Optimal Labor Contracts and Involuntary Unemployment

Under Costly and Imperfect Monitoring

Seung Chan Ahn

and

Roger L. Faith

Arizona State University

June 1993

Revised May 1994

Revised February 1995

The first author gratefully acknowledges the financial support of the College of Business Arizona State University. We also thank the referees of this journal for many helpful comments on previous versions of this paper.
This paper provides an explanation for the concurrence of rigid wages and involuntary unemployment. We consider cases in which a firm monitors its workers, but at some cost. A key assumption in the model is that the firm cannot perfectly distinguish shirkers from nonshirkers. Thus, the firm has to rely on negotiated compensation and work effort, as well as monitoring, to reduce the incentive to shirk. We find that rigid wages and involuntary unemployment arise simultaneously when monitoring costs are large and the effectiveness of monitoring is low.
REFERENCES


INTRODUCTION

We often observe that labor contracts specify rigid wages over given contract periods but, over the same periods, allow firms to lay off their workers in response to changing economic conditions. The prevalence of such contracts has been postulated as an important factor in explaining economy-wide fluctuations in employment and output. Azariadis (1974), Baily (1975) and Gordon (1974) --henceforth ABG -- have developed implicit contract theories in order to explain why such contracts could be the consequence of the rational behaviors of economic agents. They consider a simple contract model in which a risk-neutral firm makes a contract with a group of risk-averse workers before the state of nature is realized. Intuitively, less risk-averse firms offer their more risk-averse workers insurance against fluctuations in labor income with a contract which specifies a fixed wage across states of nature. To maintain profitability given fixed wages, firms reduce employment in bad states of nature, giving rise to involuntary unemployment.

While the ABG model provides rich insights into the nature of labor contracts, a disappointing aspect of the model is that it predicts more employment than a spot market equilibrium. (See Cooper [1987].) If severance pay or work effort (or effective work hours) are also negotiated in the contract, the model leads to exactly the same (productively efficient) employment level that would have been observed in the spot market. (See Akerlof and Miyazaki [1980].) Furthermore, when leisure is a normal good, any unemployment is voluntary, in the sense that the unemployed are better off than the employed.
Other contract models have considered cases where firms and workers possess asymmetric information about firms’ profit, workers’ productivity or workers’ utility. Examples in the literature are, amongst others, Grossman and Hart (1981), Azariadis (1983), Chari (1983), Green and Kahn (1983), Hart (1983), Foster and Wan (1984), Cooper (1985) and Moore (1985). If one party (firm or workers) has superior information regarding its status (e.g., worker’s productivity) the better informed party has an incentive to misreveal its true status in order to increase its welfare. In this case, a contract has to include an incentive compatible mechanism which induces all parties to reveal their true statuses. Such contracts usually lead to an inefficient allocation of resources (underemployment) relative to the symmetric information case.

Asymmetric information also provides an explanation for involuntary unemployment. Foster and Wan (1984) and Moore (1985) demonstrate that when workers have superior information about their true productivity or disutility of work, the optimal truth-telling restrictions imposed on a contract lead to involuntary unemployment. However, these models do not explain why wages are rigid (i.e., state-independent) because imposing truth-telling restrictions naturally leads to state-dependent wages.1

This paper provides an alternative explanation for how asymmetric information between contractants could result in rigid wages and unemployment, in particular involuntary unemployment. We consider a bargain over wage, employment, work effort and severance pay. In our model, production is a result of all the workers’ cooperation. Thus, the firm cannot identify each worker’s effort, and workers have an incentive to shirk. To help control shirking, the firm monitors its workers, but only at some cost which the firm regards as a cost
of employment. A key assumption in the model is that the firm cannot perfectly distinguish shirkers from nonshirkers. Thus, the firm has to rely on negotiated compensation and work effort, as well as monitoring, to reduce the incentive to shirk. With these assumptions, we derive the conditions under which rigid wages and involuntary unemployment may arise simultaneously.

The paper is organized as follows. Section I explains the basic model and derives the necessary conditions that characterize an optimal labor contract. We also consider the changes in the optimal contract implied by imperfect monitoring and workers’ incentive to shirk. Section II examines the characteristics of the optimal contract under alternative assumptions with respect to the monitoring technology, and shows that both the effectiveness and cost of monitoring have crucial effects on the properties of the optimal contract. Section III examines the conditions under which the optimal contract can generate involuntary unemployment. Some concluding remarks follow in section IV.

I. THE BASIC MODEL AND ASSUMPTIONS

We consider an extension of the ABG contract model. The model has two periods. In period 1, a firm makes a contract with N homogeneous workers, while actual employment and production occur in period 2. The contract specifies what will happen in period 2. The firm and workers bargain over wage, severance pay to unemployed workers, employment, and each worker’s required effort, under each state of nature. We assume that workers incur prohibitively high mobility costs so that moving from the current firm to another firm in
period 2 is not feasible. We also assume that the firm is risk-neutral and workers are risk-
averse.

We make two important assumptions. First, we assume that the firm bears some extra
cost (C) in addition to wages whenever it employs an extra worker. There are several
interpretations of this employment cost. One is that C represents the administrative cost of
hiring a new worker. Another interpretation is that C represents the costs of training the
worker in firm-specific production skills not available outside the firm. The other
interpretation which plays an important role in our model is based on assumption that in the
absence of monitoring the firm can only observe total output, not an individual worker’s
effort. Consequently, each worker has an incentive to shirk, and the firm must adopt, at a
cost C, a monitoring technology to control shirking.

The second important assumption is that even with monitoring the firm can observe a
shirker only with a given probability, $m \in (0,1]$, which can be thought of as the effectiveness
of the monitoring technology. (See Shapiro and Stiglitz [1984].) A worker’s effort can be
perfectly observed only if $m = 1$. We assume that both $m$ and $C$ remain fixed during the
contract period.²

For simplicity, we assume that the firm uses only labor to produce a single
commodity. Output depends on total employment ($L$) and each worker’s level of effort ($e$),
which are perfect substitutes in the production process. Revenue is given by $sf(eL)$, where $s$
$\in [s_\underline{s},s_\overline{s}]$ denotes a random output demand, or technology, shock in period 2.³ We assume that
$s_\underline{s} > 0$. While $s$ is unknown to the firm and workers in period 1, it becomes public
information in period 2. We also assume:
Assumption 1. \( f(eL) \) is a twice differentiable and strictly concave function defined for \( eL \geq 0 \), with \( f' > 0 \) and \( f'' < 0 \). Furthermore, \( f'(0) = \infty \).

The last condition ensures that the firm will produce a non-zero output in any state, i.e., \( e > 0 \) and \( L > 0 \). To produce, the firm has to pay labor costs consisting of a wage (\( w \)) to an employed worker, a severance payment (\( b \)) to an unemployed worker, and a fixed employment cost (\( C \)) per employee. Therefore, the profit of the firm in state \( s \) is:

\[
\pi(s) = qf(eL(s)) - w(s)L(s) - b(s)(N-L(s)) - CL(s).
\]

Each risk-averse worker has the same utility function, \( U = U(y,e) \), where \( y \) denotes consumption (or income). We assume:

Assumption 2. \( U \) is a twice differentiable and strictly concave function, with \( U_y > 0 \), \( U_e < 0 \), \( U_{yy} < 0 \), and \( U_{yy}U_{ee} - (U_{ye})^2 > 0 \). Furthermore, \( U_{ye} \geq 0 \).

Here, the subscripts denote partial derivatives. If \( U_{ye} < 0 \), the optimal contract we discuss later could have the empirically unrealistic property that a worker’s labor income has to fall, while work effort increases, as the state improves. Assumption 2 rules out this perverse case. Of course, separable utility functions, which have been frequently adopted in contract literature, trivially satisfy the assumption. As we shall discuss later, the assumption \( U_{ye} \geq 0 \) also plays an important role in explaining the underemployment of effort.

We further assume:
Assumption 3. \((U_e/U_y)U_{yy} \geq U_{ye}\), for any \(w\) and \(e\).

Assumption 3 implies that work effort (leisure) is not normal (inferior).\(^7\) When \((U_e/U_y)U_{yy} = U_{ye}\), the supply of work effort is independent of income. This is the case for the utility function, \(U = u(w-v(e))\), where \(u'' < 0\), \(v' > 0\) and \(v'' \geq 0\). Such separable preferences are restrictive, but have historically played an important role in the contract literature. (See, for example, Azariadis [1974, 1983], Grossman and Hart [1981], Hart [1983] and Moore [1984].) Assumption 2 and 3 are quite general in that they subsume most of the utility functions adopted in the contract literature.

For employed workers, consumption in state \(s\) is given by \(w(s)\), and for the unemployed, \(b(s)\), so that \(y(s)\) equals \(w(s)\) or \(b(s)\). For a given \(s\), a worker’s expected utility is:

\[
V(s) = \frac{L(s)}{N} U(w(s),e(s)) + \frac{(N-L(s))}{N} U(b(s),0).
\]

During the production period, the firm monitors its workers. If a worker’s effort is found to be lower than the level specified by the contract, the worker is fired and receives reduced compensation.\(^8\) If a worker shirks and is not detected, the worker continues to receives \(w\). To make the model as simple as possible, we assume that both the shirker’s work effort and compensation, if caught, are zero. Under these assumptions, a potential shirker’s expected utility is equal to \((1-m)U(w,0)+ mU(0,0)\).\(^9\) The firm will choose \(w\) and \(e\) so that the following nonshirking condition (NSC) is satisfied for all \(s\):
Since Assumption 1 ensures that \(e(s) > 0\) and \(L(s) > 0\) for any \(s\), NSC implies that \(w(s)\) must also be positive. Therefore, an optimal contract must specify a positive \(w\), \(e\) and \(L\) for any state.

We now consider the optimal contract. Let \(\delta(s) = [b(s), w(s), e(s), L(s)]\) be the vector of variables to be determined by the contract for a given \(s\); and \(\delta = \{\delta(s) \mid s \in [s_-, s_+]\}\) be a contract which specifies what will happen when each state is realized. Assuming workers are identical, the optimal contract, \(\delta^* = \{\delta^*(s) = [b^*(s), w^*(s), e^*(s), L^*(s)] \mid s \in [s_-, s_+]\}\) solves:

\[
\text{max}_{\delta} \quad E_s[\pi(s) + \theta N V(s)]
\]

subject to

\[
0 < L(s) < N, \text{ for all } s,
\]

and NSC given in (3), where \(E_s\) denotes the expectation operator over \(s\). The parameter \(\theta\) is a positive number whose magnitude depends on the relative bargaining power of a worker.\(^{10}\) (See Azariadis [1983], Kahn [1985] and Cooper [1985]).

The optimal contract \(\delta^*\) must satisfy the first-order (Kuhn-Tucker) conditions (with \(s\) suppressed):

\[
U_{\gamma}(b, 0) = 1/\theta
\]

\[
\phi G = 0, \quad G \geq 0, \quad \phi \geq 0
\]

\[
L[-1 + \theta U_{\gamma}^*] + \phi G_w = 0
\]
where $\phi = \phi(s)$ and $\beta = \beta(s)$ are state-dependent Lagrangian multipliers, $U^e = U(w,e)$, $U^u = U(b,0)$ and $G = G(w,e;m)$. For the equality of (6), we assume that $U_y(0,0) \geq 1/\theta$.

The equalities in (8) - (10) follow from Assumption 1. Necessary condition (6) implies that $b(s)$ is state-independent, i.e., $b^*(s) = \tilde{b}$ for any $s$. Note also that (6) implies that the optimal severance pay ($b$) is positively related to a worker’s bargaining power ($\theta$).

Solving equations (7) - (10) with $b = \tilde{b}$ yields $w^*(s)$, $e^*(s)$ and $L^*(s)$.

The characteristics of the optimal contract $\delta^*$ crucially depend on whether NSC binds or not. Consider the case in which NSC does not bind and $\phi = 0$. In this case, equation (8) becomes the Borch-Arrow condition for optimal risk-sharing when one of the agents is risk-neutral, and equations (6) and (8) imply that the marginal utility of income should be equalized over states. Equations (8) and (9) together imply that the marginal product of labor equals the marginal rate of substitution between income and work effort (or leisure). Therefore, the contract $\delta^*$ is not only optimal in terms of risk-sharing, but is also efficient in production. The intuition is straightforward. Suppose that a combination $(w,e)$, which is efficient in production and risk-sharing, also satisfies NSC. This case is basically equivalent to the case in which the firm can perfectly observe each worker’s effort; and thus, workers have no incentive to shirk. The information sets of the firm and the workers are intrinsically "symmetric." Thus, as Hart (1983) suggests, productive efficiency and optimal risk-sharing can be jointly satisfied rather than mutually exclusive.
When NSC binds and $\phi > 0$, $\delta^*$ exhibits neither productive efficiency nor optimal risk-sharing. Note that $G_w = U_y^e(w,e) - (1-m)U_y^e(w,0) > 0$ (because $U_{ye} \geq 0$), and $G_e = U_c^e < 0$.

Equation (8) implies that the optimal contract is suboptimal in terms of risk sharing because $1 + \theta U_y^e < 0$. Equations (8) and (9) imply that $sf' > -U_y^e/U_c^e$. Thus, when $\phi > 0$ (as shown later, this is the case whenever NSC is binding), the marginal product of labor exceeds the marginal rate of substitution between income and effort, implying that each employed worker’s effort is underemployed.

Figure 1 depicts the case of underemployment. For ease of exposition, we temporarily assume that $L$ is fixed at a certain level. Curve I is the boundary of NSC on which $G(w,e;m) = 0$, so any point below curve I violates NSC. Curve II is the iso-profit curve whose slope equals $sf'$. Curve III is the indifference curve of an employed worker which, since $|U_y^e/U_c^e| < |G_e/G_w|$, is flatter than curve I. The point $a$ denotes an optimal combination $(w,e)$ for a given $L$. It is obvious from Figure 1 that both the firm and the workers can be made better off by trading higher wage for higher effort. For example, workers as a group would be willing to move along the iso-profit curve to point $d$. However, the firm will resist such a trade because once such a "trade" occurs, each worker becomes a shirking "demon" and production drops to zero.

The possibility of underemployment may be understood in conjunction with Cooper (1985). He examines an optimal contract between a firm and a single worker where the worker, whose leisure is normal, has superior information about his reservation wage or tastes. The optimal contract requires an incentive compatibility constraint which induces the worker to reveal his true status, and that under such a constraint, labor is underemployed.
Similar reasoning applies to our framework. When monitoring is not perfect \((m < 1)\), the firm has to adjust \(w\) and \(e\) so that workers find shirking unattractive (i.e., NSC holds). However, it is worth noting that Cooper’s model requires normality of leisure, while ours does not.\(^{12}\)

In the next section, we will consider the conditions under which NSC binds with positive \(\phi\).

II. THE OPTIMAL CONTRACT UNDER COSTLY MONITORING

The previous section has shown that the characteristics of the optimal contract depend on whether NSC binds or not. In this section, we will show that the effectiveness of monitoring \((m)\), in turn, determines whether NSC binds or not. We shall consider two polar cases: NSC never binds (CASE A), and NSC binds for all states (CASE B). While the latter is the case we are primarily interested in, we begin with CASE A in order to demonstrate how CASE B arises.

CASE A is most likely to occur if monitoring is perfect \((m = 1)\). Here, no worker has an incentive to shirk. If, given perfect monitoring, NSC binds, an unrealistic case arises: any employed worker’s utility equals that of caught shirkers (see equation (3)). For simplicity, we assume that this perverse case does not arise for any state when \(m = 1\). (A sufficient condition will be given later.) We denote the optimal contract under CASE A by \(\delta^*_A \equiv \{\delta^*_A(s) = [\bar{b}, w^*_A(s), e^*_A(s), L^*_A(s)] | s \in [\underline{s}, \bar{s}]\}\)

We first consider the conditions under which unemployment can arise for CASE A.
Let \( w_A(s,L) \) and \( e_A(s,L) \) be the solutions to the problem: \( \max_{w,e} \pi(s) + \theta NV(s) \) given \( s \) and \( L \).

Both \( w_A \) and \( e_A \) must uniquely solve (8) and (9) since the maximand is globally and strictly concave given \( s \) and \( L \). Thus, we can write: \( w_A'(s) = w_A(s,L_A'(s)) \) and \( e_A'(s) = e_A(s,L_A'(s)) \). The following results will be useful in deriving the properties of \( \delta_A^* \).

**Lemma 1.** \( \partial U(w_A,e_A)/\partial s \leq 0 \) and \( \partial U(w_A,e_A)/\partial L \geq 0 \), for any \( s \) and \( L \).

*Proof:* See Appendix.

We now define:

\[
\xi_A(s,L) = \delta(s,L)e_A(s,L) - w_A(s,L) + \theta[U(w_A(s,L),e_A(s,L)) - \bar{U}(\bar{b},0)],
\]

which is equal to the left-hand side (LHS) of (10) plus \( (C + \beta(s)) \). Then, we obtain the following results:

**Lemma 2.** \( \partial \xi_A/\partial s > 0 \) and \( \partial \xi_A/\partial L < 0 \), for any \( s, L > 0 \).

*Proof:* See Appendix.

Consider the following condition:

\[
\xi_A(s,N) - C < 0.
\]

Note that under this condition, \( w_A(s,N), e_A(s,N) \) and \( N \) violate (10). Therefore, full employment cannot be optimal at any state in the neighborhood of \( s_\star \), and (12) is a sufficient condition for \( \delta_A^* \) to generate unemployment for some bad states. Condition (12) is more
likely to hold if \( s \) is small (by Lemma 2), \( N \) is large (by Lemma 2) or \( C \) is large because \( \xi_A \)
is independent of \( C \). Let \( s_A^* \) be a state such that \( \xi_A(s_A^*,N) - C = 0 \). (By Lemma 2, such a state exists if \( \xi_A(s,N) \geq C \).) Then, full employment cannot be optimal for any \( s < s_A^* \). Of course, if \( s_A^* > s \), full employment never arises under \( \delta_A^* \).

Using Lemmas 1 and 2, we can establish the following result:

**Proposition 1.** Given (12): (i) \( L_A^*(s) < N \) for \( s < s_A^* \) and \( L_A^*(s) = N \) for \( s \geq s_A^* \); (ii) for any state \( s < s_A^* \), there exist some fixed values, \( \tilde{w}_A \) and \( \tilde{e}_A \), such that \( w_A^*(s) = \tilde{w}_A \) and \( e_A^*(s) = \tilde{e}_A \); (iii) for any state \( s \geq s_A^* \), \( dw_A^*/ds \geq 0, 13 \ de_A^*/ds > 0 \) and \( dU(w_A^*,e_A^*)/ds \leq 0 \); and (iv) \( \delta_A^*(s) \) is unique for any \( s \).

**Proof:** See Appendix.

While a more rigorous proof is given in Appendix, most of the results are directly implied by the Kuhn-Tucker conditions given in the previous section. When \( L < N \), the optimal choices of \( w, e \) and \( sf' \) depend on the three equations (8) - (10) with \( \phi = \beta = 0 \). Substituting (10) into (9) eliminates \( sf' \) which implies that \( w_A^* \) and \( e_A^* \) are independent of \( s \). Then, \( L_A^* \) is obtained by solving equation (10) with fixed values of \( w_A^* \) and \( e_A^* \). Total differentiation of (10) yields \( dL_A^*/ds = -f'/(sf'')e_A^* > 0 \). Since \( L_A^* \) is increasing in \( s \), there must be a threshold state, \( s_A^* \), such that \( L_A^* = N \). Once all the workers are employed, the only way to increase output is to increase work effort, adjusting wages as the state improves. The difference in comparative statics between cases in which \( L_A^* < N \) and \( L_A^* = N \) results from the fact that \( \beta = 0 \) in (10) for the former case while \( \beta > 0 \) for the latter.
When C is sufficiently small, unemployment will never occur under $\delta^*_A$. Stated formally:

**Corollary 1.** When $C = 0$, $\delta^*_A$ specifies full employment for any state.

*Proof*: See Appendix.

Corollary 1 is simply a generalization of Akerlof and Miyazaki’s (1980) wage bill argument. Since work effort and employment are perfect substitutes in production, the risk-neutral firm is indifferent as to how a given amount of total compensation is distributed between employed and unemployed workers, as long as a given profit level is guaranteed. However, risk-averse workers will prefer certain employment to risky rationing of employment. Therefore, the optimal contract will specify full employment. Total payments to workers and total work effort necessary to produce a given output are equally distributed among workers.

The results stated in Proposition 1 are not surprising in that some previous studies, under different settings, have shown the concurrence of rigid wages and unemployment in bad states. However, one novel result of our model is that the magnitude of the employment cost (C) is an important factor in explaining unemployment.

The properties of $\delta^*_A$ are illustrated in Figure 2. Curve A is the locus of all $(w,e)$ such that the utility of an employed worker equals the utility of a laid-off worker, i.e., $U(w,e) = U(\bar{b},0)$. Curve B is the locus of all $(w,e)$ which satisfy $U_y(w,e) = U_y(\bar{b},0)$ which results from (6) and (8) when $\phi = 0$. Curve A is steeper than B by Assumption 3, and the two curves coincide if the income effect is zero. Curve C is the locus of all $(w,e)$ such that an
employed worker’s utility equals that of a caught shirker. Any point below curve C violates NSC. The optimal values of w and e remain at \( \bar{w}_A \) and \( \bar{e}_A \) up to \( s = s^*_A \), and both w and e start to increase for \( s \geq s^*_A \). The sufficient condition under which NSC does not bind under \( \delta^*_A \) is that \( U(w^*_A(s), e^*_A(s)) > U(0,0) \).

Figure 2 also illustrates two important features of \( \delta^*_A \). First, employed workers can never be better off than their laid-off co-workers, if leisure is normal. This is because curve B is below A and the solution contract \( (\bar{w}_A, \bar{e}_A) \) must lie on curve B. Both workers are equally well off only if the income effect equals zero so that curves A and B coincide.\(^{19}\)

Second, once full employment is reached, further improvement in the firm’s state adversely affects each employed worker’s welfare. Since the indifference curve A is steeper than curve B, the employed worker moves to a lower indifference curve (D, for example) as \((w, e)\) moves to the northeast along curve B. This explains part (iii) of Proposition 1. Intuitively, both the firm and its workers would want more labor supplied when labor is more productive (the state improves). But the optimal contract with a risk-neutral firm and risk-averse workers requires (when NSC does not bind) that income be distributed optimally across states in terms of risk sharing (curve B). When leisure is normal (effort is inferior), the firm can redistribute a certain lump-sum amount of labor income from "good" states to "bad" states in order to induce workers to supply more effort in the good state and less in the bad state. This income-distribution process will lower (raise) workers’ ex post level of utility in "good" ("bad") states, while it improves workers’ ex ante expected utility.\(^{20}\)

Now we want to show that the non-shirking condition should bind for any \( s \) when \( m \) is sufficiently small. Consider the function \( G(w, e; m) = U(w, e) - (1-m)U(w, 0) - mU(0,0) \) at
the point \((w_A(s,N), e_A(s,N))\). Since \(G(w_A(s,N), e_A(s,N); m)\) is strictly increasing in \(m\), and \(G(w_A(s,N), e_A(s,N); m=0) < 0\), there must exist an \(m > 0\), such that:

\[(13) \quad G(w_A(s,N), e_A(s,N); \bar{m}) < 0 .\]

Lemma 1 implies that \(U(w_A(s,N), e_A(s,N)) \geq U(w_\star(s), e_\star(s))\), and the fact that \(w_\star(s) = w_A(s, L_\star(s)) \geq w_A(s, N)\) (see A.1 and A.2 in Appendix) implies that \(U(w_A(s, N), 0) \leq U(w_\star(s), 0)\). Therefore, for any \(s\),

\[(14) \quad G(w_\star(s), e_\star(s); \bar{m}) = U(w_\star(s), e_\star(s)) - (1-\bar{m})U(w_\star(s), 0) - \bar{m}U(0,0) < 0 .\]

That is, under (13) contract \(\delta_\star_A\) is infeasible for any \(s\).

Let \(\delta_\star_B = \{\delta_\star_B(s) = [\bar{b}, w_\star_B(s), e_\star_B(s), L_\star_B(s)] \mid s \in [s_L, s_U]\}\) denote the optimal contract under (13). Given \(\delta_\star_B\), CASE B arises: NSC always binds under the optimal contract. To see why, suppose that \(\phi = 0\) (NSC does not bind) at some state \(s\). Then, the solution to (7) - (10) must be equal to \(\delta_\star_A(s)\), by (iv) of Proposition 1. However, this is a contradiction because \(\delta_\star_A(s)\) is not feasible under (13). Since this result plays an important role in determining the properties of \(\delta_\star_B\), we state it formally:

**Proposition 2.** For \(\delta_\star_B\), \(\phi > 0\) for any \(s\) (NSC always binds).

Hence, for the derivation of \(\delta_\star_B\), we can replace (7) by:

\[(7') \quad \bar{G} = G(w, e; \bar{m}) = 0 .\]

Let \(w_B(s, L)\) and \(e_B(s, L)\) be the solutions to (7'), (8) and (9). A sufficient condition for \(w_B(s, L)\) and \(e_B(s, L)\) to be optimal given \(s\) and \(L\) is that the set, \(\{(w,e) \mid G(w,e; \bar{m}) \geq 0\}\), is
convex \cite{Arrow and Enthoven 1961}, or equivalently, \( G(w,e;\bar{m}) \) is quasiconcave on the boundary of NSC. We assume this sufficient condition is satisfied, and thus:

\[
2G_wG_eG_{we} - G_e^2G_{ee} - G_w^2G_{ww} \geq 0,
\]

(15)

where \((1- \bar{m}) = \frac{[U(w,e)-\bar{m}U(0,0)]}{U(w,0)}\) and \(U_s = U(w,0)\). A separable utility function trivially satisfies (15) for any \(\bar{m}\) because \(U_{ye} = 0\) and \(U_{yy} = U_{yy}\). The quasiconcavity condition (15), while not necessary, takes on an important role in determining the properties of \(\delta^*_B\).

We now obtain the following results:

**Lemma 3.** Given (13) and (15), \(\partial U(w_B,e_B)/\partial s > 0\); \(\partial U(w_B,e_B)/\partial L < 0\), for any \(s, L > 0\).

*Proof:* See Appendix.

Note that Lemma 3 is exactly the opposite of Lemma 1. Defining

\[
\xi_B(s,L) = s\delta(e_B(s,L)L)e_B(s,L) - w_B(s,L) - b + \theta[U(w_B(s,L),e_B(s,L))] - U_B(0,0)
\]

(16)

we also obtain the following results, which are identical to those of Lemma 2:

**Lemma 4.** Given (13) and (15), \(\partial \xi_B/\partial s > 0\); \(\partial \xi_B/\partial L < 0\), for any \(s, L > 0\).

*Proof:* See Appendix.

Like CASE A, contract \(\delta^*_B\) generates unemployment under the condition:

\[
\xi_B(s,L) - C < 0,
\]

(17)
In this case, full employment can never be optimal for states in the neighborhood of \( s \). For full employment in some good states, \( \xi_B \) must increase with \( s \) after some point. Condition (15) ensures that this is the case, because \( \frac{\partial \xi_B}{\partial s} > 0 \). Now, let \( s_B^* \) be the state for which \( \xi_B(s_B^*, N) - C = 0 \). Then, using Lemmas 3 and 4, we obtain the following:

**Proposition 3.** Suppose that (13), (15) and (17) hold. Then, (i) \( L_B^*(s) < N \) for \( s < s_B^* \) and \( L_B^*(s) = N \) for \( s \geq s_B^* \); (ii) for any state \( s < s_B^* \), there exist some fixed values, \( \bar{w}_B \) and \( \bar{e}_B \), such that \( w_B^*(s) = \bar{w}_B \) and \( e_B^* = \bar{e}_B \); and (iii) for any state \( s \geq s_B^* \), \( \frac{dw_B^*}{ds} > 0 \), \( \frac{de_B^*}{ds} > 0 \) and \( \frac{dU(w_B^*, e_B^*)}{ds} > 0 \).

**Proof:** See Appendix.

We omit the formal proof because it parallels that of Proposition 1. The results stated in Proposition 3 are obtained from (7’) and (8) - (10) with \( \phi > 0 \). When \( L < N \) and \( \beta = 0 \). The solutions of \( w, e, s^* \) and \( (\phi/L) \) do not depend on \( s \), so that \( w_B^*, e_B^* \) are fixed. \( L_B^* \) is obtained by solving equation (10), and \( L_B^* \) increases with \( s \): \( \frac{dL_B^*}{ds} = -\frac{f'}{sf''}e_B^* > 0 \). Once \( L_B^* \) reaches the full employment level, work effort should increase as the state improves.

Note that both \( \delta_A^* \) and \( \delta_B^* \) specify fixed wage and effort for low \( s \). At first glance, fixed effort seems implausible. Since labor productivity is positively related to \( s \), higher effort in good states and lower effort in bad states would seem to benefit both workers and the firm.

However, this is not so. The higher level of employment in the good states causes a countervailing reduction in the marginal product of effort so that marginal product and effort remain constant.23
An interesting difference between $\delta^*_A$ and $\delta^*_B$ is that for the latter, the firm’s market status positively affects each employed worker’s welfare. Figure 3 illustrates why. Curves A and B are the same as in Figure 2. Curve E is the boundary of the non-shirking condition, $G(w,e;\bar{m}) = 0$. Since NSC binds, the optimal contract $\delta^*_B$ lies on curve E, rather than on the optimal risk-sharing condition, curve B. Further, since E is steeper than any indifference curve (F, for example) of an employed worker, the utility of an employed worker increases as $(w,e)$ moves to the northeast along curve E. This explains (iii) of Proposition 3. The difference between $\delta^*_A$ and $\delta^*_B$ results from the fact that the firm can no longer provide optimal insurance under $\delta^*_B$. The redistribution of labor income from good to bad states, which is mandated when NSC does not bind, will induce workers to shirk in good states. Thus, to increase effort in good states, the firm must offer higher wages and ex post utility.

Since by (6) and (8), $U_y(w^*_B,e^*_B) < U_y(\bar{b},0)$, contract $\delta^*_B$ must lie on that segment of curve E which lies above B. Note also that involuntary unemployment arises if $(\bar{w}_B,\bar{e}_B)$ lies above curve A. In the following section, we consider the conditions under which involuntary unemployment can occur.

III. INVOLUNTARY UNEMPLOYMENT

The previous sections have demonstrated how imperfect and costly monitoring distorts the optimal contract. This section deals with the issue of involuntary unemployment, "involuntary" in the sense that an employed worker’s utility is higher than that of an identical, but unemployed, twin. Shapiro and Stiglitz (1984) showed how imperfect monitoring can
result in involuntary unemployment. In their model, laid-off and fired workers share the same level of utility, so that firms should offer employed workers higher utility than that enjoyed by unemployed workers in order to prevent the employed from shirking. However, in our model, laid-off and fired workers are not treated as equals, so their result does not necessarily carry over to our model. In our model, caught shirkers are discharged without compensation. Laid-off workers, on the other hand, are compensated for the separation from their employers.

Based on the results of previous sections, we shall examine sufficient conditions under which an optimal contract generates involuntary unemployment. Since involuntary unemployment never arises under \( \delta^*_A \), our discussion is limited to \( \delta^*_B \). We investigate conditions for involuntary unemployment at the worst state, \( s_0 \), because under \( \delta^*_B \) employed workers’ utility is nondecreasing with \( s \).

We begin by denoting the weighted welfare function of the firm and its workers at state \( s \), \( \pi(s) + \theta NV(s) \), by:

\[
\begin{align*}
\pi(s) + \theta NV(s) = (18) Y(w,e,L) = \frac{s}{L(e)} - wL - b(N - L) - c_L + \theta [LU(w,e) + (N - L)U(b,0)]
\end{align*}
\]

Let \( L(w,e) \) be the employment level which maximizes the welfare function, \( Y \), at given \( (w,e) \); that is, \( L(w,e) \) is the solution of equation (10) with \( \beta = 0 \). Here, \( L(w,e) \) is allowed to be greater than \( N \). Substituting \( L(w,e) \) into (18), we define a concentrated welfare function by:
Some straightforward algebra shows that $Y_c(w,e)$ is quasiconcave. Note that when the condition for unemployment, (17), holds, $L^*(s) = L(\bar{w}_B, \bar{e}_B) < N$. Further, since $(\bar{w}_B, \bar{e}_B)$ satisfies the first order conditions for maximizing $Y_c$ subject to $G(w,e;\bar{m}) = 0$, maximizing $Y_c$ subject to NSC is equivalent to maximizing $Y$ subject to NSC and the employment restriction, $L \leq N$. That is, $(\bar{w}_B, \bar{e}_B)$ is determined where a contour of $Y_c$ is tangent to the boundary of NSC. This result can be used to derive conditions under which $\delta^*_B$ generates involuntary unemployment.

Figure 4 illustrates the case where $U(\bar{w}_B, \bar{e}_B) > U(\bar{b},0)$. Curves A, B and E are the same as in Figure 3. Curve H is a locus of $(w,e)$ such that $Y_c(w,e) = Y_c(\bar{w}_B, \bar{e}_B)$. Consider the point $(\tilde{w}, \tilde{e})$ at which curves A and E intersect, that is, where $G(\tilde{w}, \tilde{e};\bar{m}) = 0$ and $U(\tilde{w}, \tilde{e}) = U(\bar{b},0)$. Obviously, involuntary unemployment arises if $\bar{w}_B > \tilde{w}$ or $\bar{e}_B > \tilde{e}$. This condition is satisfied if and only if curve H', the contour of $Y_c$ where $Y_c(w,e) = Y_c(\tilde{w}, \tilde{e})$, is steeper than curve E, the boundary of NSC. The slopes of curves E and H' at $(\tilde{w}, \tilde{e})$ are given by:

\[
\gamma_0(\tilde{w}, \tilde{e}) = -\frac{G_w}{G_{w}} \frac{U_c(\bar{w}, \bar{e})}{U_j(\tilde{w}, \tilde{e}) - (1-\theta)U_j(\bar{w}, 0)},
\]

\[
\gamma_Y(\tilde{w}, \tilde{e}) = -\frac{Y_{ee}}{Y_{ew}} \frac{\partial U_y^0(\tilde{w}, \tilde{e})}{\partial \tilde{w}} = \frac{s f^0\{G(\tilde{w}, \tilde{e})\} + \theta U_y(\tilde{w}, \tilde{e})}{1 + \theta U_y(\tilde{w}, \tilde{e})}.
\]
where the last equality in (21) results from (10) with $\beta = 0$ and the fact that $U(\tilde{w}, \tilde{e}) = U(\tilde{b},0)$.

Thus, involuntary unemployment arises if and only if:

$$\gamma_d(\tilde{w}, \tilde{e}) < \gamma_f(\tilde{w}, \tilde{e}) .$$

Stated formally:

**Proposition 4.** If conditions (13), (15), (17) and (22) hold, involuntary unemployment arises under $\delta^*_B$; that is, $U(w^*_B(s), e^*_B(s)) \geq U(\tilde{w}_B, \tilde{e}_B) > U(b,0)$.

The magnitude of $C$ is an important determinant for involuntary unemployment. Observe that $(\tilde{w}, \tilde{e})$ does not depend on $C$. Equations (20) and (21) show that $\gamma_f(\tilde{w}, \tilde{e})$ increases with $C$, while $\gamma_d(\tilde{w}, \tilde{e})$ remains unchanged. This result implies that when $C$ is sufficiently large, both conditions (17) and (22) should hold, and thus, $\delta^*_B$ generates involuntary unemployment. Intuitively, when monitoring costs are large, the firm wants to substitute employment for effort in order to reduce employment costs. Asking workers to provide greater effort means that the firm must counter the increased incentive to shirk by offering employed workers a higher wage. This strategy will result in a higher level of utility for each employed worker because the boundary of NSC (curve $E$) is steeper than an employed worker’s indifference curve. (See Figure 4.)

An increase in $C$ not only results in a higher likelihood of involuntary unemployment, but also in a greater *ex post* welfare loss for unemployed workers. This is so because the *ex post* utility of an employed worker is positively related with $C$. Stated formally:
Proposition 5. $\frac{\partial U(\bar{w}, \bar{e})}{\partial \bar{e}} > 0$, for all $s < \delta^*_{\bar{e}}$.

Proof: See Appendix.

The likelihood of involuntary unemployment under $\delta^*_{\bar{e}}$ also depends on the form of utility function. An obvious example is the case when the utility function has no income effect, e.g., $U(y,e) = u(y-v(e))$. In this case, $\gamma_{y}(\bar{w}, \bar{e}) = \infty$, because $U_{y}(\bar{w}, \bar{e}) = 1/\theta = U_{y}(\bar{b}, 0)$. Therefore, condition (22) should hold. Figure 4 explains this result more explicitly. When there is no income effect, curve B coincides with curve A, so that $U(w^*, e^*_B) \geq U(\bar{w}, \bar{e}) > U(\bar{b}, 0)$, and employed workers are better off than their laid-off colleagues. This result indicates that in general, involuntary unemployment would be more likely if the income effect is small ($[U_{y}U_{yy}/U_{y} - U_{ye}]$ is near zero). Note that the less risk-averse a worker (smaller $-U_{yy}/U_{y}$), the lower the disutility of work ($U_{e}$), and the smaller the income effect.

In order to obtain further insight into the possibility of involuntary unemployment, we consider a specific example. Assume that a worker’s utility function takes the following separable form:

$$U(y,e) = y^{\alpha} - ke,$$

where $0 < \alpha < 1$ and $k > 0$. With this specification, we are assuming that each worker is risk-neutral in terms of effort while risk-averse in terms of income. The nonshirking condition (NSC) is:

$$G(w, e; m) = \bar{m}w^{\alpha} - ke \geq 0.$$
While the utility function in (23) is not strictly concave, all the results in Proposition 2 and 3 still apply. In fact, Proposition 2 holds for any $\bar{m} < 1$. To see why, suppose that $\phi = 0$. Substituting (8) and (9) into (10) yields:

$$\text{LHS of (10)} = -(C + \beta) < 0,$$

which is a contradiction if $C > 0$. The reason why NSC binds for any $\bar{m}$ is that the utility function in (23) is only weakly concave.\(^{28}\)

Condition (17) is satisfied if we choose a sufficiently large value for $C$. Substituting (23) and (24) into (22), we obtain the following condition:

$$J_B = (1 - m) \left[ 1 + \left( \frac{\alpha}{1 - \alpha} \right) C(\beta \alpha)^{-U(1 - \alpha)} \right] > 1,$$

where $J_B$ equals the ratio of utility if employed to utility if laid off, i.e., $U(\bar{w}_B, \bar{e}_B)/U(\bar{b},0)$.\(^{29}\)

As discussed below Proposition 4, involuntary unemployment occurs if the monitoring costs of each employed worker ($C$) are sufficiently large. The degree of risk-aversion is also an important factor generating involuntary unemployment. Workers become less risk-averse as $\alpha$ increases, because $-U_{yy}/U_y = [(1 - \alpha)/\alpha]/y$. Note that $J_B \to \infty$ as $\alpha \to 1$, if $\theta < 1$. Thus, the likelihood of involuntary unemployment increases as workers become less risk-averse, if their bargaining power ($\theta$) is low.

Condition (26) also shows that the lower the probability of catching a shirker ($\bar{m}$), or the lower workers’ bargaining power ($\theta$), the more likely involuntary unemployment will occur. Some intuition follows. First, when NSC [condition (24)] binds the firm must increase $w$ or decrease $e$ as $\bar{m}$ decreases, in order to suppress workers’ incentives to shirk. This may result in higher utility for an employed worker. Second, as we see from (6), the
weaker is workers' bargaining power, the smaller is severance pay. NSC implies that an employed worker's utility must be higher than that of a caught shirker. Therefore, when severance pay is so low that the difference between the utility of an unemployed and a fired worker becomes negligible, involuntary unemployment is likely to occur. This is essentially the case of Shapiro and Stiglitz (1984). However, it is not clear how the effect of $\bar{m}$ and $\theta$ on the likelihood of involuntary unemployment, which we find from (26), may be generalized to other utility functions.\(^{30}\)

IV. CONCLUSIONS AND COMMENTS

Implicit contract models were motivated to explain rigid wages and involuntary unemployment. But early models have been criticized because of the assumptions that the expected utility of an unemployed worker is fixed at a low level and individual work effort is fixed (no work sharing). Allowing for endogenous effort and severance pay often resulted in "voluntary" rather than involuntary unemployment. Asymmetric information models could explain involuntary unemployment but resulted in state-dependent wages.

Our paper attempts to avoid the above problems while explaining the forces which generate rigid wages and involuntary unemployment simultaneously. The results obtained in this paper are also consistent with some other stylized macroeconomic facts: (i) state-dependent employment and output, and (ii) procyclical labor productivity. The important idea in this paper is that when firms have to engage in costly monitoring of their workers, and the monitoring is not perfect, both the cost and effectiveness of monitoring are factors explaining.
rigid wages and involuntary unemployment. Nonetheless, we do not wish that these results be overemphasized, as alternative assumptions about the contract may result in different implications. For example, rigid wages and involuntary unemployment may not arise concurrently if firms are risk-averse or have superior information on the realized state of nature.

In some cases, monitoring may not be necessary. In small firms, for example, an individual worker’s contribution to output may be directly observable. In this case, the firm, without monitoring, may successfully induce nonshirking by adopting some incentive compatible wage scheme where a worker’s wage is tied to the worker’s output. Or, if individual effort cannot be observed, profit sharing, whereby each worker’s income is tied to profits, may be utilized to control shirking. For these small firms, the characteristics of the contracts may be quite different from ours. However, incentive compatible wage schemes and profit sharing are unlikely to be successful in firms which hire a large number of low-skilled workers (for example, manufacturing). For these firms, each worker’s contribution to production is not easily separable, so it is not possible to tie each worker’s wage to his output. Profit sharing also has a serious limitation as the penalty for shirking (reduced profit) is shared by all workers. Since the gains to shirking accrue entirely to the individual worker and the costs are dispersed across all workers, the incentive effects of profit sharing will be negligible. Thus, large firms are likely to adopt monitoring devises to assure worker’s performance thereby giving additional credence to our model and its result.
APPENDIX

Proof of Lemma 1. The comparative statics on (8) and (9), with $\phi = 0$, yield:

\[
\frac{\partial w_A}{\partial s} = \frac{f'L^2\theta U^e_y}{K_A} \geq 0; \quad \frac{\partial e_A}{\partial s} = \frac{-f'L^2\theta U^e_y}{K_A} > 0;
\]

\[
\frac{\partial w_A}{\partial L} = \frac{sf''L^2\theta U^e_y}{K_A} \leq 0; \quad \frac{\partial e_A}{\partial L} = \frac{-sf''L^2\theta U^e_y}{K_A} < 0;
\]

where $K_A = sf''L^2\theta U^e_y + L^2\theta[U^e_y U^e_{ee} - (U^e_y e)^2] > 0$. By (A.1) and (A.2), we have:

\[
\frac{\partial U(w_A, e_A)}{\partial s} = U_y^e \frac{\partial w_A}{\partial s} + U_y^e \frac{\partial e_A}{\partial s} = \frac{sf'L^2\theta U^e_y}{K_A} \geq 0.
\]

\[
\frac{\partial U(w_A, e_A)}{\partial L} = U_y^e \frac{\partial w_A}{\partial s} + U_y^e \frac{\partial e_A}{\partial s} = \frac{sf''L^2\theta U^e_y}{K_A} \leq 0.
\]

Proof of Lemma 2. Using (A.1), (A.2), (8) and (9), we have:

\[
\frac{\partial \xi_A}{\partial s} = \frac{f'L^2\theta U^e_y U^e_{ee} - (U^e_y e)^2}{K_A} > 0;
\]

\[
\frac{\partial \xi_A}{\partial L} = \frac{-sf''L^2\theta U^e_y U^e_{ee} - (U^e_y e)^2}{K_A} < 0.
\]

Proof of Proposition 1: (i) Suppose that $L_A'(s_1) = N$ for a state $s_1 < s_A^*$. Then, Lemma 2 and (10) imply that $\xi_A(s_1, N) - C < 0$ and $\beta(s_1) \leq 0$. However, note that LHS of (10) = $\xi_A(s_1, N) - C - \beta(s_1) < 0$. Therefore, (10) is not satisfied. This is a contradiction. Now,
suppose that \( L_A^*(s_2) < N \) for a state \( s_2 \geq s_A^* \). Then, \( \xi_A(s_2,N) - C \geq 0 \) and \( \beta(s_2) = 0 \).

Furthermore, \( \xi_A(s_2, L_A(s_2)) > \xi_A(s_2,N) \), by Lemma 2. Then, LHS of (10) = \( \xi_A(s_2, L_A(s_2)) - C - \beta(s_2) > 0 \), which is also a contradiction.

(ii) By (10), \( \xi_A(s, L_A(s)) - C = 0 \). Total differentiation gives us:

\[
(A.7) \quad \frac{dL_A^*}{ds} = -\frac{\partial \xi_A / \partial s}{\partial \xi_A / \partial L} = -\frac{f'}{sf''} > 0 .
\]

Then, using (A.1), (A.2) and (A.7), we obtain:

\[
(A.8) \quad \frac{dw_A^*}{ds} = \frac{dw_A(s, L_A^*(s))}{ds} = \frac{\partial w_A}{\partial s} + \frac{\partial w_A}{\partial L} \frac{dL_A^*}{ds} = 0 .
\]

Similarly, \( de_A^*/ds = 0 \).

(iii) Since \( L_A^* = N \), the results directly follow from (A.1) and Lemma 1.

(iv) Since \([\pi(s) + \theta NV(s)]\) is globally and strictly concave given \( s \) and \( L \), \( w_A \) and \( e_A \) are unique, and so are \( \dot{w}_A(s) \) and \( \dot{e}_A(s) \) given \( s \) and \( L_A^*(s) \). Hence, it suffices to show that \( L_A^*(s) \) is unique for a given \( s \). For \( s \geq s_A^* \), \( L_A^*(s) = N \) is the unique solution. Now consider a state \( s < s_A^* \). By (10), \( L_A^*(s) \) must solve: \( \xi_A(s, L_A(s)) - C = 0 \). Since \( \xi_A \) is strictly increasing with \( L \), \( L_A^*(s) \) must be unique.

Proof of Corollary 1:  We first show that \( \xi_A \) is always positive. Substituting (8) and (9), with \( \phi = 0 \), into \( \xi_A \) gives us:

\[
(A.9) \quad \xi_A(s, L) = \frac{1}{u'} \left[ U^* - U^u - U^w(s - \bar{b}) - U^e \right] \geq 0 ,
\]

because \( U \) is strictly concave. Note that the equality holds only if \( w^*(s_1) = \bar{b} \) and
e^*(s_i) = 0 (no production). However, since e_A(s,L) > 0, \xi_A must be positive. Now, suppose that L_A^* < N. Then, LHS of (10) = \xi_A(s,L_A^*) > 0, which is a contradiction.

Comment: When the utility function is weakly concave, multiple solutions are possible, so that the equality in (A.9) may hold for some non-zero e. For this case, we can still show that full employment is optimal, in a manner similar to Proposition 1 of Cooper (1987).

Proof of Lemma 3: Define:

\( M_B = \theta L(2\bar{G_w}G_sU_{yw}^e - \bar{G_s}^2U_{yw}^e - \bar{G_w}^2U_{we}^e) + \phi(2\bar{G_w}G_sG_{we} - \bar{G_s}^2G_{ww} - \bar{G_w}^2G_{we}) \).

The concavity of U and (15) imply that M_B > 0. The comparative statics on (7'), (8) and (9) yield:

\[ \frac{\partial \bar{M}_B}{\partial s} = \frac{-f' \bar{G_w} \bar{G_s}}{K_B} > 0; \quad \frac{\partial e_B}{\partial s} = \frac{f' \bar{G_w}^2}{K_B} > 0; \]

\[ \frac{\partial \bar{M}_B}{\partial L} = \frac{-sf'' \bar{G_w} \bar{G_s}}{K_B} < 0; \quad \frac{\partial e_B}{\partial L} = \frac{sf'' \bar{G_w}^2}{K_B} < 0; \]

where \( \bar{G}_w = U_y^e - (1-\bar{m})U_y^e > 0, \bar{G}_c = U_c^e < 0 \) and \( K_B = -sf'' \bar{G}_w^2 + M_B > 0 \). By (A.11) and (A.12), we have:

\[ \frac{\partial U(w_B,e_B)}{\partial s} = U_y^e \frac{\partial \bar{M}_B}{\partial s} + U_c^e \frac{\partial e_B}{\partial s} = \frac{-(1-\bar{m})U_y^e U_c^e f' L \bar{G_w}}{K_B} > 0, \]
Proof of Lemma 4. Note that:

\( \frac{\partial U(w_B, e_B)}{\partial L} = U_y^* \frac{\partial w_B}{\partial L} + U_z^* \frac{\partial e_B}{\partial L} - \frac{(1 - \bar{m}) U_z^* U_y^* \bar{f} \bar{L} \bar{e} \bar{G}_w}{K_B} < 0. \)

Proof of Proposition 5: Total differentiation of (7') and (8) - (10), with \( \beta = 0 \), gives us:

\( \frac{\partial \xi_B}{\partial \bar{s}} = \frac{f' e M_B}{K_B} > 0 \); \( \frac{\partial \xi_B}{\partial L} = -\frac{sf'' e^2 M_B}{K_B} < 0. \)

\( \frac{\partial \xi_B}{\partial L} = \frac{f' e M_B}{K_B} > 0 \); \( \frac{\partial \xi_B}{\partial L} = -\frac{sf'' e^2 M_B}{K_B} < 0. \)

Proof of Proposition 5: Total differentiation of (7') and (8) - (10), with \( \beta = 0 \), gives us:

\[
H \begin{pmatrix} d\phi \\ d\xi_B \\ d\xi_B \\ dL_B \\ dL_B \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} dC,
\]

where,

\[
H = \begin{pmatrix} 0 & \bar{G}_w & \bar{G}_e & 0 \\ \bar{G}_w & \theta L U_{y_f}^* + \phi \bar{G}_w & \theta L U_{y_e}^* + \phi \bar{G}_w & -1 + \theta U_y^* \\ \bar{G}_e & \theta L U_{y_e}^* + \phi \bar{G}_w & \theta L U_{e_e}^* + \phi \bar{G}_e & \theta L U_{e_e}^* + \phi \bar{G}_e + \bar{f} \bar{e} L \bar{e} \bar{G}_w \\ 0 & -1 + \theta U_w^* & \bar{f} \bar{e} L + \bar{f} \bar{e} + \theta U_e^* & \bar{f} \bar{e}^2 \end{pmatrix}.
\]

By virtue of the second-order condition, \( |H| < 0 \). Applying Cramer’s rule and using (8) and (9), we can show:
(A.19) \[ \frac{\partial \bar{w}_B}{\partial C} = -\frac{s \bar{f}'' \bar{L} \bar{e} \bar{G}_w \bar{G}_e}{|H|} > 0 \; ; \; \frac{\partial \bar{e}_B}{\partial C} = \frac{s \bar{f}'' \bar{L} \bar{e}^2 \bar{G}_w}{|H|} > 0 \, . \]

Therefore, we have:

(A.20) \[ \frac{\partial U(\bar{w}_B, \bar{e}_B)}{\partial C} = U_y^e \frac{\partial \bar{w}_B}{\partial C} + U_e \frac{\partial \bar{e}_B}{\partial C} = \frac{-(1 - \bar{m}) U_y^e U_y^2 s \bar{f}'' \bar{L} \bar{G}_w}{|H|} > 0 \, . \]
NOTES

1. In Foster and Wan (1984), wage is state-dependent, increasing with the production shock. Also, the expected utility of an unemployed worker is fixed at an arbitrarily low level as in Azariadis’ (1974) model. In Moore (1984), wage increases with the worker’s degree of disutility of working, while severance pay is state-independent.

2. While the assumption that m is fixed during the contract period simplifies the analysis, the assumption is not without justification. First, changing monitoring intensity during the contract period may be construed as an unfair labor practice. Thus, "evidence" of shirking obtained by an increased monitoring effort may not be legally admissible if the worker challenges his dismissal in court. Second, state-dependent monitoring may violate implicit workplace relationships and create an atmosphere of distrust and discontent among workers, reducing productivity. Third, it may simply be too costly, once the monitoring technology is in place, to alter it during the contract period.

3. If the firm is competitive in the product market, s and f can be regarded as the price of the firm’s product and the production function, respectively.

4. We assume strict concavity for expository purposes. It can be replaced by weak concavity, $U_{ys}U_{ee} - (U_{ye})^2 \geq 0$.

5. If we interpret e as effective working hours, then $U_{ye} < 0$ implies that wage rates are countercyclical under the optimal contract we consider later. Many empirical studies find that wage rates, to the contrary, are positively correlated with labor demand. See, for example, Bils (1985).

6. See, for example, Foster and Wan (1984).
7. Suppose that a self-employed worker solves the problem: \( \max e \ U(\bar{y}+re,e) \), where \( r \) is the constant marginal product of effort, and \( \bar{y} \) is non-labor income. Let \( e^* \) be the solution to this problem. Then, it can be shown that \( \partial e^*/\partial \bar{y} \leq 0 \) only if Assumption 3 holds.

8. All that we need for our results is that a state-independent penalty can be imposed on shirkers such that the utility of a nonshirker exceeds the expected utility of a shirker for some \( w \) and \( e \).

9. If the government pays unemployment benefits \( (b_g) \) to unemployed workers, fired workers may also be eligible for such benefits. In this case, the expected utility of a potential shirker is given by: \( (1-m)U(w,0) + mU(b_g,0) \). Our results are robust to this alternative specification, unless \( b_g \) is state-dependent.

10. One may alternatively define the optimal contract as the solution to: \( \max E_s [\pi(s)] \) subject to (5), (3) and \( E_s[V(s)] = \bar{U} \), where \( \bar{U} \) is the reservation utility level. In this case, \( \theta \) can be interpreted as the state-invariant Lagrangean multiplier corresponding to the reservation utility constraint. By duality, if we fix \( \theta \) in (4) at its solution value in the above problem and solve (4) subject to (5) and (3), \( E_s[V((w^*(s),e^*(s),L^*(s))] = \bar{U} \). That is, the two methods yield the same solution. We choose to define the optimal contract as the solution to the system described by (3) - (5), because it allows us to obtain straightforward comparative statics.

11. If we allow \( w \), \( e \) or \( L \) to be zero, the equalities in (8) - (10) must be replaced by weak inequalities. However, none of the optimal \( w \), \( e \) and \( L \) can be zero. To see why, suppose that \( L = 0 \). Then, \( \pi(s) + \theta NV(s) = -\delta N + \theta NU(b,0) \), for any \( w \) and \( e \). Choose \( w \) and \( e \) such that NSC holds and \( U(w,e) = U(b,0) \). Since \( f'(0) = \infty \), if the
firm increases \( L \) infinitesimally, \( \pi(s) \) increases with \( V(s) \) unchanged. This is a contradiction. Similarly, we can show that \( w \) and \( e \) must be positive, otherwise (7) - (9) imply that \( L = 0 \).

12. Other studies have also noted the possibility of underemployment in those cases in which a firm has superior information about its revenue function. Grossman and Hart (1981), Azariadis (1983) and Moore (1985) find an underemployment outcome when a worker’s utility function exhibits a zero income effect and the firm is risk-averse. Extending this result, Chari (1983) and Green and Kahn (1983) show that underemployment never occurs if leisure is normal and the firm is risk-neutral. These results differ from those of Cooper (1985) and this paper.

13. Equality holds only if \( U_{ye} = 0 \).

14. Equality holds only if income effect is zero.

15. They obtain the same result for the case of fixed work effort.

16. For example, Rosen (1985, Section V) considers the case in which working hours and employment are not perfect substitutes in production.

17. This curve is upward sloping because we assume \( U_{ye} \geq 0 \).

18. If leisure is inferior, that is, if \( (U_e/U_y)U_{yy} < U_{ye} \), \( B \) will be steeper than \( A \).

19. Employed workers can be better off than laid-off workers if leisure is inferior.

20. See Rosen (1985) for further details. The argument in the text is reversed if leisure is inferior.

21. The constraint qualification condition (Arrow and Enthoven [1961], Theorem 2) holds. Note that (i) for some \((w,e)\), \( G(w,e;\bar{m}) > 0 \), and (ii) \( G_w(w,e;\bar{m}) \neq 0 \) and \( G_e(w,e;\bar{m}) \neq 0 \) for \((w,e)\) satisfying NSC.
22. In general, the first term on the right-hand side of (15) is negative while the third term is positive. The sign of the second term is indeterminate. A sufficient condition for the second term to be nonnegative is that the degree of absolute risk-aversion with respect to income \((-U_{yy}/U_y)\) is non-decreasing in \(e\). In this case, \(U_{yy} \leq 0\) (because \(U_{ye} \geq 0\)), so that \([U_{yy}^e - (1-m)U_{yy}] > 0\). For example, \(U = [y - v(e)]^\gamma\) (\(\gamma < 1\) and \(v(0) = 0\)) has this property and satisfies (15) for any \(\bar{m}\). (Note that with this utility function, NSC is expressed as \([1-(1-\bar{m})^{1/\gamma}]w - v(e) \geq 0\), and the boundary of NSC is quasiconcave.) In short, the boundary of NSC, \(G(w,e;\bar{m}) = 0\), is likely to be quasiconcave, if (i) \(U_{ye}\) is small, (ii) \(U_{ee}\) is large, or (iii) \([U_{yy}^e - (1-m)U_{yy}]\) is positive.

23. We are indebted to an anonymous referee for this point.

24. We are indebted to an anonymous referee for pointing out this crucial difference between Shapiro and Stiglitz’ model and ours.

25. Since the utility function is concave, we have:

\[
U(w,e) - U(b,0) - U_y(b,0)(w-b) - U_e(b,0)e < 0,
\]

for any \((w,e)\). This condition in turn implies that:

\[
-w + \tilde{b} - C + \theta[U(w,e) - U(\bar{b},0)] < 0,
\]

because \(C > 0\), \(U_e < 0\) and \(U_y(\tilde{b},0) = 1/\theta\). Therefore, \(L(w,e)\) exists for any \((w,e)\) as long as \(f'(\infty) = 0\) and \(f'(0) = \infty\).

26. The bordered Hessian matrix of \(Y_{\xi}(w,e)\) is given by:
27. Note that \(-1 + \theta U_y(\hat{w}, \hat{e}) < 0\). This is so because \((\hat{w}, \hat{e})\) is located above curve B on which \(U_y(w, e) = 1/\theta = U_y(\bar{b}, 0)\).

28. In fact, for any utility function with the property that \(U_{yy} U_{ee} - (U_{ye})^2 = 0\), we can show that NSC binds for any \(\bar{m} < 1\).

29. Consider the first-order conditions for \(\delta_B^*\) when \(L_B^* < N\). Replacing \(U(w, e)\) in (6) and (8) - (10) by (23) and substituting (6), (24), (8) and (9) into (10), we can show that \(\bar{w}_B = \bar{b} + \{\alpha/(1-\alpha)\}C\), where \(\bar{b} = (\theta \alpha)^{1/(1-\alpha)}\). Then, condition (24) with inequality implies that \(U(\bar{w}_B, \bar{e}_B) = (1-\bar{m}) \bar{w}_B^{\alpha} = (1-\bar{m})(\theta \alpha)^{1/(1-\alpha)} + \{\alpha/(1-\alpha)\} C^\alpha\).

30. For a general form of utility function, condition (22) alone does not lead to clear-cut effects of \(\bar{m}\) and \(\theta\) on the likelihood of involuntary unemployment because \((\hat{w}, \hat{e})\) depends on \(\bar{m}\) and \(\theta\).