$, and then using these estimates to compute predicted values $\hat{\lambda}_{iu}$ and $\hat{\lambda}_{in}$ for each individual. Lee (1978) shows that estimation of (A2) to (A5) by OLS, with $\hat{\lambda}_{iu}$ ($\hat{\lambda}_{in}$) added as an additional explanatory variable, over a sample of libraries that are covered by (not covered by) a collective bargaining agreement, will lead to consistent estimates of the α_{ju} and B_{ju} (α_{jn} and B_{jn}). Consequently, consistent estimates of the estimated wage and output differentials associated with collective bargaining coverage may be obtained from (A6).