1984

Union Productivity Effects and Economic Efficiency

Peter Kuhn

Citation of this paper:

Kuhn, Peter. "Union Productivity Effects and Economic Efficiency." Centre for the Study of International Economic Relations Working Papers, 8424C. London, ON: Department of Economics, University of Western Ontario (1984).

CENTRE FOR THE STUDY OF INTERNATIONAL ECONOMIC RELATIONS

WORKING PAPER NO. 8424C

UNION PRODUCTIVITY EFFECTS AND ECONOMIC EFFICIENCY

Peter Kuhn

This paper contains preliminary findings from research work still in progress and should not be quoted without prior approval of the author.

DEPARTMENT OF ECONOMICS
THE UNIVERSITY OF WESTERN ONTARIO
LONDON, CANADA
N6A 5C2

UNION PRODUCTIVITY EFFECTS AND ECONOMIC EFFICIENCY*

Peter Kuhn
University of Western Ontario
London   Ontario   Canada
N6A 5C2

April 1984

Abstract

Because they lack formal theoretical foundations, existing empirical models of union productivity effects tell us little about why such effects occur and what their impact is on economic efficiency. This paper attempts to fill this gap by developing three alternative formal models of union productivity effects with contrasting implications for economic efficiency and then assessing which is most consistent with the available empirical evidence. While it turns out that an "agency" model, which predicts positive union efficiency effects, seems to be able to predict the observed results best, the evidence is far from conclusive. Further tests are suggested to distinguish the models.

UNION PRODUCTIVITY EFFECTS AND ECONOMIC EFFICIENCY

I.   Introduction

Researchers have accumulated considerable evidence indicating that unionization can increase total factor productivity in firms. Studies on this question, which yield positive productivity effects of from 6 to 20 percent include those by Mandelstamm (1965), Frantz (1976), Brown and Medoff (1978), Clark (1980a, 1980b), and Allen (forthcoming).[1] Unfortunately, since virtually no formal theoretical analyses of union productivity effects have been done, interpretation of the results is difficult. In particular, without theoretical models to test against the data, the results can tell us nothing about why union productivity effects occur or how they affect economic efficiency. This paper will provide such a set of models and use them to answer the above questions, based on existing empirical research.

Informal explanations of union productivity effects, which occur repeatedly in the empirical literature, can be summarized by three main arguments. The first proposes that union job security provisions in internal labor markets increase informal training of junior workers by senior workers (Williamson et al, 1975). This argument is not pursued here, primarily because of evidence showing that union workers report receiving less formal and informal training (Duncan and Stafford (1980)) although further evidence on that question might make this argument worth pursuing. The second argument is that more effective elicitation of preferences in making decisions for public goods in the work place by unions, who use preferences of average rather than marginal workers, has beneficial productive effects, such as a reduction in turnover--a claim formalized and evaluated in Kuhn (1983). The third argument seems to be based on "X-efficiency" considerations and argues that

1

unions tighten "managerial slack". As early as 1960, Schlichter, Healy, and Livernash (1960, p. 951) concluded: "The challenge that unions presented to management has, if viewed broadly, created superior and better-balanced management even though some exceptions must be recognized." More recently, Clark (1980a, pp. 465-67) quoted a manager who observed that "before the union, this place was run like a family; now we run it like a business." Brown and Medoff (1978, p. 359) argued that "'union-shocked' management is able to extract more output from a given amount of inputs than management which is not confronted with a union stimulus." This third possible channel of union influences on productivity is examined in this paper, primarily because, once formalized, it yields an important set of results and implications, not the least of which is that union policies that increase productivity by raising managerial quality or effort—something that seems generally accepted by authors in this area to be a "good thing"—may not necessarily be economically efficient. To know whether these policies are or are not efficient, it is necessary to distinguish between alternative models of how such effects arise.

The basic argument of this paper is simply the following. Quality of management, or coordination of resources, like any other input, is a scarce resource. Whether a union-induced increase in the use of that input raises or lowers economic efficiency depends to a large extent on whether the "correct" amount of that input, given its resource cost, is being employed initially.[2] For example, if the quality of management—understood broadly as encompassing factors such as effort and conscientiousness of managers,

ability to promote personal favorites, and use or abuse of authority over employees--is chosen optimally in the absence of a union, then introducing a union that restricts managers' ability to take on-the-job leisure or to indulge personal preferences in on-the-job decisions, is economically inefficient. But if managerial effort is chosen optimally both before and after the union but unions force managers to increase their effort because unions transfer rents from managers to workers, then union productivity effects are neutral in efficiency terms, even though they embody redistributive effects. Finally, if for some reason managerial quality is at a suboptimal level before unionization, perhaps because of an agency problem, union-induced increases in such quality might improve welfare.

Sections II, III, and IV of this paper present simple but formal theoretical models of each of the above cases. A "neoclassical" model presented in Section II derives union productivity effects with negative welfare implications and demonstrates that increases in managerial perfor-mance in response to unionism do <u>not</u> constitute evidence of "X-inefficiency" but are instead perfectly consistent with standard neoclassical analysis. A "shock", or "income-effects" model presented in Section III derives union productivity effects with neutral efficiency implications--unions redistribute rents only. Finally, the "agency" model presented in Section IV shows how positive economic efficiency effects could be associated with positive union productivity effects. Section V then confronts each of the three alternative models with the empirical results to see which of the three can best explain them. Interestingly, the "agency" model seems to be most consistent with the existing evidence, although further tests are surely necessary to resolve the issue more satisfactorily.

II. A Simple Neoclassical Model

The simplest way to illustrate the intuition behind the neoclassical explanation of union productivity effects that operate through increases in managerial quality is to consider a firm that employs only two inputs, managerial services, M, and production labor, L. Input M can be thought of in very broad terms, including not only managers' effort in coordinating and supervising the production process, but such things as the extent to which managers refrain from promoting personal favorites or from abusing their authority over workers. (Extensions to more factors including capital and the model's precise empirical implications are explored in Section V.)

The production function is assumed to be homogeneous of degree 1 and can be written as

$$Q = F(L,M) \tag{1}$$

or

$$Q/L = f(M/L) . \tag{2}$$

Homogeneity implies that the first-order conditions for cost-minimization by this firm can be written in the form

$$\log(M/L) = g(\log[w/s]) \tag{3}$$

where w is the wage of production labor, s is the marginal cost of managerial services, and g is a positive, monotonically increasing function, with the characteristic $g' = \sigma$, the elasticity of substitution between production labor and managerial services. Our assumption that s is constant implies a perfectly competitive market for managerial services, which can be adjusted either by contracting with existing managers for more or less effort or by purchasing more of these services on the open market.

Equation (3) implies that a wage increase brought about by unionism leads the neoclassical, cost-minimizing firm to increase the amount of managerial services used per employee. This will imply, via equation (2), an observed increase in "productivity", defined as output that is produced from _measured_ inputs, where measured inputs do _not_ typically include managerial effort. Since this definition of productivity is used in all empirical work to date (productivity changes are simply output changes that cannot be explained by measured inputs, and are attributed to "better coordination of resources", i.e., more effective management) it is the definition we shall adopt in this paper.

The economic efficiency effects of a union that simply raises wages above the social marginal cost of labor and that causes firms to substitute toward other scarce inputs, such as capital and quality of management, can be nothing but negative. They are also very well understood, since they can be analyzed in terms of traditional two-sector models of unionism, such as those proposed by Johnson and Mieszkowski (1970) and DeFina (1983), simply by replacing the capital input in those models by M. All we shall add to the discussion in those papers here is a set of reminders that (1) at least in the "K,L" version of those models, the estimated magnitude of this loss is always found to be quite small relative to GNP; (2) the effect of unions in these models is to induce the choice of an economically inefficient consumption bundle as well as production technique; and (3) unionism will not involve changes in "technical" efficiency of production (if indeed this term has any meaning once the quality of management is defined as an input) even though measured productivity from

observable inputs increases. In even simpler terms, if all unions do is raise wages in a pure neoclassical model, they will cause an increase in measured productivity in union firms because union firms use more of the unmeasured input, M, per worker. This causes an economically inefficient allocation of M and L between the sectors of the economy and leads society to consume too little of the union-produced good.

III. Income Effect or "Shock" Models

Another possible model of union productivity effects arises when we try to formalize some of the comments made in the institutional literature on this topic. In general, the notion that managers are somehow made worse off by unions--by no longer being able to promote favorites or by cutting back on secretarial staff and on-the-job leisure, for example-- seems to play a major role, suggesting that, according to these analyses, there is some kind of surplus-sharing arrangement between owners and managers of the firm, and that increases in managerial effort arise from income, not factor-substitution effects. This section examines these possibilities through a series of simple models that make varying assumptions about rent-sharing between owners and managers: (1) a single entrepreneur owns and manages the firm; (2) owners get all the surplus; (3) managers get all the surplus; (4) bargaining occurs over how the surplus is shared; and ( ) a general, fixed surplus-sharing rule exists.

Throughout Sections III and IV we will use a key simplifying assumption in order both to present the main results of the models in the simplest possible form and to keep the notation fairly consistent with that in related literature. The contribution of managers to the output of the firm can be written as

$$Q(\underline{a}) \ , \ Q' > 0 \ , \ Q'' < 0 \ , \tag{4}$$

where $\underline{a}$ represents managers' effort and temporarily replaces M/L as the focus of the analysis.

Assumption (4) can be interpreted in two ways. First, the firm's production function is of the form $F(\underline{a},L)$, with $F_{\underline{a}L}$ unrestricted in magnitude; but the effect of unions is simply to raise the total wage bill without affecting employment, L. This is because unions have been able to achieve perfect wage-discrimination or efficient featherbedding rules, as suggested in Leontief (1946), McDonald and Solow (1981), and Hall and Lilien (1976). The other interpretation of assumption (4) is that labor is variable and that unions do alter the marginal costs of labor, but the production function is separable in labor and managerial effort, i.e., $F_{\underline{a}L} = 0$. The analysis in Sections III and IV is generalized to the case where L is variable and the production function cannot be separated in Section V.

1. <u>A single-owner-manager</u>. Our simple firm's profits can now be written as:

$$Q(\underline{a}) - c \ , \tag{5}$$

where c gives the size of the wage bill, and assumption (4) allows us to model the effect of a union purely as an increase in c. If the firm is owned and managed by the same individual, that entrepreneur's utility is given by

$$W_e = U[Q(\underline{a})-c] - V(\underline{a}) \qquad \begin{matrix} U' > 0, & U'' < 0 \\ V' > 0, & V'' \geq 0 \end{matrix} \ , \tag{6}$$

where U is utility of income and V is disutility of effort.

The first-order condition for the maximization of equation (6) with respect to $\underline{a}$ is

$$U'[Q(\underline{a})-c]Q'(a) = V'(\underline{a}) \ . \tag{7}$$

Since the left-hand side of equation (7) is increasing in c but decreasing in $\underline{a}$ and the right-hand side is increasing in $\underline{a}$, it is clear that an increase in fixed costs, c, brought about by the union will make the entrepreneur worse off and induce him to work harder (higher $\underline{a}$). Leisure is a normal good (as is implicitly assumed here), so it is optimal to take this reduction of utility in terms of both lower income and leisure. Unions have no negative or positive efficiency effects, but they increase their own utility at the entrepreneur's expense, with an increase in measured productivity due to greater managerial effort occurring as a byproduct. This redistributive but efficiency-neutral impact of unions is a characteristic of all the examples of shock-effects models discussed below.

2. Pure profit-maximizing firm. Suppose that management and ownership are separated, in the usual manner where managers do not share in the surplus produced by the firm (as in the neoclassical model presented in Section II). Define s as the compensation paid to managers so that the profits (and, hence, utility) of owners are

$$W_o = Q(\underline{a}) - c - s \ . \tag{8}$$

Managers' utility is fixed by the market for managers and is given by:

$$W_m = U(s) - V(\underline{a}) = \overline{W}_m \ . \tag{9}$$

First-order conditions for a maximum of (8) given (9) are simply

$$U'(s) = V'(\underline{a})/Q'(\underline{a}) \ , \text{ and} \tag{10}$$

$$U(s) - V(\underline{a}) = \overline{W}_m \ , \tag{11}$$

which are independent of c. No union productivity effect occurs when managers

do not share in the surplus of the organization and when there is no incentive for factor substitution between L and $\underline{a}$, as we have assumed here.

3. "Managerial" firm. "Managerial" theories of the firm differ from traditional theories in that they are typically, if implicitly, framed in a context where what occurs is maximization of managerial utility (9) subject to a capital-market constraint embodying (8), rather than vice-versa. That this constraint is a basic ingredient in such models is evident in Marris (1964), who uses a minimum valuation ratio in his models, and in Leibenstein (1966), who invokes possible "failure of the firm" as a constraint on managerial satisficing behavior. Similarly, Williamson (1963, p. 243), who sees the firm as a coalition of owners, managers, and workers, defines: "'minimum' after-tax profits...[as] profits negotiated by the other (i.e., non-managerial) members of the coalition that are just sufficient to satisfy their demands", while Machlup (1967, p. 23) in his review of this literature concludes that while many alternative maximands have been proposed for managers, "...all maximanda are subject to the constraint of some minimum benefits to the owners of the business".

Maximization of (9) subject to a minimum level of $\overline{W}_o$ in (8) yields the conditions:

$$U'(s) = V'(\underline{a})/Q'(\underline{a}) \text{ , and} \tag{12}$$

$$Q(\underline{a}) - c - s = \overline{W}_o \text{ .} \tag{13}$$

Note that the "efficiency" criterion, (12), is still the same; but now the actual solution does depend on c. In fact, higher c will entail a higher $\underline{a}$ by managers, as is apparent when (12) is solved for s and substituted into (13):

$$Q(\underline{a}) - c - U'^{-1}[V'(\underline{a})/Q'(\underline{a})] = \overline{W}_o \text{ .} \tag{14}$$

Since $Q(\underline{a})$ and $-s = -U'^{-1}[V'(\underline{a})/Q'(\underline{a})]$ are both increasing in $\underline{a}$, higher c must be accompanied by a higher $\underline{a}$ and a lower s. Managers are made worse off by the increase in c since it is now harder for them to meet the minimum profit constraint. As Williamson pointed out (1963, p. 249), increases in costs, such as unionization, can lead to higher management effort, greater professionalism, lower perquisites (such as offices and favoritism) in such models. Or, as Leibenstein (1966, pp. 408-409) put it, "Thus we have instances where competitive pressures from other firms or adversity lead to efforts towards cost reduction."

    4. A "bargaining" model. It is easy to see that both the profit-maximizing and managerial models are special cases of a more general model. What equations (8) and (9) for the owners' and managers' utility really give us, as $\overline{W}_m$ is varied parametrically, is a utility-possibility frontier in W_o and W_m, of the shape shown in Figures 1 and 2. Moving to the right along the frontier, $\underline{a}$ is increasing and s is decreasing. This is easily seen by deriving

$$ds/d\underline{a} = (V''Q' - V'Q'')/U''Q'^2 < 0 \tag{15}$$

from (10) or (12) and noting that utility of owners increases to the right. The slope of the utility-possibility frontier is

$$dW_m/dW_o = -\lambda < 0 \tag{16}$$

where λ is the Lagrange multiplier in the managerial model, i.e., from the maximization of (9) with respect to $W_o \geq \overline{W}_o$. Its second derivative can be written as

$$d^2W_m/dW_o^2 = -d\lambda/dW_o = -1/D(\lambda Q''-V'')U'' < 0 \tag{17}$$

where $D > 0$ is the bordered Hersian determinant of the Lagrangean. The utility-possibility function is thus concave to the origin . If $\overline{W}_o$ and $\overline{W}_u$ in Figure 1 are interpreted as the next-best alternatives to owners and managers in the market, then it is clear that the profit-maximizing and managerial models of the firm are represented, respectively, by points A and B in Figure 1.
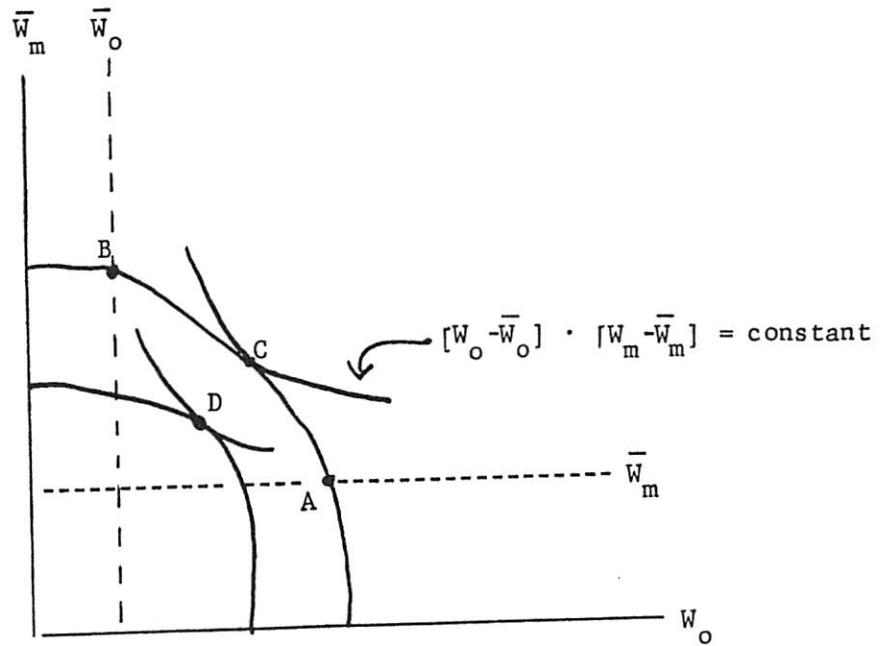
$$[W_o - \overline{W}_o] \cdot [W_m - \overline{W}_m] = \text{constant}$$

FIGURE 1

as

A more likely solution than either point A or B is probably that predicted by Zeuthen's (1930) well-known bargaining model. Consider this bargaining problem as a two-person cooperative game between owners and managers, as in Aoki (1980).[3] If the two parties agree to settle on a point on the utility-possibility frontier, they receive the rewards specified by that point. If they fail to agree, they get $\overline{W}_o$ and $\overline{W}_m$, respectively. These "alternative" utility levels can be interpreted either as "market" returns to capital and to managerial skills or as the well-being of both parties in the case of a takeover bid. If bargaining occurs so that each side makes a concession to the other when, in a well-defined sense, the opponent's willingness to tolerate the risk of a conflict arising is greater, as in Zeuthen's (1930, Chapter IV) bargaining model, the solution turns out to be the $\underline{a}$ and s that maximize $[W_o - \overline{W}_o] \cdot [W_m - \overline{W}_m]$ (see also Aoki, 1980, pp. 605-06). This is the same as the Nash solution to the bargaining problem, as shown in Harsanyi (1977, Chapter 8) and is depicted in Figure 1 as point C before unionization and as point D afterward. What happens to $\underline{a}$ as unions raise costs and the bargaining solution moves from point C to point D? This is easily derived from the first-order conditions for the maximization of $[W_m - \overline{W}_m] \cdot [W_o - \overline{W}_o]$, which after some simple manipulation yield the efficiency condition (10) and

$$Q'(\underline{a})[U\{U'^{-1}(V'(\underline{a})/Q'(\underline{a}))\} - V(\underline{a}) - \overline{W}_m] = V'(\underline{a})[Q(\underline{a}) - U'^{-1}(V'(\underline{a})/Q'(\underline{a}))$$
$$- c - \overline{W}_o] . \qquad (18)$$

Inspection of equation (18) again yields the conclusion that higher c leads to higher $\underline{a}$, giving us a positive union productivity effect.

5. _A general surplus-sharing rule_. To generalize our model further, we note that in reality there are many reasons why the Nash solution might not always be arrived at because of factors like variation in bargaining skill, the possibility of threats and bluffs, and the failure of some of the "stronger" rationality postulates required to achieve solutions in these cases. Suppose that what owners and managers are able to agree on, embodied in established convention and other means, is a general rule according to which surplus produced by the organization is shared. For simplicity, assume that this rule is linear and of the form

$$W_m = \alpha + \beta W_o \qquad \beta \geq 0 . \qquad (19)$$

An example of such a rule is shown in Figure 2. This example, with $\alpha = 0$, implies that a constant ratio of utilities of the two parties has been agreed on. Interesting polar cases are $\beta = 0$, which implies a profit-maximizing firm, and $\beta = \infty$, which implies a managerial firm.

The solution will be the $\underline{a}$ and s that are on the utility-possibility frontier, thus satisfying (10), which also satisfy (19). Points C and D in Figure 2 are two such solutions for different levels of c. The level of $\underline{a}$ at the solution is given by

$$U[U'^{-1}(V'(\underline{a})/Q'(\underline{a}))] - V(\underline{a}) = \alpha + \beta[Q(\underline{a}) - U'^{-1}(V'(\underline{a})/Q'(\underline{a})) - c] . \qquad (20)$$

For $\beta = 0$, this gives the profit-maximizing solution with $\overline{W}_m = \alpha$ and, of course, no union shock effect occurs. $\beta = \infty$ gives the managerial solution with $\overline{W}_o = 0$ and does show a shock effect. The interesting result, however, is that for all intermediate cases $0 < \beta < \infty$ as well, a union productivity effect occurs. Whenever managers participate in the changing fortunes of the firm

(i.e., $\beta \neq 0$), there is a positive union productivity effect. Interestingly, profit-sharing--a device often used to entice managers to behave in a more profit-maximizing way--causes precisely the kind of departure from a "pure" profit-maximizing model of the firm that implies that union productivity effects will occur.
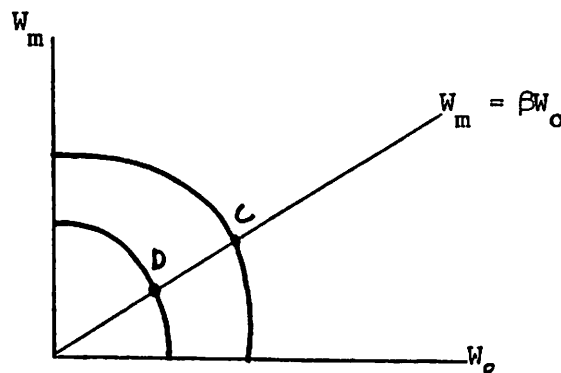

FIGURE 2

IV. An Agency Model

This section presents a crude agency model, adopted from Jensen and Meckling (1976), which shows how union productivity effects can be generated from yet another source. In this model, the reward structure facing managers (i.e., the "entrepreneur" in Jensen and Meckling's terminology) is fixed by the ownership structure of the firm, and direct monitoring is the only device available to principals to induce managers to change their effort.[4]

We begin by considering a firm that, for exogenous reasons, is wholly financed by equity, a proportion, α, of which is retained by the original entrepreneur and now manager of the firm, and $(1-\alpha)$ has been sold to "owners".

Property rights are all well defined, and the question is, given α, to determine the optimal level of monitoring of managers (and hence, of managerial effort) in a union versus a nonunion situation. When this preliminary analysis is completed, we then proceed to characterize the more realistic and interesting cases of what happens when property rights are ill defined and α is endogenous.

We assume that the production function is still that given in (5), or just $Q(\underline{a})$, and write the original entrepreneur's utility as

$$Q(\underline{a}) - \delta C(\Delta \underline{a}) - \underline{a} , \tag{21}$$

where the first two terms, consisting of output minus monitoring costs, $\delta C(\Delta \underline{a})$ (defined below) give the value of the firm, or the entrepreneur's wealth. It should be noted that a fraction, α, of this, accrues in direct "dividends" from the firm, while $(1-\alpha)$ is the proceeds from the sale of shares to outside owners. Disutility of effort is given by the last term, $\underline{a}$. Managers' utility has been assumed to be linear in wealth and leisure in order to eliminate the income effects focused on earlier as a possible explanation of changes in $\underline{a}$.

The agency problem faced by our manager is that, once $(1-\alpha)$ of the firm has been sold, his incentives to supply effort are reduced, and his optimizing condition becomes:

$$\alpha Q'(\underline{a}) = 1 , \tag{22}$$

illustrated in Figure 3 as point n'. This response is assumed to be known and foreseen by all parties, including shareholders. The manager's effort will be $\underline{a}_n$, and his wealth will be n; both are less than what could be achieved if he owned the entire firm: $\underline{a}_m$ and m. Without monitoring,
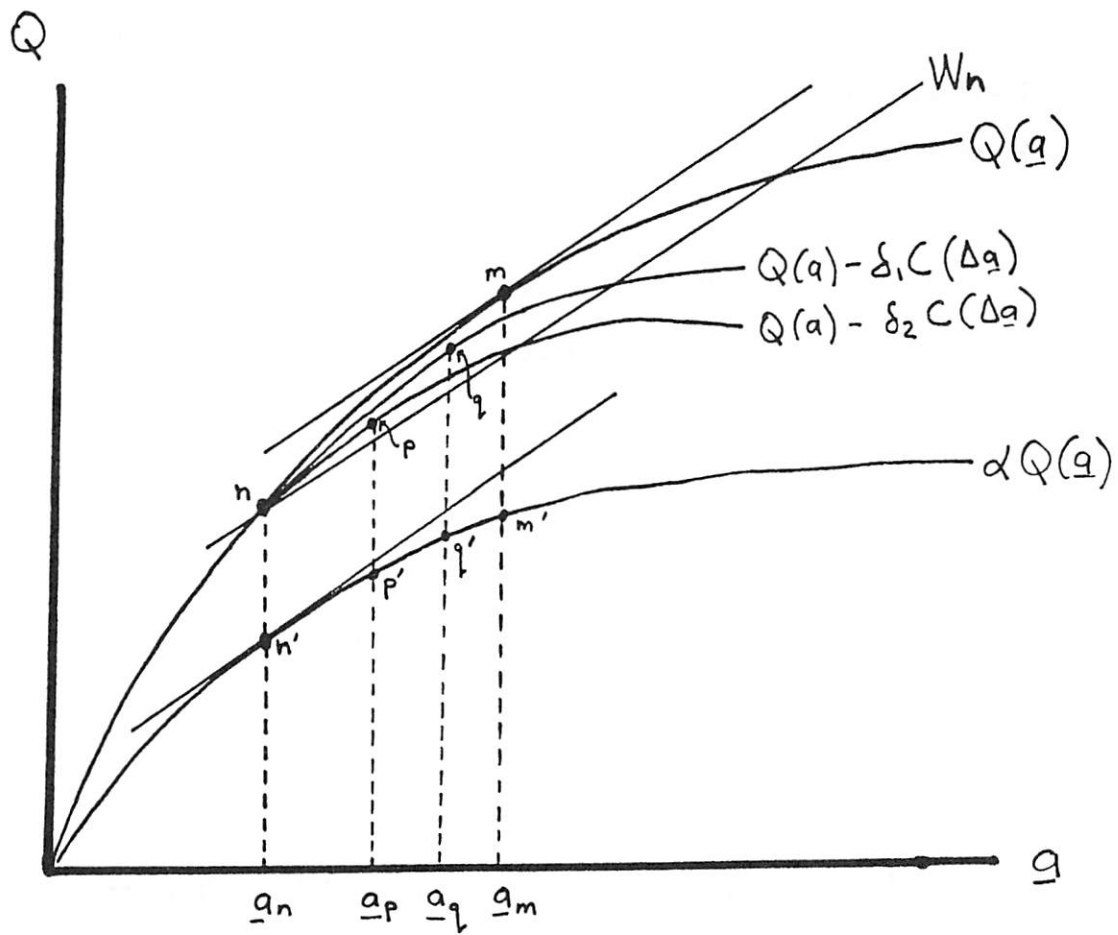
FIGURE 3

the proceeds from the sale of the firm will be nn', and his utility is given
by the indifference curve W_n. This agency problem is partially mitigated,
however, by the fact that the monitoring of managers by owners is possible,
at a convex cost function of $\delta C(\Delta \underline{a})$, where $\Delta \underline{a} = \underline{a} - \underline{a}_n \geq 0$, which is the increase
in effort above the level that managers would choose in the absence
of monitoring. When a given level of monitoring can be written into the
contract whereby shares are sold to owners, and δ, a "shift" parameter in the
monitoring cost function equals δ_1, the optimal amount of monitoring and of
manager effort is given in Figure 3 by point p, with effort $\underline{a}_p$. This move to
p is Pareto superior because owners are still indifferent (paying pp' to get
pp') while managers achieve $W > W_n$.

Finally, suppose that unions, or associations of workers, can monitor
managers better than owners can; specifically, they face $\delta = \delta_2 < \delta_1$. It is
possible that owners may be more numerous and less well organized than
workers. In addition, they (or their higher-level representatives) may rarely
visit the production site, supervising lower-level managers only through a
cumbersome hierarchy, while workers and their union representatives are
always in the plant and can easily observe managers' performance on a daily
or hourly basis (as a joint product with supplying L). The industrial union's
system of stewards, who comprise a hierarchy parallel to the firm's and
constantly monitor the performance of lower-level managers and foremen, has
often been interpreted in precisely this way.[5]

Clearly, management can do better by having unions rather than the
new "absentee" owners monitor them. This can be achieved by either: (1) selling
the $(1-\alpha)$ shares directly to unions instead of owners, or (2) signing a

contract with unions to purchase their monitoring skills for a price $\delta_2 C(\Delta\underline{a})$

and leaving ownership as it is. In either case, the new value of the firm

will be at q and managers' effort will be at $\underline{a}_q$, both of which are greater

than under shareholder supervision. The new union arrangement will be

Pareto superior to the nonunion case and will involve paying union workers

more than nonunion workers to compensate them for their monitoring costs.

This result can be derived more formally by substituting the solution for

$\underline{a}_n$ from (22):

$$\underline{a}_n = Q'^{-1}(1/\alpha) \tag{23}$$

into (21), yielding

$$Q(\underline{a}) - \delta C(\underline{a} - Q'^{-1}[1/\alpha]) - \underline{a}_) \tag{24}$$

then maximizing with respect to a and showing the sensitivity of the result

to δ, but the diagrammatic treatment seems straightforward enough here.

We have illustrated a situation where a union's lower monitoring costs

in an agency setting leads to higher managerial effort (hence higher measured

productivity) and a positive union wage effect, and both predictions appear

to be consistent with reality. Our model, however, still has two unattractive,

unrealistic features; namely (1) the amount of the firm financed by outside

equity, $\alpha < 1$, is exogenously imposed, and yet in the absence of other

constraints the best level of α is clearly zero in Figure 3; and (2) in this world

managers should prefer having unions to not having them.

The first problem is dealt with fairly easily in the same way that

Jensen and Meckling do. Write the production function as $Q(\underline{a}) + G(I)$, where I

is the amount of capital invested in the project, $G' > 0$, $G'' < 0$. In the absence

of wealth constraints, optimal I is given by $G' = r$, where r is the opportunity cost of capital. This situation will have $\alpha = 1$ and $I = I^*$. If the entrepreneur has initial wealth $\overline{I} < I^*$, however, he must raise outside equity[6] to finance his project of the amount $(I - \overline{I})$ for all $I > \overline{I}$. This means that we can write

$$\alpha = \min(\overline{I}/I, 1) , \tag{25}$$

or, in the interesting case where "one" is not binding in (25) and $\overline{I}$ is normalized to equal unity,

$$\underline{a}_n = Q'^{-1}(I) . \tag{26}$$

The entrepreneur's problem now becomes that of choosing both a and I in

$$\underset{\underline{a}, I}{\text{Max}} \; Q(\underline{a}) + G(I) - \delta C[\underline{a} - Q'^{-1}(I)] - \underline{a} - rI , \tag{27}$$

which generalizes (24) and where r is the opportunity cost of capital (assumed to be the same for the entrepreneur as for owners). The first-order conditions for (27) are

$$Q'(\underline{a}) - \delta C'(\underline{a} - Q'^{-1}(I)) = 1 \quad \text{and} \tag{28}$$

$$G'(I) - \delta C'(\underline{a} - Q'^{-1}(I))[-Q'^{-1'}(I)] = r , \tag{29}$$

which have a fairly intuitive interpretation. The first terms in both (28) and (29) are the marginal products of a and I, whereas the right-hand side gives their resource costs. The remaining term gives the marginal agency costs of a and I respectively, both of which are increasing in δ.

An insight into the firm's behavior in this more general case can be had by considering first the case where marginal monitoring costs are constant,

so that $C'' = 0$. Now the firm's choices of $\underline{a}$ and I are independent of each other, and a fall in δ will mean exactly the same rise in $\underline{a}$ as in the simpler version of the model where I was fixed. This fall in δ will also lower I as long as the second-order condition

$$G'' + \delta C' Q'^{-1''} < 0 \tag{30}$$

is satisfied.

When the monitoring cost function is convex, as we have assumed, the analysis is not quite so simple, since induced declines in I lower the marginal agency costs of $\underline{a}$, and vice versa, thus to some extent "damping" the effects described above. To verify that these interactions do not change the direction of the effect of δ, however, we continue to assume that (30) holds (for comparability), differentiate (28) and (29), and solve for $d\underline{a}/d\delta$ and $dI/d\delta$, yielding

$$d\underline{a}/d\delta = (C'/D) \cdot (G'' + \delta C' Q'^{-1''}) < 0 \tag{31}$$

$$dI/d\delta = -(C'/D)Q''Q'^{-1'} < 0 \tag{32}$$

where $D > 0$ is the Hessian determinant of (27). Thus, under fairly general conditions, the results derived earlier regarding effects of monitoring costs on managerial effort continue to hold when ownership structure (and, hence, firm size) are made endogenous in this simple model.

The second problem with the present model is easily solved by considering the actual nature of property rights in this situation. Entrepreneurs will only prefer the socially more efficient union arrangement if they can

appropriate the gains to it by selling shares to unions or by purchasing unions' monitoring services at or near cost. But if unions can appropriate all these gains simply by a vote of the majority of workers in a bargaining unit, we may have a situation where unions increase the size of the pie but make managers and owners worse off, which is consistent with positive union wage and productivity effects as well as the opposition of managers and owners to unionism.

V.   Reconciling and Testing the Models

Of the models presented, the first used a constant-returns-to-scale production function written in intensive form, in which the amount of labor used could have some impact on the marginal product of managerial effort. The remaining two models, for simplicity of illustration and consistency with related literature, were developed in terms of optimal absolute levels of managerial effort that were independent of the amount of labor used because of technology or because unions had efficient featherbedding rules that held labor constant. Since the bulk of the empirical work estimates Cobb-Douglas production functions in intensive form and because it is important to see the implications of L-M substitution for the "income effects" and agency models, our first step in comparing the three models and the evidence is to translate the second and third models into an "intensive" form. This is done below and is then followed by a comparison of the predictions of all three models with the regression results in Brown and Medoff (1978) and Clark (1980a, 1980b). We will refer to situations in which L is variable as the "long run" and to situations where it is fixed as the "short run".

1. <u>Management-labor substitution in the income-effects model</u>. We return now to a CRS production function of the form of (1) and (2), which can also be written as

$$Q = Lf(m) , \qquad (33)$$

where m is the ratio of managerial input per worker. Since we know that any surplus-sharing by managers produces income effects, we can see the implications of such effects for the model's predictions by examining the pure case where the firm is owned and operated by the same individual. This individual maximizes

$$\underset{m,L}{\text{Max}} \ U[Lf(m) - wL] - V(mL), \qquad (34)$$

whose first-order conditions are

$$U' f' - V' = 0 \quad \text{and} \qquad (35)$$

$$U' (f-w) - mV' = 0 . \qquad (36)$$

Dividing (35) and (36) yields

$$w = f - mf' , \qquad (37)$$

which determines the management-labor ratio independently of the size of the firm and of any income effects (embodied in U'), since it amounts to setting the marginal product of labor equal to the wage, i.e., using the optimal factor input combination for any given output. This yields the following:

<u>RESULT 1</u>:

<u>In the "long run", when the firm is free to vary the amount of labor used, union-induced changes in M/L cannot be explained by income effects, since only factor-substitution effects occur in (37).</u>

Managers' behavior will still embody income effects, but these will affect only the decision on firm size, L, not the factor input ratio. Only if unions restrict the firm's adjustment of L through featherbedding rules or wage discrimination schemes will M/L be affected by managers' income effects (consider equation (35) with L exogenously fixed, which gives exactly the same model and results as equations (6) and (7)), and only then might we be able to interpret estimated union efficiency effects as arising from "income" or "shock" effects on managers.

2. <u>Management-labor substitution in the agency model</u>. Using the same production function as above, we can write the manager's maximization problem in the agency model as

$$\text{Max}_{M,L} \quad Q(M,L) - M - B \ . \tag{38}$$

In (38), B represents the total bill for wages and monitoring services paid out by managers. In the absence of a union, this is

$$\bar{w}L + \delta_1 C(M-M^*) \ , \tag{39}$$

where $\bar{w}$ is the nonunion wage and M^* replaces $\underline{a}_n$ in Section IV as the level of managerial services that would be supplied in the absence of monitoring. The first term is paid to workers and the second to owners. Under unionism, we assume that the union does the monitoring and is compensated for doing so through a higher union wage, which is

$$w_u = \bar{w} + (\delta_2 C(M-M^*)/L) \ . \tag{40}$$

The total bill for wages and monitoring comes to

$$w_u L = \bar{w}L + \delta_2 C(M-M^*) \ , \tag{41}$$

which is the same as (39) except for unions' lower monitoring costs. Thus, it is legitimate to model the effects of unions on entrepreneurs' decisions simply as a change in δ.

If we consider the simple case where ownership structure is exogenous, M* will be a fixed parameter and the first-order conditions for a maximum of (38) are

$$Q_M = 1 + \delta C' (M-M^*) \tag{42}$$

$$Q_L = \bar{w} \tag{43}$$

or, in intensive form when Q is CRS

$$M/L = h[\bar{w}/(1+\delta C')] , \qquad h' > 0. \tag{44}$$

Conditions (43) and (44) give us the following:

RESULT 2:

<u>In the long run, when the firm is free to vary the amount of labor hired, agency effects are sufficient to explain union-induced increases in M/L associated with higher wages.</u>

In the pure agency model with M-L substitution, the higher level of M/L is <u>not</u> a neoclassical response to higher union wage demands, because the firm continues to hire labor to the point where its marginal product equals the nonunion wage, $\bar{w}$, as (43) indicates. Managers see the higher union wage bill simply as a compensation for monitoring activity, which is independent of L, so M/L is determined only by the ratio of the marginal costs of monitoring to the nonunion wage in (44). Our conclusion holds <u>a fortiori</u> if the firm is not free to vary L.

3. <u>Confronting the data</u>. It is difficult to interpret the empirical results on union productivity effects because all the econometric work in question estimates firm or industry production functions without incorporating

constraints due to optimizing behavior of firms in response to factor prices. To see this, we note first that most of the production functions estimated in the literature can be seen as special cases of the Cobb-Douglas form

$$Q = (1 + bU)AK^{\alpha}L^{\beta}S^{\delta}M^{1-\alpha-\beta-\delta} , \tag{45}$$

where U = unionism (either as a 0-1 variable or a percentage of the industry); K = capital; L = production labor, typically adjusted for quality; S = non-production labor, not adjusted for quality or effort; and M gives components of supervisor effort and quality not measured in S. The "true" regression equation based on (45) is

$$\log(Q/L) \approx \log A + bU + \alpha \log(K/L) + \delta \log(S/L) + (1-\alpha-\beta-\delta)\log(M/L), \tag{46}$$

in which, for example, Clark omits M/L, and Brown and Medoff omit both M/L and S/L.

The cost-minimization conditions corresponding to (45) and (46), which must be satisfied in all three alternative models presented above (if factor prices are interpreted correctly), are

$$(\beta/\alpha)(K/L) = w/r \tag{47}$$

$$(\delta/\alpha)(S/L) = w/q, \text{ and} \tag{48}$$

$$((1-\alpha-\beta-\delta)/\alpha)(M/L) = w/s , \tag{49}$$

where r, w, q, and s are the market prices of all the factors in (45), in order.

From (47) through (49), two things are immediately clear. First, if all firms had the same production functions and faced the same input prices as we would expect in a competitive market, there would be no differences in

input ratios across firms, and all the estimates are reflecting is some complex combination of things such as differences in technology among firms and measurement errors in input ratios. Second, it becomes clear that in order to interpret clearly the empirical results in question, what is really needed is a theory of how and why w, r, q, and s differ across firms. Fortunately, the theoretical models examined here do just this. We examine the implications of each below.

(a) <u>Neoclassical model</u>. The main contention of the neoclassical model of union productivity effects is that an exogenously imposed wage increase by unions leads firms to substitute away from labor and toward unmeasured factors like managerial quality, causing an upward bias in the coefficient b in (46). Will this be true in the present case? If unions raise w but leave all other input prices the same, as the neoclassical model suggests, we can write

$$w = r[\beta/\alpha](K/L) = q[\delta/\alpha](S/L) = s[(1-\alpha-\beta-\delta)/\alpha](M/L) \ . \qquad (50)$$

This means not only that a higher union wage will induce a higher K/L, S/L, and M/L, but that any one of the measures of factor substitution (log K/L, log S/L, or log M/L) is as good as any other, since the model predicts them to be perfectly collinear. Omitting M/L or S/L or both will <u>not</u> cause a bias in b due to pure neoclassical factor-substitution effects if we believe that the production function is Cobb-Douglas and other factor prices are unaffected by unionism. More complex explanations can be concocted, of course, based on market imperfections that fail to equalize the prices of other factors across sectors and more complex production functions.[7] But it seems at a first pass that the neoclassical model is not as good a candidate for explaining these effects as others are.

(b) Income-effects model. Like the neoclassical model, the income-effects model is based on the assumption that the effect of unions is to raise wages, w. Thus for the reasons discussed above and because of Result 1, which indicated that in the long run all changes in M/L in income effects models are due only to neoclassical factor substitution, this theory cannot explain our observed result in the long run. Its main difference from the neoclassical model is that it can explain short-run effects. In the "short run", when L is fixed, we would expect K/L to be unchanged by unionization but we expect managers to raise S and M purely for income-related reasons. Thus, the collinearity problem in (50) does not arise, so income effects are a possible explanation, at least of Clark's before-after unionism results, but not of any of the cross-sectional results in both Brown and Medoff and Clark.

(c) Agency model. The main difference of the agency model from the two preceding ones is that it is based on a different assumption regarding the effect of unions on factor prices. Instead of raising w, unions are assumed to lower the marginal costs of managerial inputs, i.e., q or s or both, while leaving the actual marginal cost of labor, w, unaffected. Thus, even when the cost-minimization conditions (47) through (49) hold (as we argued they would in equation (44)), the multicollinearity problem noted in (50) does not arise. Only the agency model predicts variation in the ratio of managerial input to labor that is independent of included variables, such as the capital-labor ratio; for this reason it is the only model presented here that can explain estimated positive cross-sectional union productivity effects in the long run.

VI.   Conclusions

This paper has attempted to fill a considerable gap in the literature

on union productivity effects by proposing three alternative formal models

of that phenomenon to test against the available data. Several useful

conclusions emerged from that exercise. First, the almost automatic

identification of union productivity effects with improvements in economic

efficiency is simply incorrect on theoretical grounds; whether these observed

changes constitute efficiency improvements depends on which of the

theoretical models more accurately describes reality.

Second, of the three theoretical models proposed, the one that

predicts positive efficiency consequences of union productivity effects

does seem to be most consistent with the existing empirical results.

This is because (1) our "neoclassical" model, which only applies to long-

run factor-substitution effects, does _not_ predict a positive union productivity

coefficient when a Cobb-Douglas production function is estimated in intensive

form; (2) our income-effects model predicts a productivity effect only in

the "short run" when labor is fixed; while (3) the agency model predicts a

positive union coefficient in both the short and long run.

Third, and perhaps most important, the analysis highlighted some

of the problems inherent in interpreting empirical estimates of production

functions that treat input ratios as exogenous without incorporating

optimizing behavior by firms. This suggests that further work in the area--

in order to better distinguish various behavioral hypotheses about

unions, such as the three models presented here--should proceed in at

least two new directions. One is to choose econometric techniques that

get away from the assumption of exogenous input levels, such as estimating

a cost function that treats prices as given or estimating a set of factor demand equations that does the same and also incorporates optimizing behavior by firms. As Binswanger (1974) has shown, this approach can identify all the parameters of the production function, including a union productivity effect, as well as incorporate the important possibility of nonconstant elasticities of substitution as possible influences in the neoclassical model outlined earlier. One may wish to go even further and endogenize the size of the union wage effect given unions' preferences and the shape of the production function, if that direction appears to be a fruitful one.

The second area of research is to try to test some of the other predictions of the three models directly. For example, the neoclassical and income-effects models can be distinguished by the fact that in one, managers always get their market alternative utility level, which means that the union-induced increase in effort must be compensated for by an increase in income. In the other, effort increases because managers' well-being must fall, and so long as income is a normal good this means that managers' compensation should fall after unionization.

Perhaps some other easily identifiable distinguishing characteristics of the models can be derived (for example in the case of a firm with diminishing returns in its production function the pure neoclassical model predicts that output should contract as a result of unionism while the pure agency model predicts that it should expand). We hope that, now that a more formal theoretical basis for such research has been established, interest in pursuing the matter in these directions will increase.

Notes

[1]Recent evidence on the relation between unionism and productivity growth (see, for example, Hirsch and Link, 1984) as well as the negative productivity effects found in the U.K. by Pencavel (1977) suggest that the results of these careful analyses may not apply universally over firms or over time. While resolving the issue of when and where union productivity effects do occur is an important empirical question, it is not analyzed here, since our main purpose is to distinguish between possible causes of positive union productivity effects when they occur.

[2]This same point applies to the question of whether union-induced declines in turnover or changes in other inputs like worker effort and quality are economically efficient.

[3]Managers in our model take the place of workers in Aoki's model.

[4]In a more general formulation of the agency problem, another tool would be available to mitigate this incentive problem; namely, the design of a reward schedule facing managers that is not necessarily a constant proportion of output, i.e., αQ, as we have assumed. As several authors have shown, under certain circumstances, such as perfectly observable output, optimal nonlinear reward schedules can completely eliminate the incentive problem (e.g., Harris and Raviv, 1978). One could, however, imagine incorporating optimal reward schedules into the present model without much change in the basic results by having output unobservable and managers' effort imperfectly observable. Then, "cost of monitoring" can be seen as devoting resources to increasing the reliability of observation, η, of effort, while the optimal reward schedule is being chosen for each possible η. We can imagine that unions could produce higher η for the same resource

cost, and the same analysis as here would go through. The present model, however, seems adequate enough for the purpose at hand, which is to distinguish the implications of three general types of models for the empirical results in question

[5]As an anonymous referee has pointed out, it is also possible that unions monitor worker effort better than owners or managers do, at least in craft unions. This possibility is not treated here but we should be able to analyze it in a very similar model.

[6]Debt financing is ruled out in our simple model. It should be noted, however, that because debt has its own agency costs, the solution will not in general involve 100 percent debt but a combination of debt and equity, as Jensen and Meckling derive.

[7]It is interesting to ask what would have to be true of a production function $f(L,K,M)$ for an increase in wages to cause an increase in log M/L, even controlling for log K/L, since this would have to be true if the positive union coefficient is to be explainable by neoclassical factor-substitution effects. If the production function is homogeneous of any degree and we normalize $s = r = 1$ (since these prices are not changing), this condition can be expressed as the requirement that $[(d \log M/L)/(d \log w/s)]/[(d \log K/L)/(d \log w/r)] \equiv (\sigma_{ML})/(\sigma_{KL})$ be increasing in K/L. In other words, for factor substitution effects due to union wage increases to help explain estimates of union productivity effects when the production function is homogeneous, it must be true that management becomes relatively more substitutable with labor (in the traditional Hicks sense) than does capital as the capital-labor ratio rises. While this is certainly possible, it seems a rather tenuous premise on which to base a theory until further evidence is available.

References

Allen, S. G. "Unionized Construction Workers Are More Productive."
Quarterly Journal of Economics, forthcoming.

Aoki, M. "A Model of the Firm as a Stockholder-Employee Game." American
Economic Review (September 1980), pp. 600-10.

Binswanger, H. P. "A Cost Function Approach to the Measurement of Elasticities
of Factor Demand and Elasticities of Substitution." American Journal
of Agricultural Economics (May 1974), pp. 377-86.

Brown, C. and J. Medoff. "Trade Unions in the Production Process." Journal
of Political Economy 86:3 (1978), pp. 355-78.

Clark, Kim. "The Impact of Unionization on Productivity: A Case Study."
Industrial and Labor Relations Review (July 1980a), pp. 451-69.

_____. "Unionization and Productivity: Micro-Econometric Evidence."
Quarterly Journal of Economics (December 1980b), pp. 613-39.

DeFina, R. H. "Unions, Relative Wages, and Economic Efficiency." Journal of
Labor Economics (October 1983), pp. 408-29.

Frantz, J. "The Impact of Trade Unions on Production in the Wooden Household
Furniture Industry." Senior Honors Thesis, Harvard College, 1976.

Hall, R. and D. Lilien. "Efficient Wage Bargains under Uncertain Supply
and Demand." American Economic Review (December 1979).

Harris, M. and A. Raviv. "Some Results on Incentive Contracts with
Applications to Education and Employment, Health Insurance, and Law
Enforcement." American Economic Review (March 1978), pp. 20-30.

Harsanyi, J. Rational Behavior and Bargaining Equilibrium in Games and
Social Situations. Cambridge, Mass., 1977.

Hirsch, B. T. and A. N. Link. "Unions, Productivity, and Productivity

    Growth." <u>Journal of Labor Research</u> (Winter 1984), pp. 29-37.

Jensen, M. and W. Meckling. "Theory of the Firm: Managerial Behaviour,

    Agency Costs, and Ownership Structure." <u>Journal of Financial</u>

    <u>Economics</u> (1976), pp. 305-60.

Johnson, H. and P. Mieszkowski. "The Effects of Unionization on the

    Distribution of Income: A General Equilibrium Approach." <u>Quarterly</u>

    <u>Journal of Economics</u> (November 1970), pp. 539-61.

Kuhn, P. "Implicit Contracts, Union Politics, and the Theory of Intra-

    Firm Compensation Structure: A Synthesis." Mimeograph. University

    of Western Ontario, 1983.

Leibenstein, M. "Allocative Efficiency vs. X-Efficiency." <u>American Economic</u>

    <u>Review</u> (June 1966), pp. 392-415.

Leontief, W. "A Pure Theory of the Guaranteed Annual Wage Contract."

    <u>Journal of Political Economy</u> (February 1946).

McDonald, I. and R. Solow. "Wage Bargaining and Employment." <u>American Economic</u>

    <u>Review</u> (December 1981), pp. 896-908.

Machlup, F. "Theories of the Firm: Marginalist, Behavioral, Managerial."

    <u>American Economic Review</u> (March 1967), pp. 1-33.

Mandelstamm, A. "The Effects of Unions on Efficiency in the Residential

    Construction Industry." <u>Industrial and Labor Relations Review</u>

    (July 1965), pp. 503-21.

Marris, R. <u>The Economic Theory of 'Managerial' Capitalism</u>. New York:

    Macmillan, 1964.

Pencavel, J. H. "The Distributional and Efficiency Effects of Trade Unions

    in Britain." <u>British Journal of Industrial Relations</u> (July 1977),

    pp. 137-56.

Slichter, S. The Impact of Collective Bargaining on Management. Washington: Brookings, 1960.

Stafford, F. P. and G. Duncan. "Do Union Members Receive Compensating Wage Differentials?" American Economic Review (June 1980), pp. 355-71.

Williamson, O. E. "A Model of Rational Managerial Behavior." In R. M. Cyert and J. G. March (eds.), A Behavioral Theory of the Firm. Englewood Cliffs, N.J.: Prentice-Hall, 1963.

Williamson, O. E. et al. "Understanding the Employment Relation: The Analysis of Idiosyncratic Exchange." Bell Journal of Economics (Spring 1975).

Zeuthen, F. Problems of Monopoly and Economic Warfare. London, 1930.