Union Effects on Productivity, Profits, and Growth: Has the Long Run Arrived?

John T. Addison, University of South Carolina and Universität Bamberg

Barry T. Hirsch, University of North Carolina at Greensboro

This article interprets literature examining union effects on economic performance. Production function studies indicate small overall union impacts on productivity; positive effects, where they exist, appear to result from management response to decreased profit expectations and from a natural selection process. Lower profitability among unionized firms is well established; more interesting is the possibility that unions appropriate quasi rents deriving from long-lived tangible and intangible capital. The connection between unions, investment behavior, and productivity growth emerges as a particularly fruitful line of empirical inquiry, although it does not encourage a sanguine view of unionism’s long-run impact.

I. Introduction

Controversy continues to surround the nature and direction of union effects on economic performance. Much of the impetus behind this debate stems from the work of Richard Freeman and James Medoff (1984), who point to potential increases in productivity resulting from union-induced changes in the workplace. Such increases are effected through the exercise...
of "collective voice," coupled with an appropriate institutional response from management. Unions, it is argued, lower turnover and establish more efficient governance structures in workplaces characterized by public goods, complementarities in production, and long-term contractual relations. Although recent literature continues as before to investigate productivity differentials, increasing attention is being given to an examination of union effects on profitability, investment, and growth.

This article interprets recent literature examining union effects on economic performance. We first address the ambiguity surrounding the production function test itself, prior to offering an interpretation of observed productivity differences. The newer literature examining union effects on profitability is next analyzed. Issues yet to be resolved in this area include the magnitude of the union profit effect, the consistency of union productivity and profitability estimates, and the sources from which unions capture profits. Finally, we turn to what may be the most interesting aspect of the literature; namely, the dynamic effect of unions on productivity growth, investment activity, and long-run performance. It is concluded that the debate over what unions do will focus increasingly on the longer-term consequences of union rent seeking.

II. Productivity and the Production Function Test

Measurement Issues

The majority of the union productivity studies follow Brown and Medoff (1978) in employing some variant of the Cobb-Douglas production function

\[ Q = AK^a(L_u + cL_n)^{1-a}, \] (1)

where \( Q \) is output, \( K \) is capital, \( L_u \) and \( L_n \) are union and nonunion labor, respectively, \( A \) is a constant of proportionality, and \( \alpha \) and \( (1 - \alpha) \) are the output elasticities with respect to capital and labor. The parameter \( c \) reflects productivity differences between union and nonunion labor. If \( c > 1 \), then union labor is more productive, in line with the collective voice model; if \( c < 1 \), then union labor is less productive, in line with conventional arguments concerning the deleterious impact of such things as union work rules and constraints on merit-based wage dispersion. Manipulation of equation (1) yields the estimating equation (we ignore the error term)

\[ \ln(Q/L) \simeq \ln A + \alpha \ln(K/L) + (1 - \alpha)(c - 1)P, \] (2)

where \( P \) is union density \((L_u/L)\).

Equation (2) assumes constant returns to scale. This assumption may be relaxed by including a \( \ln L \) variable as a measure of establishment size. The coefficient on \( P \) measures the logarithmic productivity differential
of unionized establishments; if it is assumed that the union productivity effect solely reflects the differential efficiency of labor inputs, the union labor productivity effect is calculated by dividing the coefficient on $P$ by $(1 - \alpha)$. Also, to anticipate what follows, $c > 1$ implies that unions have higher total factor productivity (TFP), as obtains after subtracting $\alpha \ln(K/L)$ from both sides of (2),

$$\ln Q - \alpha \ln K - (1 - \alpha)\ln L = \text{TFP} \simeq \ln A + (1 - \alpha)(c - 1)P. \quad (3)$$

A number of limitations attach to the production function test. As Brown and Medoff themselves noted, the use of value added as an output measure confounds price and quantity effects. That is, part of the measured union productivity differential may result from higher prices in the unionized sector. Not surprisingly, estimated union productivity effects tend to be lower when price adjustments are made (e.g., Allen 1984b) and are rarely large in studies where $Q$ is measured explicitly in physical units. Union firms can more easily pass through higher costs in product markets sheltered from nonunion and foreign competition. Use of value added, therefore, is most likely to confound price and output effects in aggregate analyses relating industry value added to industry union density. It is less likely to be a problem in firm-level analyses that at least allow the possibility of controlling for a firm’s union status and industry density. A related possibility is that the union coefficient in equation (2) may crudely be tracking the union-nonunion wage differential.\(^1\)

The most pointed criticism leveled at the Brown-Medoff approach has come from Morgan Reynolds, who argues: “Estimates of $c$ are of no consequence in discovering the independent impact of unionization on productivity because there is no way to statistically separate observed marginal productivity differentials from union/nonunion price [wage] differentials” (Reynolds 1986, pp. 445–46). He points out that if profit-maximizing firms operate on the labor demand curve, they will adjust employment such that the marginal value product, $\partial Y/\partial L$, equals the wage. If union establishments face a higher wage, they will lower employment so that they have a correspondingly higher marginal value product. According to Reynolds (1986, p. 445), we thus will observe

$$w_u/w_n = (\partial Y/\partial L_u)/(\partial Y/\partial L_n) = c, \quad (4)$$

\(^1\)This argument was suggested to us by Gregg Lewis in correspondence. It relies on the argument that, by definition, value added is the sum of union and nonunion labor cost and a residual that chiefly comprises the return to owned capital. Following Lewis, a Taylor expansion of such an equation yields a $\ln(Q/L)$ function in which union density ($P$) enters in conjunction with labor’s share times the union wage premium. The suggestion is that the coefficient on $P$ from eq. (2) varies directly with the union-nonunion wage differential.
where subscripts $u$ and $n$ denote union and nonunion, respectively. Regardless of whether or not capital and labor quality are controlled for statistically, union firms facing higher wage rates must have a correspondingly higher marginal value product in long-run equilibrium. The production function test, according to Reynolds, proves only that union wages are higher, not that unions increase productivity.

How valid is Reynolds's criticism? He is correct in stating that firms operating on their labor demand curves will adjust employment such that $\frac{\partial Y}{\partial L} = w$. Reynolds's case, then, rests on the question of what in fact is measured econometrically in the production function studies and whether union wage-employment settlements are located on labor demand curves. Because the production function studies control for capital inputs and labor quality, it can be argued that they measure vertical shifts of the labor demand schedule resulting from union productivity effects, namely, a comparison of union and nonunion ($\frac{\partial Y}{\partial L})\_K$. By contrast, movements along a nonshifted demand schedule imply a changing ratio of labor to other inputs. The production function approach, then, need not be an inappropriate test of the union productivity effect if other inputs in the production process can be measured accurately. (Other problems may of course make the approach inappropriate.) But, as Brown and Medoff, as well as Reynolds, point out, union firms facing higher wage rates must be more productive if they are to survive in the long run. Hence the union productivity effect is not being measured across a representative sample of firms. Rather, only those union firms that have survived by increasing productivity (a shift in the demand schedule) sufficiently to offset higher union wages are observed. Measurement of union productivity differentials from among a sample of surviving firms thus overstates the productivity effect of unions on a representative firm. It does measure a productivity rather than a wage effect, however, although in the long run these two effects should tend to converge.

The Reynolds criticism must be further qualified if wage-employment outcomes are not located on the labor demand schedule. As is well known, settlements on the demand curve are generally inefficient (McDonald and Solow 1981). There exist settlements off the labor demand curve, at lower wage and higher employment levels than would obtain through sequential wage and employment determination, preferred by both the union and management. A settlement on the contract curve (formed by the tangencies of the union indifference and firm isoprofit curves) requires explicit or implicit contracting that jointly determines wages and employment; otherwise, once a wage was determined, the firm could maximize short-run profitability by adjusting employment to the corresponding point on the demand schedule. No static inefficiency would result in the event of settlements on a vertical contract curve, referred to as the "strong efficiency" case (Brown and Ashenfelter 1986). Both parties agree to maximize the
joint value of the enterprise to owners and labor and then bargain over its division. Despite skepticism as to the actual union political model and contractual mechanisms that would produce settlements off the demand schedule (Oswald 1984), empirical studies generally reject the hypothesis of wage-employment outcomes on the demand schedule and find mixed support for the hypothesis of strongly efficient settlements (Brown and Ashenfelter 1986; Card 1986; Eberts and Stone 1986; MaCurdy and Pencavel 1986; Svejnar 1986; and Abowd 1987).

If contract settlements off the labor demand schedule obtain, the link between union-nonunion relative wages and marginal products is severed. We do know, however, that the marginal value product of union labor (measured on the demand schedule) will be less than the union wage, whereas the two will be equated in nonunion firms (ignoring the possibility of sufficiently large threat effects whereby nonunion outcomes mimic union outcomes). The equality stated by Reynolds in equation (7) need no longer hold; the union-nonunion productivity differential will be less than the union relative wage effect. The appropriateness of the production function test, therefore, continues to rest largely on econometric issues.

Also troubling is the assumption made in most studies of identical production functions in the union and nonunion sectors (i.e., \( \alpha_u = \alpha_n \)). Such an assumption is necessitated in aggregate studies by the absence of separate data on the union and nonunion sectors, although union density can be interacted with other right-hand side variables in aggregate models. Brown and Medoff (1978) investigate the consequences of relaxing this assumption and find the estimated union coefficient sensitive to assumed differences in \( \alpha_u \) and \( \alpha_n \), being sharply lower in most cases. The assumption of equivalent production parameters is probably less serious—and less necessary—in industry-specific studies (see Clark 1980a, 1980b). In addition, estimation of the production function in log-linear form, made possible by use of a first-order Taylor-series approximation, biases upward the absolute value of the estimated union productivity effect (Lovell, Sickles, and Warren 1988). The implication is that existing Cobb-Douglas studies have over-stated (by an apparently modest degree) both positive and negative estimates of union productivity effects.

The production function test also may depend crucially on the ability to control accurately for all inputs in the production process. Union and nonunion establishments may differ systematically in the quality of unmeasured organizational factors so that "firm effects" are not independent of union status. For example, inputs such as managerial supervision and the quality of labor relations may be correlated with unionism, and omission of these factors may bias the union coefficient. That being said, neither Brown and Medoff nor Clark find their estimates to be sensitive to the exclusion of labor quality controls.

One response to the broader problem of a restrictive functional form has been to assume a translog rather than a Cobb-Douglas technology
Union Effects 77

(Boal 1985; Bemmels 1987). The translog form allows one to test for restrictions on technology made by Cobb-Douglas—for example, unitary elasticity of substitution between inputs and homogeneity. Translog studies have decisively rejected the Cobb-Douglas specification. Yet as Allen (1987) points out, estimation of any production function by ordinary least squares (OLS) is strictly appropriate only when input quantities are exogenous (see Nerlove [1965], pp. 29–34, for a more general analysis of this point). This is clearly not the case in the production function test, posing a simultaneity problem for all such studies. One response to the problem has been to measure the relative efficiency of union and nonunion establishments using (translog) cost and profit functions (e.g., Allen 1987), in which the union effect is captured via intercept and slope parameters that vary by union status. Both functions have the advantage of permitting a direct estimate of the net effect of unions on profits or costs, in contrast to the alternative practice of comparing union coefficients from separate production and wage equations.2

Evidence and Interpretation

In light of the concerns regarding the production function test, critical evaluation of the empirical evidence is required. We make no attempt to provide a complete review of what is now a large literature (for surveys, see Freeman and Medoff [1984]; Hirsch and Addison [1986]); rather, we examine closely what we believe are the most important studies and patterns evident in this literature. The original Brown and Medoff (1978) paper will be our benchmark. Using aggregate two-digit manufacturing industry data cross-classified by state groups for 1972, Brown and Medoff obtain coefficients on union density of from .22 to .24, implying values of \(c - 1\) (obtained by dividing the union coefficient by \(1 - \alpha\)) of from .30 to .31. In a separate analysis using the 1973–75 Current Population Survey (CPS) files, they estimate a logarithmic union-nonunion wage differential of .23 (averaged over males and females).

Most of the potential measurement problems previously discussed apply with some force to the Brown-Medoff study. While these problems are largely unavoidable and recognized by the authors, any generalization of their results must meet the dual criteria of plausibility and consistency.

2 An alternative approach to measuring productivity differences is the estimation of nonstochastic, nonparametric frontier production functions (Forsund, Lovell, and Schmidt 1980). These estimation procedures allow production efficiency to be decomposed into allocative (price) and technical efficiency, the latter being further decomposed into a Farrell measure of efficiency (a firm is technically inefficient if it operates on the interior of its production set), a measure of input “congestion,” and a measure of scale efficiency. In principle, such estimation results would facilitate inferences as to how unions affect productivity. This approach may prove promising for industry-specific union productivity studies.
with subsequent findings. By either standard, their estimate of the union productivity effect is too high. Labor's share in value added is .53 (Brown and Medoff 1978, p. 367), implying an increase in unit cost of 12% (.23 times .53) stemming from the union wage premium. A .22 to .24 increase in total factor productivity, however, in turn would decrease unit costs by over 20%. Hence, the net effect of union wage and productivity effects would be to decrease unit costs substantially. Although union effects on the profit rate cannot be ascertained with precision from the change in unit costs (Clark 1984), these parameter estimates would almost certainly imply an increase in profitability resulting from unionism. But the profits literature provides unambiguous evidence of lower profitability in unionized industries and firms.

Wessels (1985) casts further doubt on the plausibility of large union productivity estimates because of inconsistencies between this result and employment effects. Specifically, he argues that if wage setting is constrained by the demand function and labor is measured in efficiency units in the production function, it is impossible to reconcile equal wage and productivity increases with conventional (below unity) estimates of the elasticity of labor demand. The possible exception is where unions exclusively enhance capital productivity, but even here it is shown that the elasticity of substitution between capital and labor would have to fall considerably below unity to produce the result that unions raise productivity substantially. Again, conventional estimates of the substitution elasticity are too high to reconcile apparently small union employment effects with the measured impact of unions on productivity. Wessels concludes that one of the effects must be wrong: either unions do not substantially increase productivity or they substantially reduce employment.3

There are surprisingly few manufacturing- and economy-wide productivity studies and, except for Brown and Medoff, none reports evidence of a positive overall union productivity effect. In perhaps the best study to date, Clark (1984) uses the Profit Impact of Market Strategy (PIMS) data for 902 manufacturing businesses from 1970–80 (over 4,600 usable observations) to estimate value-added (and scales) equations similar to (2). He obtains marginally significant coefficients on the union variable from −.02 to −.03, in sharp contrast to the results in Brown and Medoff. The Clark study has the advantage of a large sample size over multiple years, business-specific information on union coverage, and a detailed set of control variables (although the union coefficient is little affected by the inclusion of

3 Although his basic point is sound, limitations attach to Wessels's analysis. He does not address the possibility of efficient bargaining outcomes off the demand curve, he appears to say (Wessels 1985, p. 103) that unit costs do not change if productivity and wages increase by equal percentages (this ignores the fact that labor's share is less than one), and he rather glosses over the diversity of research findings.
the latter). The study can otherwise be subjected to many of the same criticisms that apply to the Brown-Medoff analysis. In Clark’s separate two-digit industry regressions, positive union productivity effects are found only for textiles, furniture, and petroleum (estimates range from 6% to 17%).

Studies by Bemmels (1987), who estimates a translog production function using a limited sample of 46 surveyed manufacturing plants in 1982, Hirsch (in press), who estimates a variant of equation (2) using a sample of 315 Fortune 1,000 companies, and Lovell, Sickles, and Warren (1988), who estimate an economy-wide annual time-series model of equation (2), all conclude that unions decrease productivity (the estimate of $c$ in the latter study is implausibly low). Based on the extant evidence, we conclude that the average union productivity effect is probably quite small and, indeed, is just as likely to be negative as positive.

As expected, results from industry- and firm-specific productivity studies produce a varied picture. The primary advantage of these studies is that many of the econometric problems inherent in the aggregate studies are avoided. For example, output can be measured in physical units rather than value added, information on firm-level union status is more readily available, and coefficients can be allowed to differ between the union and nonunion sectors. Moreover, recent studies have moved away from the overly restrictive Cobb-Douglas specification or have attempted to avoid input endogeneity problems by estimation of cost rather than production functions. But these advantages are achieved at the price of a loss in generality. Taken as a group, however, we believe that within the diversity of these results are to be found systematic patterns that increase our understanding of how unions affect the workplace.

From a methodological perspective, two of the best analyses are Clark’s studies of the cement industry (Clark 1980a, 1980b) and Allen’s most recent analysis of the construction industry (Allen 1987). Clark’s studies are notable for the use of physical output measures, for allowing production function parameters to vary between union and nonunion plants, in controlling for firm effects through the study of plants changing from nonunion to union status, and in introducing a supervisory labor input measure. In his wider study, Clark (1980b) investigates productivity differences between nine nonunion (29 observations) and 119 union (436 observations) cement plants over the period 1973–76. He finds marginally significant (using one-tailed tests) union coefficients in the range .07–.10. An interesting result is that union proproductivity effects are most pronounced in the Southwest region of his sample, where nonunion firms are most prevalent, although, as before, the union coefficients are estimated with considerable imprecision. In a novel study of six cement plants that changed from nonunion to union status between 1953 and 1976, Clark (1980a) reports a positive union productivity differential of from 6% to 8%. Although significance
levels are again weak, evidence from his fixed effects model indicates a small positive productivity effect that is apparently sustained through time.

In perhaps the most ambitious study to date, Allen (1987) attempts to measure the relative efficiency of union and nonunion establishments using cost and profit functions. Translog cost function are estimated for three separate samples of building construction projects (commercial office buildings, schools, and hospitals). As noted earlier, this approach allows for the endogeneity of input quantities and also sidesteps problems attaching to the comparison between union productivity and wage effect estimates from separate production and earnings functions containing different control variables or estimated at different levels of aggregation. Allen finds higher costs for union contractors in school and hospital construction, but lower costs in commercial office buildings, ceteris paribus. The hypothesis of equal cost function parameters is soundly rejected across all specifications in the schools sample; however, differences between coefficients are less evident in the commercial office buildings sample (while the number of nonunion hospitals is too limited to allow such tests in the hospital sample).

A striking result from this study is the finding of greater scale economies in the union sector; diseconomies appear in the nonunion sector at lower output levels and increase more rapidly with output. Union office building construction costs less once buildings exceed a threshold size and, thereafter, becomes steadily less expensive with output. For schools, on the other hand, even though marginal costs fall with output in the union sample (and rise in the nonunion sample), nonunion contractors are more cost efficient at all output levels. Taken in conjunction with his profits results, these findings support a competitive union explanation for office buildings and a monopoly union explanation for schools and hospitals (see, relatedly, Allen 1986a). Office buildings are built for the private sector while schools and some hospitals are built for the public sector. Allen argues that market segmentation characterizes public sector contracts, allowing union construction to have higher costs without sacrifice in profits. In commercial office buildings, by contrast, the absence of market segmentation and greater cost consciousness require competitive unionism. Scale economies in unionized construction are attributed to the effectiveness of union hiring halls in guaranteeing large supplies of (already screened) skilled labor and, in part, to the irrelevance of certain restrictive practices (e.g., minimum crew sizes) on large projects. He is unable to quantify these effects, although it does appear that his results reflect more than just compositional effects (see Allen [1986b] for further discussion of scale economies).

We have dwelled at some length on the Allen study because not only does it provide a glimpse inside the union productivity black box, but also demonstrates palpably that the union impact is not a datum, hints at the potentially important role of market structure, and addresses a number of
deficiencies in the literature. We note that the rationale for positive union productivity effects is not rooted in collective voice considerations because of the short-term employment relation that obtains in construction.

Despite substantial diversity in the industry-specific studies, several systematic patterns are revealed. As previously noted in Hirsch and Addison (1986, chap. 7), estimated productivity effects tend to be largest in those industries where union-nonunion wage differentials are most pronounced. Of course, such a pattern is exactly what would be predicted by critics of the production function test. We contend that these results also support a traditional “shock effect” interpretation of unionization. In industries or firms where unions obtain large wage premiums, management must respond to the increase in labor costs by organizing more efficiently, reducing slack, and increasing measured productivity.4

Second, positive union productivity effects are largely restricted to the private sector and appear to be most significant where competitive pressures are intense. This latter point was taken up by Clark (1980b), who found the largest union productivity effect in the region where nonunion competition was most pronounced. Similarly, in an early and widely cited study of residential construction in two cities, Mandelstamm (1965) identified competition from outside contractors in a nearby metropolitan area as the chief source of the greater efficiency observed in his more highly unionized city. And in Clark’s (1984) analysis of lines of business, the largest positive union-nonunion productivity differentials were found in industries having relatively large nonunion sectors (textiles, furniture, and petroleum). The apparent absence of a sizable productivity effect in public libraries (Ehrenberg, Sherman, and Schwarz 1983), government bureaus (Noam 1983), schools (Eberts and Stone 1987), and hospitals (Sloan and Adamache 1984), despite significant union wage premiums in all but the library sample, supports the thesis that product market competition is a spur to greater efficiency in unionized markets (this theme is examined directly by Allen [1987, 1986a]).

Perhaps the most direct evidence of a shock effect is reported by Clark (1980a), who finds that management adjustments to unionism exhibit fairly consistent patterns in the wake of unionization. In all cases new plant managers were hired, and in most cases there was increased emphasis on cutting costs, establishing production targets and goals, and improving monitoring and communications. Yet a problem of interpretation remains.

4 Causation could be argued to work in the opposite direction, with union-induced productivity increases leading to large union wage premiums. This appears unlikely. If unionism raised productivity directly, there would be mutual gains from unionization, with labor and shareholders sharing in the gains. Thus we would observe management in nonunion firms encouraging union organizing and higher profits in union firms. Neither outcome is observed. Moreover, firm-specific productivity changes should not affect competitive, market-clearing wages.
The evidence indicates that a relatively competitive, cost-conscious economic environment is a necessary condition for a managerial response to unions that will produce a positive productivity gain. Such a response should also be larger, ceteris paribus, the larger the union wage gain (or, alternatively, the greater the pressure on profits). It is precisely in such competitive environments, however, that there should be relatively little managerial slack and the least scope for union organizing and wage gains. This reinforces our earlier conclusion that the possibility for sizable union productivity effects is limited. Clearly, any judgment taken on the potential scope for union productivity enhancement hinges crucially on one’s priors regarding the degree of competitiveness and slack in the economy.\(^5\)

For this reason, a complementary and perhaps more appealing explanation of the empirical regularities in the literature would be that the productivity studies suffer from a selectivity or survival bias. As discussed in the previous section, union firms unable to increase productivity in response to union wage increases are less likely to survive and be included in researchers’ data samples. Thus, estimates of union effects on productivity (or profits and costs) are biased estimates of the union effect on a representative sample of nonunion firms. Not all firms are “shocked” or able to respond sufficiently to offset union cost increases. In the long run these firms will not survive.

Direct evidence on the routes through which unions affect the workplace remains meager. As Brown and Medoff (1978, p. 374) state: “The idea that unions make firms . . . more productive would be more persuasive if the mechanisms by which productivity is improved could be isolated.” A central tenet of the collective-voice framework motivating these analyses is that the substitution of voice for exit will reduce quits and improve morale and cooperation among workers. Related arguments suggest that improved governance structures can lead to union productivity gains. Yet the evidence is opaque. While cross-section and time-series evidence reveals that quits are significantly lower among union workers, ceteris paribus (Freeman 1980; Blau and Kahn 1983; and Mincer 1983), there is little to suggest that this advantage significantly contributes to productivity.\(^6\) Clark (1980a) reports either no change or even higher quits following unionization in three of the six plants he followed over time. One route through which union voice might improve productivity is via the establishment of procedures

---

\(^5\) Unions initially may organize and gain bargaining power during a period where a firm or industry has significant market power. Union impact on profitability is likely to intensify over time as foreign and domestic nonunion competition increase and union wage increases cannot be passed through to consumers.

\(^6\) Brown and Medoff introduce a quits variable into their production function and attribute the reduction in the union coefficient to the effects of lower quits on productivity. This procedure reduces the union coefficient by one-fifth; the residual four-fifths are loosely attributed to “better management, morale, motivation, communication, etc.” (Brown and Medoff 1978, p. 74).
to handle grievances, but there is no direct evidence to support this view. Ichniowski (1986), using data from 10 unionized paper mills, finds the number of grievances filed to be inversely related to productivity (see also Katz, Kochan, and Gobeille 1983); left unanswered in this treatment, however, is the effectiveness of union voice in reducing and/or arbitrating grievances relative to implicit grievance procedures occurring in similar unobserved nonunion plants.

In sum, we should expect to find rather more positive and direct indications of the operation of collective voice. Further ambiguity is occasioned by Allen’s (1984a) finding that absenteeism is at least 30% higher among union workers (even though he reports that the decline in productivity from this source is only a fraction of 1 percent). The greater dissatisfaction apparently expressed by union workers need not be inconsistent with union-induced improvements in efficiency via collective voice (Freeman and Medoff 1984, chap. 9; Leigh 1986). But neither is it inconsistent with a shock effect explanation whereby the work pace is increased in unionized settings, leading to greater worker dissatisfaction.

Finally, there remains the unresolved problem of union endogeneity. We know that union workplaces differ substantially from nonunion workplaces (Duncan and Stafford 1980; Freeman and Medoff 1984). Can unions more successfully organize where there exist special cost advantages? Or, as suggested by Brown and Medoff (1978, p. 368), are unions more heavily organized in less productive (two-digit) industries? Do U.S. companies’ newer and largely nonunion plants have higher productivity than their older and more highly organized plants? We know of no study that has satisfactorily addressed the issue of union endogeneity in estimating union-nonunion productivity differentials.

We conclude that there is no compelling evidence that, in general, the net effect of unions on productivity is positive or negative. Rather by default, the explanation for positive union productivity effects, where they are observed, is best explained by either shock effect arguments or by deficiencies in the production function test. Actual union productivity gains appear to be generated by management responses to sizable wage premiums, significant nonunion competition, and deteriorating profit expectations. Estimated union-nonunion productivity differentials may result from an entangling of union productivity, price, and wage effects. Or we may be observing the outcome of a competitive process in which less productive unionized enterprises have been selected out of the system. Accordingly, our analysis of union effects must examine dynamic aspects of economic performance.

III. Unions and Profitability

Closely related to the interpretation of the productivity evidence is the issue of how labor unions affect profitability. One of the more troubling
puzzles in the literature has been the seeming inconsistency between the empirical suggestion of a substantial union productivity effect and mounting evidence that unionism significantly reduces profitability. Information on union wage and productivity effects, coupled with knowledge of labor’s share, makes possible a calculation of union effects on unit costs. For example, a productivity effect of 9% would leave a firm’s unit costs unchanged if labor’s share were .60 and the union wage premium 15% (it is assumed that the firm is fully unionized or, alternatively, that nonunion employees also receive a 15% premium). The link between unit costs and profit rates, however, is not transparent. In general, it depends on the nature of the bargaining settlement (e.g., on the demand curve or strong efficiency), the product demand elasticity, market structure, and the elasticity of substitution between capital and labor (Clark 1984).

Perhaps the simplest case is that of sequential wage and employment determination on the labor demand curve, with offsetting productivity and wage effects that leave unit costs unchanged. For a product market monopolist, Clark (1984, pp. 896–97) shows that the rate of return on capital will decrease (increase) if the elasticity of substitution is greater (less) than one, whereas the return on sales will remain unchanged (because \( Q \) and \( P \) do not change).

More interesting is the strong efficiency case. For purposes of comparison, examine, initially, the case of inefficient sequential wage and employment determination on the demand curve, in which (subject to constraints) the union maximizes “rents,” \( R \), and the firm maximizes profits, \( \Pi \), given \( w_u \). That is,

\[
\max R = (w_u - w_0)L \quad \text{(union maximand),} \tag{5}
\]

and

\[
\max \Pi = PQ - rK - w_uL \quad \text{(firm maximand),} \tag{6}
\]

where \( Q \) is output, \( P \) product price, \( K \) capital, \( r \) the price of capital, \( w_u \) is the realized union wage, \( w_0 \) is the wage in the absence of unions, and \( L \) is employment. A measure of the excess of the union wage bill over the competitive wage bill, \( R \), has been a common maximand assumed in the literature (e.g., Rosen 1969).

Such an outcome is inefficient: settlements preferred by both parties exist off the demand curve, at a lower wage, and a higher employment level. An efficient bargaining situation on a vertical contract curve implies that the two parties will maximize (borrowing terminology from Abowd [1987]) the total value of the enterprise \( V \), being the sum of firm profits (\( \Pi \)) and union rents (\( R \)), and then bargain over division of the surplus. Maximizing \( V \) results in the same output, price, and input usage as obtains
in the case where the firm maximizes $\Pi$ subject to the competitive wage or opportunity cost wage, $w_0$; that is,

$$\max V = \Pi + R = PQ - rK - w_uL + (w_u - w_0)L = PQ - rK - w_0L. \tag{7}$$

Thus, the firm adjusts employment according to the opportunity cost wage and not its "own" wage. Here, the union has no short-run real effects (unions will have long-run effects to which we turn subsequently). Since in this case, $Q$, $P$, $K$, and $L$ are not affected by the union, we can state unambiguously that a union wage increase ($w_u > w_0$), with no offsetting productivity increase, will decrease the profit rate, whether measured by rates of return on sales or capital, the price-cost margin, or market-value measures.

A positive union productivity effect, taken in conjunction with a union wage increase that leaves unit costs unchanged, will also result in a lower profit rate (Clark 1984). Continuing to assume that the firm acts as if it faces competitive factor prices, a factor-neutral union productivity effect would have the same effect as a decrease in marginal costs, increasing optimal output and lowering price. Input usage will increase or decrease ($K/L$ remains constant) depending on whether the product demand elasticity $\eta$ is greater or less than unity (Clark's monopoly model implies $\eta > 1$). The profit rate will decrease since $P$ falls, unit costs remain constant, and the use of $K$ varies proportionately with output.

The point of the above discussion is that a comparison of union wage and productivity effects provides only an approximate estimate of the union effect on profit rates. By the same token, evidence on union wage and profit effects does not allow a precise estimate of the union productivity effect. Empirical evidence on union-nonunion profit rate differentials, however, does permit us to gauge the plausibility of productivity estimates. More fundamentally, profitability is important in its own right since it is profitability that should most affect firm survival, management behavior, and long-run resource allocation.

Before turning to the empirical evidence on union profit effects, some preliminary comments are in order. The empirical studies are surprisingly recent, especially considering the long-standing attention given union wage effects by labor economists and the plethora of profitability studies from the industrial organization literature. Studies have employed as their unit of observation either aggregate industry, firm, or line-of-business (LB) data. Several profitability measures have been examined. The industry studies most frequently use the price-cost margin (PCM), defined as (total revenue - variable costs)/total revenue, typically measured by

$$\text{PCM} = (\text{value added} - \text{payroll} - \text{advertising})/\text{shipments}. \tag{8}$$
Firm-level analyses, by contrast, are likely to use accounting measures of the rate of return on sales \( r_s = \text{earnings/sales} \) or the rate of return on capital \( r_s = \text{earnings/assets} \).

Several of the firm-level analyses focus on publicly traded firms and use a stock market value measure of profitability, the most frequent being Tobin’s \( q \), defined as

\[
q = \frac{\text{market value}}{\text{replacement costs of assets}}.
\]

Firm-level studies also have employed “events” analyses that calculate changes in a firm’s market value attributable to not fully anticipated events involving unionization (e.g., representation elections or changes in contract provisions). A typical measure here is that of cumulative abnormal returns, CAR, defined as

\[
\text{CAR} = \sum r_{\text{act}} - r_{\text{exp}},
\]

where \( r_{\text{act}} \) is a firm’s actual rate of return (including dividends and capital gains) and \( r_{\text{exp}} \) is the firm’s predicted rate of return based on a “market” equation (which regresses a firm’s \( r \) on a broad-based market portfolio’s \( r \)) estimated over some period well before and/or after the event. CAR are summed over the period for which it is believed public information about, or anticipation of, the event might affect the stock price. CAR typically are averaged over a sample of firms in order to draw inferences as to average union effects.\(^7\)

None of the measures used in the literature is entirely satisfactory. Price-cost margins (PCMs) are readily measurable with industry data and correspond most closely to the excess of price over variable (or marginal) cost, on which so much industrial organization theory centers. In practice, however, PCMs may not correspond closely to the profit measures they are intended to represent (Liebowitz 1982). At the firm level, accounting rates of return are readily available from financial reports and measure actual or realized, as opposed to expected, earnings at particular points in time. Unlike accounting measures, stock market indicators such as \( q \) are forward-looking rather than historical, measure performance over time rather than for single periods, adjust for risk differences across firms, and are less likely to be influenced by endogenous accounting procedures. The event studies also use stock market measures and have the added advantage of identifying the effects of specific events on market value. Yet isolating the effects of unanticipated events is often elusive (Binder 1985), and it

\(^7\) For a detailed review of union events studies, see Becker and Olson (1987). Here, we do not consider events analyses examining the effect of strikes on security prices.
Union Effects 87

may be difficult to generalize from "marginal" union effects (e.g., certification elections and concession bargaining) to average union effects on profitability.

A summary of U.S. estimates of union effects on profitability is presented in table 1. The most striking result of the studies is the common theme of lower profitability in union regimes, despite substantial differences among analyses in methodology, data source, unit of observation, time period, and measure of profitability. Moreover, the magnitude of the reduction in profits is large. Although observed percentage reductions are not strictly comparable, there are some striking similarities. Of the aggregate manufacturing industry studies using Census of Manufactures and Annual Survey of Manufactures data, Freeman (1983) finds that moving from 0% to 100% union coverage lowers price-cost margins by 13%-19%. Corresponding estimates by Voos and Mishel (1986a) and Domowitz, Hubbard, and Petersen (1986, 1987) point to reductions in PCMs of 23% and 22% to 25%, respectively.

Among the firm-level analyses, Clark (1984) reports that union businesses have rates of return on capital and sales that are, respectively, 19% and 18% lower than those of their nonunion counterparts, while Hirsch and Connolly (1987) estimate union effects on $r$, ranging from 11% to 17% and on Tobin's $q$ from 13% to 20%. Hirsch (in press) reports sizable differences in $r_k$ and $q$ between companies with low and high levels of collective bargaining coverage. Ruback and Zimmerman (1984) obtain highly variable results, but find on average that union wins in representation elections reduce market value by 3.8%, whereas union losses lower market value by 1.3%. Abowd (1987) calculates the cost of union contracts for a sample of publicly traded firms and finds that unanticipated changes in union wealth are offset, roughly dollar for dollar, by changes in shareholder wealth. An anomaly among the studies is Allen's (1987) finding that profits on union construction projects are at least as high as those on nonunion projects, despite higher union costs in two of his three sectors.8

Some perspective on the magnitude of these estimates is in order. Typically, the reported estimates (or partial derivatives) are evaluated on the basis of a change from zero to complete union coverage, when in fact most industries and large firms have partial coverage. On the other hand, the negative union coefficients may be biased upward (toward zero) in many of these studies for reasons having to do with simultaneity bias, measurement error, and selectivity bias. Unions are more likely to organize in

8 Allen attributes the ability of union projects to pass through higher prices to the existence of geographical market segmentation whereby union and nonunion firms in school and hospital construction rarely compete directly against each other. Such segmentation is possible because of the lack of cost pressure in these sectors and prevailing wage legislation that typically applies here but not in his private office building sample.
<table>
<thead>
<tr>
<th>Study</th>
<th>Primary Data Sources</th>
<th>Profitability Measures</th>
<th>Selected Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aggregate manufacturing:</td>
<td>U.S. manufacturing industries:</td>
<td>Natural logs of price-cost margin (PCM) and the rate of return to capital.</td>
<td>Unionism lowers both measures significantly; moving from 0% to 100% coverage, profits fall 13%–19% in SOM sample and 39%–44% in IRS sample. Profit reduction is restricted to highly concentrated industries.</td>
</tr>
<tr>
<td>Karier (1985)</td>
<td>U.S. manufacturing; state by two-digit SIC industries; primary source is Census of Manufactures: 1972.</td>
<td>PCM</td>
<td>Unionism lowers PCM significantly; reduction is restricted to moderately and highly concentrated industries. Unions capture about 68% of &quot;potential&quot; monopoly rents.</td>
</tr>
<tr>
<td>Domowitz, Hubbard, and Petersen (1986)</td>
<td>Panel of 284 four-digit U.S. manufacturing industries; Census of Manufactures and Annual Survey of Manufactures: 1958–81.</td>
<td>PCM (measured in levels and, alternatively, first-differenced in fixed effects models).</td>
<td>Unionism lowers PCM significantly; average reduction moving from 0% to 100% coverage about 25%. PCM-concentration relation procyclical. Little evidence, except during periods of high unemployment, to support claim that union profit effect restricted to highly concentrated industries.</td>
</tr>
<tr>
<td>Domowitz, Hubbard, and Petersen (1987)</td>
<td>Panel of 284 four-digit U.S. manufacturing industries; Census of Manufactures and Annual Survey of Manufactures: 1958–81.</td>
<td>Alternative measurement of PCM incorporating use of intermediate inputs (materials) in production.</td>
<td>Unionism lowers PCM significantly; average reduction moving from 0% to 100% coverage about 22%. Unionism reduces PCMs in all categories of industries, with largest effect in durable goods industries. Concentration increases PCM in consumer and durable</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>Salinger (1984)</td>
<td>U.S. manufacturing firms; Compustat tape: 1976</td>
<td>Tobin’s <em>q</em> and the rate of return on capital.</td>
<td>Firms in more unionized industries have monopoly rents captured. Nonlinear least squares estimation with three-way interaction between industry unionism, concentration, and entry barrier variables. Union parameter significant in <em>q</em> equations; insignificant in rate of return equations.</td>
</tr>
<tr>
<td>Hirsch and Connolly (1987)</td>
<td>U.S. Fortune 500 manufacturing firms; Compustat tape: 1977.</td>
<td>Tobin’s <em>q</em> and the rate of return on sales (<em>r</em>).</td>
<td>Unionism decreases profits, but estimated magnitude is sensitive to specification. Reduces <em>q</em> by 13%–20%; <em>r</em> by 11%–17%. Concludes unions do not capture profits associated with concentration; returns from market share, research and development (R&amp;D), and weak foreign competition are more likely sources for union rents.</td>
</tr>
<tr>
<td>Study</td>
<td>Primary Data Sources</td>
<td>Profitability Measures</td>
<td>Selected Results</td>
</tr>
<tr>
<td>----------------------------</td>
<td>--------------------------------------------------------------------------------------</td>
<td>------------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Liberty and Zimmerman (1986)</td>
<td>Publicly traded firms, matched with labor contracts covering 5,000+ workers; The Wage Calendar or Wall Street Journal for contract data, Compustat</td>
<td>Accounting earnings and market value of firms.</td>
<td>Find no evidence that firms reduce reported earnings during contract negotiations to influence union demands. Unionized firms exhibit a poor earnings performance during this period and have slower sales and market</td>
</tr>
<tr>
<td>---------------------------</td>
<td>----------------------------------------------------------------------------------------------------------------------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Allen (1987a)</td>
<td>Profit per project. Value of contract minus sum of costs for labor, material, and capital, all divided by value of the contract.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Similar profits on union and nonunion projects. Despite higher union costs in school and hospital construction, higher project prices fully, or more than fully, offset these higher costs.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Union chains have significantly lower profits (76% on average). Unions shown to reduce profit associated with concentration (CR) and market share (MS) when the union variable entered only in interaction with CR or MS; the union effect not found to vary with CR or MS when both a union dummy and an interaction variable are included.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>No significant union productivity effect, but significant compensation premium in UAW-organized firms, implying lower profitability. In a supplementary analysis, probit survival equation is estimated for 213 plants (mostly failed). Results indicate UAW-organized plants significantly more likely to have closed.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
industries and firms where potential profits are high, biasing toward zero the (negative) union coefficient in OLS profit equations. Voos and Mishel (1986a) find evidence of simultaneity bias, obtaining estimates of a 35% reduction in PCM when unionism is endogenous, as compared with 23% with OLS (Hirsch and Connolly [1987] provide similar qualitative evidence). Substantial measurement error in the union variable may arise from matching three-digit industry union measures to four-digit industry observations or to individual firms. Confidence in the union-profits results has been enhanced, however, by studies measuring union coverage at the line of business or company level (Clark 1984; Abowd 1987; Hirsch, in press). Finally, there is a potential selectivity problem. The data reflect the performance only of those establishments that have survived unionization; thus, union effects on surviving firms may be a biased estimate of actual union effects on nonsurviving firms and potential effects on current nonunion firms.

Ruback and Zimmerman’s (1984) events analysis, and the closely related study by Bronars and Deere (1987), provide perhaps the most direct measures of the expected effect of union organizing and representation on market values. Inferences drawn from these studies must be qualified, however. The union elections (covering at least 750 workers) typically do not include the entire firm, the profit effects for new (marginal) unionized firms may not be representative of union effects in existing companies, and results are highly variable across firms (Ruback and Zimmerman find that equity values increased in 35% of the firms where unions won representation).

Although the magnitude of the profit reduction due to unionization cannot be identified precisely, it does appear to be substantial. This in turn raises the difficult issue of the means through which such significant profit reductions are effected. To address this issue, we must first identify those firms or industries most affected by unions and, second, locate the sources of union rents. To the extent that unions reduce profits below competitive levels, the long-run survival of union firms is threatened. For this reason, the most likely sources of union rents are economic profits deriving from market power and quasi rents associated with special firm advantages or long-lived physical and intangible capital investments. Note that monopoly profits need not be fully appropriable where firms can transfer, at low cost, production to a nonunion environment. On the other hand, appropriable quasi rents need not be associated with monopoly power or supra-competitive returns; rather, they may simply represent normal returns to long-lived investments (Klein, Crawford, and Alchian 1978).

Much attention has focused on the relationship between unions, profits, and market structure. Clark (1984) finds for the PIMS sample that the large reduction in profitability associated with unionization is entirely restricted to businesses with low market shares despite much higher profit rates in firms with high market shares. Clark rationalizes this surprising
result by postulating a transmission mechanism by which roughly equal
union wage increases are generated throughout an industry, with large
percentage profit effects on low market share, low-margin firms and neg-
ligible effects on high market share, high-margin firms. This result is not
corroborated elsewhere (e.g., Hirsch and Connolly 1987) and would seem
to be specific to the PIMS sample of businesses.

In what appears to be the leitmotiv of this literature, Freeman (1983),
Karier (1985), and Salinger (1984) conclude that negative union profit
effects are restricted to highly concentrated industries (but see Domowitz,
Hubbard, and Petersen 1986; Voos and Mishel 1986b). Freeman and Karier
base their conclusions on industry PCM equations in which an industry
unionization variable interacts negatively with concentration (or concen-
tration dummies). Both authors report large and significant profit reduc-
tions in highly concentrated industries and an absence of profit reduction
in lowy concentrated industries. Salinger's evidence is less straightfor-
ward. He estimates a nonlinear least squares profit equation (with $q$ as the profits
measure) in which industry unionism enters in a three-way interaction
with concentration and a vector of profit determinants. His union coef-
efficient can be treated as the “tax rate” on the returns produced jointly by
concentration and the profit determinants. He, too, concludes that unions
capture most of the monopoly profits associated with concentration.

The above studies appear to show that union labor captures a significant
share of potential monopoly profits, while having little effect in relatively
more competitive industries. They also establish that the omission of union
variables in traditional profit models is a serious flaw. Despite the appeal
of these findings, the conclusion that industry concentration provides the
major source for union rents would appear to be incorrect (for a detailed
critique, see Hirsch and Connolly [1987]). For example, it is far from
settled whether concentration does in fact produce any excess profits from
which unions can extract rents. Few firm-level analyses find concentration
to be a significant determinant of profits, and many find a negative relation
between the variables (Bothwell, Cooley, and Hall 1984). 9 Indeed, Ra-
venscraft (1983) and others have shown that the positive coefficient on
concentration in industry-level profit studies may instead capture the effect
of firm-level market share. Using firm or line-of-business-level data, market
share is an important positive determinant of profits, whereas concentration
is, if anything, negatively related to profits.

Moreover, if concentration is the major source for union gains, there
should exist corroborative evidence from the labor market indicating a

---

9 Salinger finds little evidence that concentration generates profits for unions to
capture yet interprets his results in a manner similar to Freeman (1983) and Karier
results using a similar data base. They show that Salinger’s functional form is
unduly restrictive and that there is little support for the contention that concen-
tration-related profits provide a source for union rents.
larger union-nonunion compensation differential in more concentrated industries (Hirsch and Connolly 1987). Yet earnings function studies generally find little evidence that concentration increases either union or nonunion wages, and there exists no significant microlevel evidence of larger union premiums in more concentrated industries (Lewis 1986, pp. 154–55). If anything, union premiums appear somewhat smaller in more concentrated industries (e.g., Mellow 1983). And firm-level evidence on Tobin’s \( q \) and \( r \) (Hirsch and Connolly 1987) suggests that the results obtained in earlier studies may have been sensitive to specification and omitted variables.\(^{10}\)

Recently, attention has shifted away from a focus on market structure as a source of union gains toward quasi rents on firm investments. In Baldwin’s (1983) theoretical model, it is argued that when the capital cycle is long relative to the union’s time horizon, the surplus that provides the return on durable and specialized capital is vulnerable to capture by organized labor. Once such capital is in place, the firm will share quasi rents with the union rather than allow a strike to shut the firm down. Baldwin contends that, in response, a union firm is likely to keep in place inefficient capital as a means of moderating union wage demands. Subsequent papers have emphasized that union behavior is unlikely to be compatible with efficiency; that is, the joint wealth of shareholders plus current and potential future union members will not be maximized as long as union membership is nontransferable.

Connolly, Hirsch, and Hirschey (1986) extend Baldwin’s rent-seeking model to the case of intangible capital, treating unions as a tax on the returns from nontransferable research and development (R&D) capital. Consistent with that model, R&D investments are found to add less to the market value of firms in highly unionized industries than to the value of firms in lowly unionized industries. Lawrence and Lawrence (1985) develop a model in which firms in a declining industry with high fixed costs from long-lived capital will face increased union bargaining power. Their argument is that union bargaining power may be enhanced in the short run by a secular decline in demand. Such a decline may decrease the elasticity of labor demand due to a decrease in the elasticity of substitution stemming from the inability to substitute lumpy labor-saving capital for union labor. The model is applied specifically to the steel industry to explain why an increased union wage premium accompanied falling steel output

\(^{10}\) Hirsch and Connolly (1987) conclude that returns to market share, R&D, and limited foreign competition may be more likely sources of union rents than is concentration. Their market share results conflict with Clark’s (1984) rather counterintuitive findings. Note that a likely source for union gains are rents made possible by government protection and regulation; for example, in industries such as postal delivery and, prior to deregulation, airlines and trucking.
during the seventies, followed by rapid employment declines during the eighties (see, relatedly, Linneman and Wachter [1986]).

A prediction of the union rent-seeking models is that the union tax on a firm’s quasi rents will affect investment behavior and, subsequently, that firm’s long-run economic performance. We return to these themes in the next section. But note that evidence supporting strong efficiency or value maximization, given levels of quasi rents (Abowd 1987), does not imply long-run allocative efficiency if union bargaining power affects firm investment behavior.

Finally, it would be particularly illuminating to effect a comparison of survival rates between union and nonunion firms. Unfortunately, virtually no information is available to us in this area. An exception is the limited analysis by Kaufman and Kaufman (1987) of the automotive engine and body parts industry, concluding that firms organized by the United Auto Workers (UAW) are significantly more likely to have failed than nonunion and non-UAW union firms. This finding is consistent with estimates derived from their sample of surviving firms of a large UAW compensation (but not productivity) premium relative to non-UAW and nonunion firms.

The above evidence points unambiguously to lower profitability under unionism. Despite a variety of statistical problems that tend to bias estimated profit effects toward zero, unions’ true impact is likely to reside within the range of estimates presented in table 1. The productivity profits puzzle stated at the outset of this section—that evidence of large union productivity effects is incompatible with evidence of large profit reductions—may be more apparent than real. Our reading of the evidence is that reductions in profitability seem likely to result from a combination of union wage increases and rather small (and possibly negative) union effects on productivity. (Note again the variability in each effect across firms, industries, and time.) Although it is not in doubt that unions reduce profitability over long periods of time, competitive pressures in labor, product, and capital markets must ultimately limit the profit spread between union and nonunion firms or lead to reductions in the size of the union sector. The steep decline in private sector unionism may have resulted in no small part from the relatively poor profit performance of U.S. companies during the 1970s. The magnitude of union-nonunion differences in economic performance during the 1980s awaits detailed empirical scrutiny.

IV. Unions, Investment Behavior, and Productivity Growth

The traditional on-the-demand-curve model treats a union wage increase as an exogenous increase in the price of labor and a relative decrease in the price of capital. Profit-maximizing firms respond by freely varying their input mix, increasing the relative use of capital and higher-quality labor. Investment activity may increase or decrease, depending on the relative strength of scale and substitution effects. By contrast, the union rent-
seeking model views union wage increases as an outcome made possible by a tax on the firm’s returns or quasi rents accruing from market power and long-lived investments. A firm facing a powerful union is less likely to put in place new long-lived capital (even of a labor-saving type), because a union with a credible strike threat can appropriate a share of the normal returns from that investment. In the face of effective rent seeking, firms reduce investments in tangible and intangible capital whose returns are most vulnerable to union capture.11

Empirical evidence with direct bearing on the union rent-seeking hypothesis remains meager. Baldwin (1983) contends that her theoretical model applies particularly well to the steel industry (see, also, Lawrence and Lawrence 1985). Connolly, Hirsch, and Hirschey (1986) hypothesize that firms in highly unionized industries will invest less in R&D because such investments add less to the market value of union than nonunion firms. Based on evidence from a 1976 sample of 367 Fortune 500 firms, they find that firms in highly unionized industries have significantly lower R&D intensity (R&D investment divided by sales). Union and union-profitability interaction variables enter negatively in their R&D intensity equation, suggesting that unions decrease R&D investment both directly and also through a redirection of earnings away from vulnerable R&D.

Two recent studies also conclude that innovative activity is less pronounced in unionized firms. Acs and Audretsch (1987) use 1982 manufacturing data and find significantly fewer large- and small-firm innovations in highly unionized industries.12 Hirsch and Link (1987) provide similar evidence from a survey of 315 small- and medium-sized New York state manufacturing businesses in 1985. Their ordered probit models, in which Likert-scale responses to survey questions on the importance of product

11 In addition to the papers by Baldwin (1983), Lawrence and Lawrence (1985), and Connolly, Hirsch, and Hirschey (1986), see Grout (1984) and Tauman and Weiss (1987). Using a Nash bargaining model, Grout shows that firms facing a union will have lower levels of investment when contracts are nonbinding than when they are binding. Tauman and Weiss develop a duopoly model in which either the union or nonunion firm may choose the most productive technology, depending on the set of assumptions chosen. In the case where the union chooses a wage after firms select their technologies, the union firm is unlikely to adopt the most productive technology. Note that even if cooperative (i.e., jointly maximizing) settlements obtain, investment will be below the competitive level if the union’s time horizon is shorter than the expected life of capital. For a fuller discussion, see Hirsch (in press).

12 The number of innovations are calculated separately for large and small firms within industries. The data were provided by the Small Business Administration, which defines an innovation as “a process that begins with an invention, proceeds with the development of the invention, and results in introduction of a new product, process or service to the marketplace” (Acs and Audretsch 1987, p. 110, n. 1).
innovation (relative to their competitors) are the dependent variables, indicate that innovative activity is significantly less important for union than for nonunion businesses.

Recent work by Bronars and Deere (1986) and Hirsch (in press) offer more comprehensive analyses of union effects on firm investment and financial behavior. Bronars and Deere develop a dynamic model in which joint wealth maximization by shareholders and current union members does not maximize rents accruing to shareholders and all current and future union members. Because current union members cannot be compensated by future union members (e.g., through the purchase of union membership), both they and shareholders have an incentive to shift positive cash flows closer to the present and defer negative cash flows further into the future. Hence, Bronars and Deere predict that unionized firms will have lower tangible and intangible capital investments, offer a higher payout rate to shareholders, and rely more heavily on debt financing. Empirical results from an unbalanced panel of 756 publicly traded firms on the Compustat tapes between 1972 and 1976, matched to industry union coverage data, provide support for these hypotheses.\(^\text{13}\) Other things equal, they find that firms in more unionized industries have (1) lower investments in capital and lower capital-to-labor ratios, (2) lower R&D investments, (3) lower advertising expenditures, (4) higher debt and debt-equity ratios, and (5) higher payouts (dividends and stock repurchases) and a higher ratio of payouts to retained earnings. All union results are highly significant, the magnitude of the union coefficients in the R&D and debt-equity equations being particularly large.

Hirsch (in press) provides a similar analysis for data on 315 Fortune 1,000 firms during the 1970s. This study employs a firm-level measure of union coverage (based on a 1972 survey) and includes detailed industry controls in the regression equations. Hirsch finds that unionized firms invest significantly less in research and development and physical capital than do similar nonunion firms. And in a recent study using the Ruback and Zimmerman (1984) election data, Bronars and Deere (1987) find that firms decrease capital investment and employment, while increasing debt, following union representation elections (regardless of the outcome) but find no conclusive evidence that R&D or advertising expenditures are affected.

Lower profitability and decreased tangible and intangible capital investments among union firms are likely to have a negative impact on the long-run productivity growth of firms. This brings us to what at first appears to be a puzzle in the literature: the sharply differentiated results

\(^\text{13}\) Bronars and Deere (1986) do not include industry dummies or detailed industry control variables in their regressions.
of production function studies indicating positive union productivity effects on the other. Most R&D studies that include a union argument report total factor productivity growth to be negatively associated (but not always significantly so) with the level of unionism (e.g., Terleckyj 1974; Kendrick and Grossman 1980; Mansfield 1980; Link 1981; and Sveikauskas and Sveikauskas 1982; but for an exception, see Clark and Griliches [1984]).

But there need be no inconsistency between the two sets of findings. From equation (3), a positive coefficient on the union variable in the production function test implies that unionized establishments have higher total factor productivity. Accordingly, an attempted reconciliation of the productivity and productivity growth literatures might proceed as follows: unions increase the level of productivity (perhaps through shock effects) but subsequently retard productivity growth, possibly reflecting long-run responses to decreased profit expectations. An alternative but more conventional argument would be that the legacy or heritage effects of union work rules are negative. This is the argument espoused by Maki (1983), in an analysis of the growth of total factor productivity in Canadian manufacturing during the period 1926–78. Ignoring the error term, Maki estimates the equation

$$\rho = a_0 + a_1 P + a_2 \Delta P + a_3 S,$$

where $\rho$ is the annual growth rate of total factor productivity, $P$ is union density, and $S$ is days lost through strikes divided by total employment (we ignore control variables). The first union term is supposed to pick up the long-term effects of unionism on growth, while first differences in $P$ are intended to proxy impact or shock effects. Maki finds that the shock effects are positive and the long-term effects are negative. For the various specifications, it takes from 5 to 8 years (coefficient $a_2$ divided by $a_1$) for the one-time positive effect of an increase in unionization to be offset by the continuing long-term effect of unionism in slowing growth.

In a related analysis of U.S. productivity growth in two-digit manufacturing industries from 1957 to 1973, Hirsch and Link (1984) show that if one specifies the standard Cobb-Douglas model in difference form, then changes in total factor productivity, $\rho$, should be a function of changes in union density, $\Delta P$ (see eq. 3). Since Hirsch and Link employ a three-

---

14 Hirsch (in press) finds productivity growth during the 1970s to be slower for companies with high levels of firm union coverage in 1972, although much of the slower growth can be accounted for by controlling in detail for industry effects. The union coefficient in a productivity growth equation measures only the direct union effect, however; unions affect growth indirectly through their impact on investments in capital and R&D.
factor Cobb-Douglas function to include technical capital, their estimating equation (ignoring control variables) is

$$\rho = \lambda + \Phi(dT/dt)/Q + (1 - \alpha)(c - 1)\Delta P,$$

(12)

where $\lambda$ is the rate of disembodied growth, $\Phi$ is $\partial Q/\partial T$ (the marginal product of technical capital), $dT/dt$ approximates net investments into stock $T$, and $(dT/dt)/Q$ is proxied by R&D intensity. They report a negative rather than a positive coefficient on the union change variable, contrary to the prediction of the collective voice model. When Hirsch and Link also include a level of unionism variable, both $P$ and $\Delta P$ are negatively and significantly related to productivity growth. The inference is, then, that unionism not only reduces total factor productivity (as reflected in the sign on $\Delta P$) but also slows the rate of productivity increase.\(^{15}\)

Finally, note that studies estimating union effects on productivity growth attempt to hold constant changes in stocks of R&D and physical capital. The specific routes through which unions slow productivity growth have not been examined thoroughly. To the extent that unions decrease investment in tangible and intangible capital, these studies understate the negative impact of unionism on firm and industry growth. A full accounting of the union role requires measurement not only of unions' direct effects on productivity growth, but also of any indirect effects deriving from their impact on investment decisions (Allen 1986c; Hirsch, in press).\(^{16}\)

The rather divergent conclusions reached in the production function and R&D literatures cast further doubt on the hypothesis that unions raise productivity. If unionized firms and industries are significantly more productive, ceteris paribus, should not there be evidence of faster productivity growth in industries and firms where unionization has grown most or declined least? Evidence of lower investment, reduced innovative activity, and slower productivity growth in more unionized sectors should not only engender caution in evaluating the evidence from production function studies, but also should give rise to concern about the long-run impact of unions on economic performance. Perhaps the long run has arrived.

\(^{15}\) We are reluctant to place undue emphasis on the Hirsch and Link findings because of the level of aggregation, the limited number of observations, the possible endogeneity of unionism, and the fragility of the conventional production function test. In an analysis of productivity growth in construction, however, Allen (1986c) also finds productivity change inversely related to both the initial level and change in unionization. Freeman and Medoff (1984, pp. 169–71) contend that the relationship between productivity growth and unionism is too weak to draw any definite conclusions.

\(^{16}\) It is difficult to draw firm conclusions about unionism's economy-wide impact, since decreased investment and growth among union firms frees resources for non-union firms. Such resource movement is neither instantaneous nor costless, however; if it were, there would not exist such significant profit differences between union and nonunion companies.
V. Conclusions

Our analysis of the burgeoning literature investigating union effects on productivity, profits, and growth has exposed a number of unresolved issues. Statistical shortcomings and considerable ambiguity attach to the production function tests adopted in the bulk of the productivity studies. While empirical evidence is mixed, the contention that unions, on average, significantly raise productivity cannot be sustained. Several empirical regularities in this literature, however, may be exploited. First, positive union productivity effects tend to be most pronounced in industries where union-nonunion wage differentials are large. Second, positive effects are more evident where competitive pressures are present. And, relatedly, positive effects are absent or quite small in the public and less competitive sectors— with or without a union wage premium. Although the link between these observed regularities is somewhat shaky, since unions are least likely to obtain large wage premiums in more competitive settings, the findings are broadly consistent with both shock effect and selectivity explanations. Firms in competitive environments are “shocked” into productivity improvements in the face of wage increases and decreased profit expectations. And competitive pressures ensure that in the very long run, firms must increase productivity in order to survive. Thus, any cross-section sample of firms will be nonrepresentative, since unionized firms not increasing their productivity are least likely to survive.

The finding that unions reduce profitability further calls into question the sizable union productivity differentials reported in some of the production function studies. Freeman and Medoff argue that unions reduce profitability in general because their productivity effects, though substantive, are nevertheless insufficient to offset increases in wage costs and greater capital intensity. To this is added the rider that since union profit effects are largely restricted to highly concentrated industries the longer-term consequences of this transfer are benign. Our own reading of this literature is that many of the productivity and profit results are simply inconsistent; “true” productivity effects are more likely to be close to zero and negative profit effects may be less severe than suggested in the empirical literature. The conclusion that unions capture concentration-related profits is based largely on a fragile econometric relationship between profitability, concentration, and unionism at the industry level. Corroborative evidence from the labor market is lacking. Returns accruing from other correlates of market power (e.g., market share, foreign competition, and government entry restrictions) and from long-lived capital appear to be more important sources of union rents.

The manner in which union rent seeking proceeds is of some importance. Such evidence as we have been able to uncover does not encourage an optimistic view of unionism’s longer-term consequences. Union rent seeking at the expense of long-lived tangible and intangible capital appears to
lower firms’ investment in physical capital, as well as to decrease R&D and other innovative and risk-taking activities. As a consequence, productivity growth tends to be slower in unionized firms and industries. Increased management opposition to unions, and declining union coverage and employment within most sectors of the U.S. economy, appear to be predictable responses to the relatively poor performance of highly unionized companies during the 1970s. More work is required here, particularly in the construction of improved data sets. But already the locus of the debate has shifted toward a consideration of unionism’s dynamic effects on long-run economic performance.

References


———. “Productivity Levels and Productivity Change under Unionism.” Mimeographed. Raleigh: North Carolina State University, March 1986. (c)


Boal, William March. “Unionism and Productivity in West Virginia Coal


Lawrence, Colin, and Lawrence, Robert Z. "Manufacturing Wage Dis-


