THE WORKER DISCIPLINE EFFECT: A DISAGGREGATIVE ANALYSIS

Francis Green and Thomas E. Weisskopf*

Abstract—We test for the presence of a “worker discipline effect,” wherein macroeconomic conditions influence worker effort, and we examine inter-industry variation in its strength. An employment function analysis is first used to find evidence of a worker discipline effect in the majority of U.S. 3-digit manufacturing industries. A factor analysis of industry firm and labor market characteristics is then used to identify several underlying factors by which industries can be distinguished. We find that the strength of the worker discipline effect is positively and significantly correlated with the degree to which industries have “secondary” characteristics.

I. Introduction

The relationship between macroeconomic conditions and the rate at which workers expend effort in production has been studied from a range of perspectives. While the idea that unemployment acts as a discipline on workers affecting their work intensity relates back to Marx's concept of the “reserve army of labor” and his distinction between “labor” and “labor power,” it has recently been modelled both from an information-theoretic neoclassical perspective (Shapiro and Stiglitz, 1984) and from a neo-Marxian viewpoint (Bowles, 1985).1

In empirical studies this relationship has been used to help explain the slowdown in U.S. productivity growth (Weisskopf, Bowles and Gordon, 1983), the incidence of strikes in the United States (Schor and Bowles, 1987) and the rise in work intensity in the United Kingdom in the 1980s (Schor, 1987). International comparisons suggest that the effect of unemployment on productivity varies considerably in importance across countries, with the effects being strongest where industrial relations are the most confrontational—as in the United States (Weisskopf, 1987). Studies of U.S. manufacturing have suggested that the effect varies also across industries, being weaker where unionization and long-term employment relations are most pronounced (Oster, 1980, and Rebitzer, 1987).

In this paper we undertake further tests for the existence of a disciplinary effect of unemployment on work intensity and productivity, and we investigate how the strength of such a “worker discipline effect” varies across industries. We apply the basic employment function framework utilized by Oster (1980) at a 3-digit level of disaggregation, and we deploy a rich set of industry characteristic data in order to analyze systematically the relationship between the size of the worker discipline effect and various industry characteristics—including, in particular, the extent to which an industry can be considered “primary” as opposed to “secondary.”

II. Theory

The argument that unemployment can serve to discipline workers in such a way as to increase their work intensity needs to be qualified in order to obtain a more realistic account of worker motivation in modern economies. While the ultimate threat of dismissal is never completely absent, alternative negative sanctions are frequently applied—such as the withholding of wage increases, denial of promotion, imposition of fines, or sometimes demotion to less skilled jobs. Workers may also be motivated to work hard by means of various positive incentives, such as regular wage and benefit increases, favorable working conditions, and/or a corporate ideology that promotes high worker morale and calls forth loyalty to the company’s aims. To examine the ways in which unemployment may affect work intensity and labor productivity, we begin by formulating a simple model.

Let a firm’s production function be

\[ Q = f(K, L) \quad f_K, \quad f_L > 0 \quad (1) \]

with the usual notation except that \( L \) refers to...
“effective labor” measured in efficiency units. We define work intensity, the average effective labor input per hour of labor employed, as

\[ h = \frac{L}{H} = h(X, Y, Z) \quad h_X, h_Y, h_Z \geq 0 \]  

(2)

where \( X \) is the negative sanction against shirking arising from the possibility of job loss, \( Y \) is a set of variables representing other negative sanctions against shirking, \( Z \) represents the array of positive incentives for greater work effort, and \( H \) is the number of labor-hours employed. The overall rate of unemployment is likely to affect work intensity through its impact on \( X \), in a way that differs according to certain aspects of the firm’s work environment.\(^2\) Drawing on recent “effort regulation” models (Rebitzer, 1987; Bowles, 1985), we may write

\[ X = X(\text{PC}, \text{PD}, W^*) \quad X_{\text{PC}}, X_{\text{PD}}, X_{W^*} > 0 \]  

(3)

where \( \text{PC} \) is the probability of being caught if a worker shirks, \( \text{PD} \) is the probability of being dismissed if so caught, and \( W^* \) is the cost associated with job loss. \( \text{PC} \) can be expressed as follows:

\[ \text{PC} = \text{PC}(S, T) \quad \text{PC}_S > 0; \text{PC}_T \leq 0 \]  

(4)

where \( S \) is the ratio of supervisors to production workers, and \( T \) represents the complexity of the technology being used. (Whether more complex technology implies ceteris paribus a higher or lower probability of being caught is immaterial to our subsequent analysis). \( \text{PD} \) is analyzed as

\[ \text{PD} = \text{PD}(E, R) \quad \text{PD}_E, \text{PD}_R < 0 \]  

(5)

where \( E \) represents the costs associated with finding and training a replacement for a dismissed worker, and \( R \) captures the ability of workers individually or collectively to resist dismissal by imposing further costs on the employer. Standard search theory implies that

\[ E = E(U) \quad E_U < 0 \]  

(6)

where \( U \) is the unemployment rate. Finally, assuming a single-period framework, the cost of job loss to a worker is given by

\[ W^* = W - r(U)W^a - \left[ 1 - r(U) \right]B \quad r_U < 0, \quad W^a > B \]  

(7)

where \( W \) is the worker’s present real wage, \( W^a \) is the alternative real wage if the worker is hired by another employer, \( r \) is the probability of being so hired, and \( B \) is the real welfare benefits received when unemployed (assumed to be less than the alternative real wage). Combining equations (2) through (7) we obtain the impact of unemployment on work intensity:\(^3\)

\[ \frac{dh}{dU} = h_X\left[ X_{\text{PD}}PD_EE_U - X_{W^*}r_U(W^a - B) \right]. \]  

(8)

It follows from the signs of the relevant partial derivatives (and \( W^a > B \)) that \( \frac{dh}{dU} \) is unambiguously positive whenever \( h_X > 0 \), but the magnitude of the effect depends on the magnitudes of the partial derivatives and \( (W^a - B) \).

The recent theoretical literature on worker motivation and effort regulation is divided on a critical issue: whether the worker discipline effect plays a more important role in primary or secondary sector industries. On the one hand, Bulow and Summers (1986) have developed a theory of dual labor markets based upon the premise that only in the primary sector is monitoring of labor costly and difficult; in the secondary sector workers’ effort is perfectly observed and workers are indifferent between employment and unemployment. Bulow and Summers are assuming (in our notation) that in the secondary sector \( \text{PC} = \text{PD} = 1 \) and \( W^* = 0 \), for any \( U \), so that \( \frac{dh}{dU} = 0 \); their approach implies that in the primary sector \( \frac{dh}{dU} \) is positive and significant.

In sharp contrast, Edwards (1979) has argued that the primary sector work effort is elicited mainly by the positive motivation associated with

\(^2\) We see no reason to hypothesize that either \( Y \) or \( Z \) is systematically influenced by the rate of unemployment.

\(^3\) We assume that supervision intensity, technology, and the ability to resist dismissal change only slowly and not with the business cycle; hence \( dS/dU = dT/dU = dR/dU = 0 \). We also assume that the real wage \( W \) is not affected by short-run variation in the rate of unemployment \( U \). In some effort-regulation models firms have been assumed to adjust \( W \) in response to changes in \( U \), in such a way that low unemployment equilibria are associated with high wages and high work intensity. We believe, however, that cyclical variations in unemployment are largely disequilibrium phenomena and that wages, for institutional reasons, vary little over the cycle; indeed, Akerlof and Yellen (1985) have provided theoretical support for this presumption. In any event, as Rebitzer (1987, footnote 7) argues, conclusions about the relative magnitude of the worker discipline effect in different industries are not altered by the assumption that \( dW/dU = 0 \).
“bureaucratic control,” so that the threat of dismissal is used much less frequently than in the secondary sector. Edwards’ analysis implies (in our notation) that in the primary sector \( h_2 \) is higher, and \( dh/dU \) are lower, than in the secondary sector. So too, with Akerlof (1982), whose conception of labor market duality is based on employer use of the dismissal threat but on worker motivation to donate collective work effort above the required minimum in return for the gift of wage premia in the primary sector. Unlike Edwards and Akerlof, Rebitzer (1987) models worker motivation exclusively within the framework of the threat of dismissal; but he also reaches the conclusion that primary sector industries (as the locus of long-term employment relations) will tend to have lower values of \( h_X \) and \( dh/dU \) than secondary sector industries.\(^4\)

In the following section we will use our empirical estimates of the size of \( dh/dU \) in different types of industries to shed light on the applicability of the analyses of Bulow and Summers, on the one hand, and Edwards, Akerlof and Rebitzer, on the other, to the U.S. manufacturing sector.

III. Empirical Analysis

(a) Worker Discipline Effects

In order to investigate how the impact of unemployment on work intensity and productivity varies in magnitude across industries and sectors, we begin by estimating that magnitude for each industry, using a conventional employment function framework.\(^6\) Assuming cost minimisation and a log-linear specification for the production function (1) and adding terms in time \( t \) to capture trend labor-saving technical progress, we obtain the equation for desired labor hours \( H^* \):

\[
\log H^*_t = a_0 + a_1 \log Q_t + a_2 \log K_t + a_3 t \\
+ a_4 t^2 + a_5 \log U_t + a_6 \log H_{t-1}. 
\]  

(9)

Assuming a constant speed of adjustment \( s \) from actual to desired labor-hours, and, on the basis of (8), replacing work intensity \( h \) with the unemployment rate \( U \), we obtain the estimating employment function:

\[
\log H_t = b_0 + b_1 \log Q_t + b_2 \log K_t + b_3 t \\
+ b_4 t^2 + b_5 \log U_t + b_6 \log H_{t-1}. 
\]  

(10)

From the estimated coefficients we can derive an estimate of the speed of adjustment \( s \) and a measure of the magnitude of the worker discipline effect (labelled \( WD \)) as follows:

\[ s = (1 - b_6) \]

and

\[ WD = d \log h/d \log U = -b_5/s. \]

Equation (10) was estimated for each of the 143 3-digit U.S. manufacturing industries, using annual data from 1958 to 1981.\(^7\) The estimates were repeated using an alternative measure of labor market slack—the mean level of unemployment duration for job losers \( UD \)—instead of \( U \), since it is possible that this measure more closely captures the probability of a job loser getting rehired within a given period of time.\(^8\) Since our

\(^4\) A parallel analysis by Williamson et al. (1975) examines the advantages of “consummate cooperation,” described as “an affirmative job attitude—to include the use of judgment, filling gaps, and taking initiative in an instrumental way,” (p. 266), when jobs are idiosyncratic, as they are most likely to be in the primary sector.

\(^5\) Rebitzer argues that \( dh/dU \) is lower in the primary than in the secondary sector because (i) \( PD \) (the probability of dismissal) is less affected by unemployment—since the grievance and arbitration procedures negotiated by unions diminish the impact of changes in replacement cost on the probability of dismissal and/or the costs of reneging on implicit contracts make it unwise to raise the rate of dismissal when unemployment is high (Lazear, 1981); and (ii) \( W^* \) (the cost of job loss) is higher, and \( X_{W^*} \) is negative, so \( X_{UP} \) is lower. Rebitzer assumes somewhat unrealistically that \( W^o \) (the alternative real wage) and \( B \) (real unemployment benefits) are invariant across sectors; although it seems likely that \( (W^o - B) \) would be greater in the primary than in the secondary sector, it also seems unlikely that this effect could outweigh the other factors implying a lower \( dh/dU \) in the primary sector.

\(^6\) Our approach follows broadly the same methodology as Oster (1980), but it differs in being more disaggregated, in covering a more recent period and in utilizing a richer set of industry characteristics in the second stage of the analysis.

\(^7\) If (10) were estimated at an aggregate level, it might be argued that the unemployment coefficient \( b_5 \) would reflect cyclical labor hoarding in response to cyclical variations in demand conditions, to the extent that this was not already captured in the coefficient \( b_6 \), though the modeling of the lagged adjustment of actual to desired labor hours. Note that any such effect would bias downward the absolute magnitude of the estimated worker discipline effect. The greater the level of disaggregation, however, the less is the linkage between industry demand conditions and aggregate unemployment; thus there is little reason to believe that for 3-digit manufacturing industries the estimate of \( b_5 \) will be seriously contaminated by cyclical demand effects.

\(^8\) We tried as further alternatives to the overall civilian unemployment rate the unemployment rate for experienced workers, and the unemployment rate including the armed forces; but these made no notable difference to the results.
main objective at this stage was to compare the worker discipline effect in different industries, rather than to obtain the best possible specification for the hours equation in any one industry, the same specification (9) was imposed, after some experimentation, on all 143 industries. The coefficient estimates for \( b_0, b_1, b_2, b_3 \) and \( b_4 \) ranged over values broadly similar to those normally obtained with employment functions, and, as usual, the estimated equation generally gave a good explanation of industry labor-hours.

The results of greater interest here are summarized in table 1, which focuses on the estimation of the magnitude of the worker discipline effect \((WD)\). The evidence in the first row of the table shows that in the vast majority of industries the estimate of \( WD \) is of the expected sign, irrespective of which measure of labor market slack is used. The second and third rows provide evidence of the frequency with which the estimates of \( WD \) are both of the expected sign and statistically significant.

11 The significance levels of worker discipline effects summarized in table 1 are based on the \( t \)-statistics for the industry regression coefficient estimates of \( b_5 \). Strictly speaking, one ought to adjust this test to take into account the role of \( b_6 \) in the determination of the full worker discipline effect; but since we are not concerned with a precise statistical inference in any one particular industry there was no need to make this adjustment in the present context. We applied one-tailed tests, since our hypothesis is for a positive worker discipline effect. In a very few cases (3 out of 143 for the unemployment rate regressions, 2 out of 143 for the unemployment duration regressions) the coefficient was of the "wrong" sign and significant at 5%. We find it more reasonable to interpret these few cases as statistical outliers than as instances in which low unemployment actually leads to high work intensity; the latter is a theoretical possibility in effort-regulation models only if wages are assumed—quite unrealistically—to be continuously adjusted to their optimal level in response to changes in labor market conditions.

12 It might be counter-argued that estimates of a negative \( b_5 \) are merely attributable to adverse selection, in that rising unemployment allows firms to lay off the least productive workers in their work force. If so, one would expect that the strength of what we have interpreted as the worker discipline effect would tend to be highest in industries for which the unemployment rate \( U \) was most strongly negatively associated with annual industry employment growth. To test for this possibility we computed the cross-industry correlation of (1) the estimated worker discipline effect \( WD \) with (2) the intra-industry correlation of \( U \) with employment growth; a negative cross-industry correlation would support the counter-argument. In fact, the resulting correlation coefficient was insignificant even at the 20% level.
worker discipline effect to different types of industries.

(b) Industry Characteristics

To carry out an inter-industry analysis of the relationship between the size of the worker discipline effect and various industry characteristics, we need quantitative estimates of a variety of characteristics. Table 2 lists characteristics for which we have been able to obtain point estimates for 100 3-digit U.S. manufacturing industries. The industry characteristics are grouped into two categories: those associated with the types of firms and product markets prevalent in an industry, and those associated with the predominant labor relations and types of jobs in the industry.13 There is a high degree of correlation among many of these variables, so it is virtually impossible to distinguish statistically the separate relationship of each of them with the strength of worker discipline effects. We therefore conducted a factor analysis to describe the industry characteristic data set in an economical fashion and to isolate a small number of identifiable underlying factors.

13 We are extremely grateful to David Howell for providing us with his figures for the first 18 industry characteristics listed in table 2. In basing our subsequent analysis on point estimates for these industry characteristic variables, we are assuming (reasonably) that the values of the variables have not varied so much over the time period 1958–81 as to alter significantly the ordering of industries with respect to the characteristics.
Table 2 presents the results of our analysis, which isolated four distinct factors. The first factor is identified as **PRIMARY** because it is highly positively correlated with characteristics normally associated with the primary sector of the economy: large firm size, concentrated product markets, high wages and high union coverage. Appropriately, this factor is negatively correlated with the quit rate—a key indicator of secondary sector status. The second factor is labelled **PHYSICAL** because it is most strongly associated with two variables reflecting the physical demands of a job; industries with a high score (e.g., logging camps, blast furnaces) tend to have a large proportion of jobs requiring heavy physical activity and/or suffering from unusually high injury and illness rates. Not surprisingly, the proportion of female workers loads strongly and negatively on this second factor. The third factor—**SKILLED**—is strongly associated with variables reflecting job skill requirements, such as indexes of the level of general educational development and the level of specific vocational preparation associated with jobs, as well as the proportion of non-production workers in the labor force and the real wage level. The fourth factor—**MECHANIZED**—is strongly positively correlated with characteristics associated with a highly mechanized production process, such as high energy use and high capital intensity.

---

**Table 2** presents the results of our analysis, which isolated four distinct factors. The first factor is identified as **PRIMARY** because it is highly positively correlated with characteristics normally associated with the primary sector of the economy: large firm size, concentrated product markets, high wages and high union coverage. Appropriately, this factor is negatively correlated with the quit rate—a key indicator of secondary sector status. The second factor is labelled **PHYSICAL** because it is most strongly associated with two variables reflecting the physical demands of a job; industries with a high score (e.g., logging camps, blast furnaces) tend to have a large proportion of jobs requiring heavy physical activity and/or suffering from unusually high injury and illness rates. Not surprisingly, the proportion of female workers loads strongly and negatively on this second factor. The third factor—**SKILLED**—is strongly associated with variables reflecting job skill requirements, such as indexes of the level of general educational development and the level of specific vocational preparation associated with jobs, as well as the proportion of non-production workers in the labor force and the real wage level. The fourth factor—**MECHANIZED**—is strongly positively correlated with characteristics associated with a highly mechanized production process, such as high energy use and high capital intensity.

---

The size of the worker discipline effect, using either measure of WD, is very significantly negatively correlated with the **PRIMARY** factor; this supports the hypothesis that negative sanctions are likely to be most effective and most widely employed in secondary industries. Further confirmation is provided by examination of the correlations with specific industry characteristic variables from which the factors were extracted.

The general education development (GED) index measures the reasoning and the mathematical and language development required for jobs prevalent in each industry; the specific vocational preparation (SVP) index measures the time required to train for particular occupations. This SVP index appears to reflect a considerable element of general skills as well as firm-specific training, and it is highly correlated with the GED index; thus it appears that our **SKILLED** factor captures primarily general rather than firm-specific skills.

Our factor analysis results differ from the results of many other efforts to apply factor analytic methods to distinguish between primary and secondary industries (e.g., Dickens and Katz, 1987) in that we have isolated more than one key dimension of industry characteristics; it is for this reason that we have identified separate **SKILLED** and **MECHANIZED** factors, which are closely associated with variables that in other analyses tend to be associated with a **PRIMARY**-type factor.

We also calculated product moment correlations corresponding to each rank correlation presented in table 3; while these showed different degrees of correlation, they described an overall picture much the same as did the rank correlations discussed in the text.
primary sector status as product market concentration ratio, union coverage and the real wage level, and positively correlated with a key indicator of secondary status; the quit rate.

To illustrate the orders of magnitude involved with the worker discipline effect we split the sample of 100 industries into three segments according to their scores on the PRIMARY factor—distinguishing the ten “most primary,” the ten “most secondary,” and the remaining industries in between. Table 4 displays for each of these three segments the simple average of the size of the estimated worker discipline effects. In the top (most primary) segment the worker discipline effect is relatively small, though positive; in the bottom (most secondary) segment it is much larger. On these estimates, if the unemployment rate rose by 50%, work intensity would rise by just over 9% in a typical “most secondary” industry but by only 3% in a typical “most primary” industry.

Returning to the results in table 3, we find that the size of the worker discipline effect is very significantly positively correlated with the PHYSICAL factor. This is further confirmed by the positive correlations with the industry characteristic, “physical demands of job.” Note that the size of the worker discipline effect is not significantly correlated with percentage of female workers, even though this latter industry characteristic loads significantly (negatively) onto the PHYSICAL factor. Thus it is really the physical environment of jobs, and not the composition of the labor force, which is at issue here. Moreover, the orthogonality of factors means that we can not interpret the PHYSICAL factor in terms of traditional primary/secondary distinctions.

We did not initially hypothesize that the magnitude of worker discipline effects would be positively associated with the extent to which jobs are physically demanding. To explain our findings in this regard one might suggest that positive means of worker motivation are most successfully employed in situations where working conditions are relatively favorable and comfortable; and one might argue that jobs involving hard physical work, prone to injury, and carried out under generally adverse environmental conditions, would inhibit the development of loyalty to the enterprise, so that the employers would be more likely to rely on negative sanctions (and hence the WD effect) to elicit greater effort on the job. Such a conjecture is consistent with our results, but firm conclusions in this area must await further research.

Our findings show no significant correlations of the size of the worker discipline effect with either the SKILLED or the MECHANIZED factor, nor with any of the industry characteristic variables with which these factors are closely associated. The result in the case of the SKILLED factor might at first appear surprising. One would expect the size of the WD effect to be negatively correlated with this factor insofar as employers are likely to use positive rather than negative motivation strategies with highly-skilled workers and, in cases where the threat of job loss is nonetheless used, the probability of dismissal is likely to be lessened by the greater cost to the employer of replacing more highly-skilled workers. One would expect an opposite (positive) correlation, however, insofar as the threat of job loss is invoked and a highly-skilled worker finds the possibility of unemployment all the more threatening because of a relatively high cost of job loss when dismissed. Our findings raise the possibility that these two hypothesized effects simply tend to offset one another, leaving no significant correlation between the size of the WD effect and the SKILLED factor.
Employers in general have a spectrum of worker motivation strategies to choose from, ranging from positive incentives associated with real wage growth, promotion and better working conditions to negative sanctions associated with the threat of dismissal if workers are caught shirking. Our empirical analysis of U.S. manufacturing industries has provided new evidence for the role of the threat of dismissal as a worker discipline device. In applying models of worker discipline at the 3-digit-industry level of disaggregation for the first time, we have strengthened the conclusions of previous studies that have highlighted the use of negative sanctions associated with unemployment to elicit greater work intensity and productivity in U.S. workplaces.

We have also found strong evidence that the use of the threat of dismissal as a worker discipline device is more prevalent in some U.S. industries than in others. Specifically, our results suggest that worker discipline effects tend to be strongest in industries characterized by low labor force unionization, low product market concentration, and other indicators of secondary sector status. Our findings are consistent with the analyses of Edwards (1979), Akerlof (1982) and Rebitzer (1987), which imply that the responsiveness of work intensity to changes in the dismissal threat is likely to be relatively low in primary sector industries; the findings are quite inconsistent with Bulow and Summers’ (1986) assumption that efficiency-wage strategies operate only in the primary sector.

APPENDIX A

Data Sources

(a) Table 1

Q, real manufacturing gross output, and H, total hours worked by production workers, were derived from an OECD datatpe, “MICRODATA,” which contains 4-digit industry data from the Annual Survey of Manufacturers and corresponding (unpublished) deflators from the Bureau of Economic Analysis. Q was calculated as the deflated value of shipments plus net change in inventories; both Q and H were aggregated from the 4-digit to the 3-digit level for the purposes of our study. K, real capital stock, was obtained at the 3-digit level from a tape prepared by the Office of Business Analysis of the U.S. Department of Commerce. Industries are classified according to SIC 1972. U, the overall civilian unemployment rate, was obtained from the Economic Report of the President (1988). Our UD estimates are for job losers only and for completed spells of unemployment, adjusted to treat spells separated by labor force withdrawals as single episodes, as discussed in Schor and Bowles (1987, pp. 585–586); we are grateful to these authors for providing these data. Over the period 1958–81, the correlation coefficient between U and UD was 0.74.

(b) Table 2

The real wage level, real wage growth, percentage of non-production workers, and growth of non-production workers in each industry (average values over the period 1958–81) were calculated from the relevant annual data on the OECD MICRODATA tape. The other 18 variables apply to 1972 (or nearby years) and are taken from Howell (1982), which may be consulted for details on sources and methods.

REFERENCES


