working paper
department
of economics
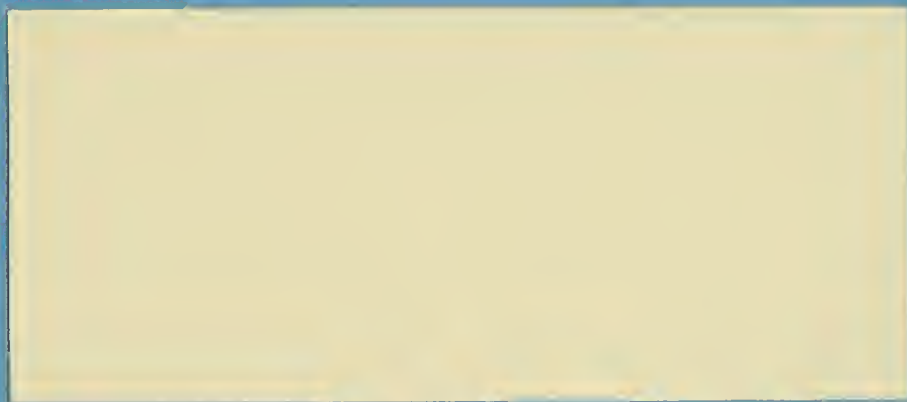
THE LABOR MARKET AND CORPORATE STRUCTURE

Daron Acemoglu
Andrew F. Newman

No. 97-8   June, 1997

massachusetts
institute of
technology

50 memorial drive
cambridge, mass. 02139

WORKING PAPER
DEPARTMENT
OF ECONOMICS

THE LABOR MARKET AND CORPORATE STRUCTURE

Daron Acemoglu
Andrew F. Newman

MASSACHUSETTS
INSTITUTE OF
TECHNOLOGY

50 MEMORIAL DRIVE
CAMBRIDGE, MASS. 021329

The Labor Market and Corporate Structure[*]

Daron Acemoglu[†]   Andrew F. Newman[‡]

June 1997

Abstract

We analyze the impact of labor demand and labor market regulations on the corporate structure of firms. We find that higher wages are associated with lower monitoring, irrespective of whether these high wages are caused by labor market regulations, unions or higher labor demand. These comparative static results are in line with the broad trends in the data. We also find that the organization of firms has important macroeconomic implications. In particular, monitoring is a type of "rent-seeking" activity and the decentralized equilibrium spends excessive resources on monitoring. Labor market regulations that reduce monitoring by pushing wages up may increase net output or reduce it only by a small amount even though they reduce employment.

Keywords: Corporate Structure, Efficiency Wages, Labor Market Regulations, Monitoring, Moral Hazard.
JEL Classification: J41, L23.

1

1   Introduction

The way firms are organized to provide incentives to their employees varies across countries and changes over time. Consider the measure of organizational form depicted in Figure 1, which plots the ratio of managerial to production workers in six countries over the past couple of decades.[1] The U.S. and Canada have more managers per worker than the other countries. Moreover, while the ratio of managers to workers is constant or increasing only slightly in Italy, Spain, Japan and Norway, it appears to increase rapidly in the U.S. and Canada. Crude as this measure of organization may be, it is consistent with what a large industrial relations and business history literature has depicted about international and temporal variations in business practice and organization.[2]

What could account for these differences in organization? One possibility is technology. But are the technologies in use across industries in the advanced industrial nations so different that they could explain the dramatic variations depicted in Figure 1? While certainly a logical possibility, there appears to be no evidence to suggest that these organizational differences are merely consequences of exogenous technological differences. In this paper we therefore take an alternative approach and argue that labor demand and regulations lead to endogenous differences in organizational forms.

Corporate structure itself is a choice variable for firms, and like many of their decisions, will be determined partly by market conditions. We construct a simple general equilibrium model in which conditions in the labor

---

[1]The data plotted are ratios of managers to non-agricultural, non-managerial workers, calculated from the Labor Statistics of the International Labor Organization. The sources of the data are the labor force surveys of the respective countries (Current Population Survey for the U.S.). These data have to be interpreted carefully, as the definition of a manager may vary across countries. Other countries do not have enough years to construct a time-series in the ILO data set, but cross-sectional comparsions are in line with the data reported here: with the exception of the U.K., all other countries appear to have lower ratios of managerial workers in their workforces than the U.S. and Canada.

[2]See, for example, Appelbaum and Batt (1995) for an overview, Chandler (1977) for a history of U.S. frms, and Freeman and Lazear (1994) on the contrast of some aspects of U.S. and German labor relations.
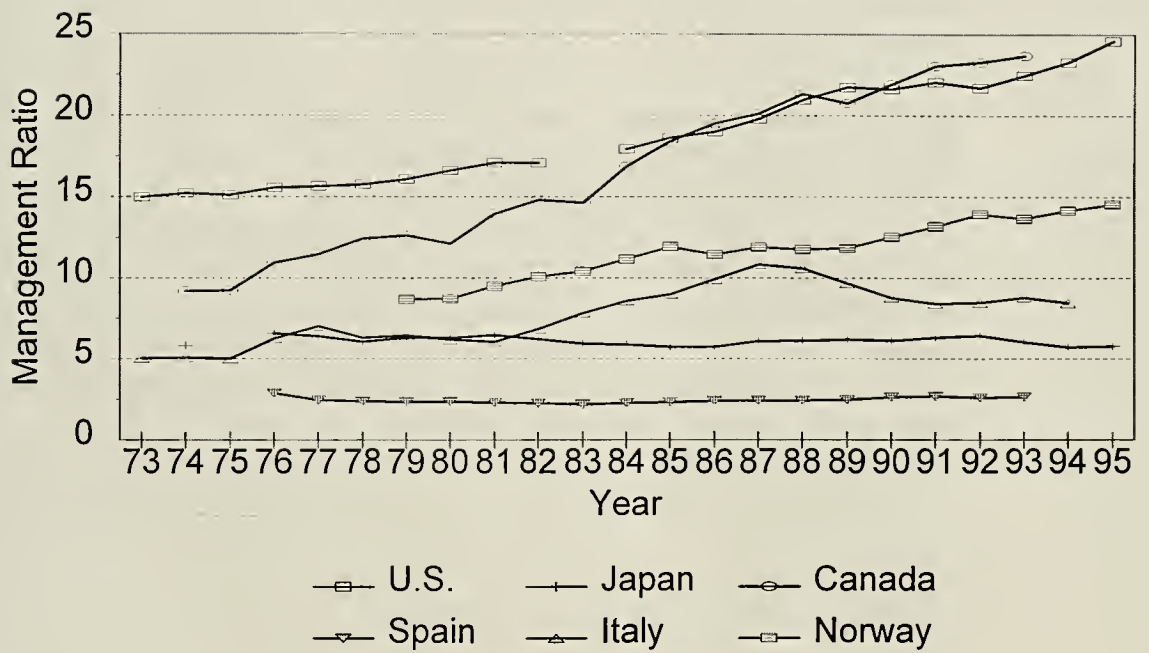
Figure 1: Trends in the ratio of managerial employees to non-managerial, non-agricultural workers in six countries. Source ILO Labor Statistics.

market — both supply-and-demand and regulatory — will lead firms to make different organizational choices. Moreover, we will find that firms' responses to changes in labor market fundamentals are in line with the time trends indicated in Figure 1. We will also show that the organizational choices of firms not only respond to the state of the macroeconomy, but can also have a substantial influence on its performance. This is because firms spend considerable resources on monitoring, which could otherwise be used for directly productive activities. Therefore, it is important to take the organizational implications of labor market policies into account in calculating their welfare consequences.

The logic of our approach is best understood by considering the incentives of a single worker in a profit-maximizing firm. After signing on with the firm, the worker takes an unobserved effort decision: he may work or shirk. Although this effort choice is not directly observed, the firm can detect it with a certain probability that depends on the amount of resources devoted to monitoring.[3] The contract between worker and firm will specify a compensation level which depends on whether he has been detected shirking. Crucially, we assume that there are liability limitations on workers: no contract can punish a worker arbitrarily severely. As is well-known, this will lead to equilibrium rents (efficiency-wages) for workers in order to induce them to exert effort.

In making his effort decision, the worker takes account of three factors: (i) the wage (rent) he is risking to lose by shirking; (ii) his payoff if fired for having shirked; and (iii) the probability of being detected when shirking. The key point is that all three will be affected not merely by the technologies of the firm or legal contractual restrictions, *but also by market conditions*. The main market conditions that we will focus on in this paper are the state of labor demand and labor market regulations.[4]

---

[3]In this context, monitoring should be interpreted broadly: anything which provides some information about worker effort is valuable to the firm (Holmstrom, 1979). A host of organizational variables, such as the number of management and supervisory personnel, the amount of discretion given to workers, the employment of accountants and consultants, or the use of certain kinds of production or information technologies are all measures of the degree of monitoring. This degree of monitoring, and more specifically the ratio of supervisors and managers to production workers, will be our measure of corporate structure in this paper.

[4]Since organizational forms are costly to restructure, changes in labor market conditions that we refer to should probably be thought of as long-lived — a decade or so — rather

First consider the impact of labor demand on incentives and monitoring. An increase in labor demand creates three effects on worker's incentives corresponding to three factors that the worker takes into account in making his effort decision. The first is the *ex ante utility* effect: in a tighter labor market, the *ex ante utility* and the equilibrium wages of workers are higher because firms are competing in order to attract workers.[5] In our world where limited liability constraints prevent negative wages, a high *level* of compensation naturally translates into *high powered incentives.* In other words, when firms are forced to pay high wages to workers because of *market conditions,* they can use these attractive wages to provide them with the right incentives and do not need a high level of monitoring.

The second force is the *ex-post reservation utility* effect. When labor demand is high workers know that being fired is not a harsh punishment because they can get a new job relatively easily (Shapiro and Stiglitz, 1984). This implies that firms will need to monitor their employees closely when labor demand is high. Finally, in a tight labor market, the demand for the resources used for monitoring will also increase. For example, when workers are used to monitor other workers, the cost of monitoring will increase with the level wages. This *cost-of-monitoring* effect also works in the direction of reducing monitoring in tight labor markets: when the cost of monitoring is high as in times of buoyant labor demand, firms will want to use less of it.

We will show that the first and the third effects always dominate the second: *when higher labor demand increases wages, the amount monitoring is reduced.* Another set of variables that vary across countries and time periods is labor market regulations and institutions. In particular, unions and minimum wage type regulations increase wages relative to labor productivity. Again the *ex ante utility* effect comes into action and predicts that labor market institutions that increase wages should lead to less monitoring.[6]

One reason to be interested in the choice of corporate structure is that it has important macroeconomic implications in our economy. The basic

than of business-cycle frequency.

[5]The firm is offering a high ex ante utility, rather than an ex post utility, because even if the worker is hired with an attractive contract, he will receive a low wage and will be fired if he is caught shirking. This distinction between ex-ante and ex-post values will be important in our analysis.

[6]The exception of course is regulations that directly or indirectly prevent firms from firing workers that shirk.

result we obtain here is that the decentralized equilibrium spends excessive resources on monitoring, and since these resources could have been used more productively, it fails to maximize net output. The intuitive reason is that monitoring is at some level a type of "rent-seeking" activity: it enables the firm to reduce wages, transferring resources from workers to firms. A social planner who cares only about aggregate output would want to raise payments to workers in order to save on monitoring costs.

Now, consider the implications of our model for the cross-country trends shown in Figure 1. Many economists believe that due labor market regulations and more powerful unions, wages are higher relative to labor productivity in Europe than in North America (e.g. Layard et al., 1991, OECD, 1994). The immediate implication of our model is that European firms should spend less on monitoring, and in Figure 1, it appears that Canada and the U.S. have many more managers than other countries. Also surprisingly, despite the more intense wage pressure and the stagnant employment, output has grown at the same rate in Europe as in the U.S., and labor productivity has grown faster (e.g. Houseman, 1995). This is consistent with our model which predicts that an economy that spends a large fraction of its productive resources on monitoring should have relatively low productivity, because monitoring is partly unproductive. Therefore, when their impact on corporate structure is taken into account, labor market regulations, which are inefficient in a number of dimensions, may have less detrimental effects than the conventional wisdom suggests, and may even increase total output, even though they will always reduce employment.

Next, consider the changes in the U.S. labor market over the past two decades. Wages for unskilled and production workers have fallen by as much as 30% (e.g. Freeman, 1995) and Figure 1 suggests that in the mean time, the ratio of managers in total employment has increased rapidly. The conventional wisdom is that a combination of globalization and technological changes has reduced demand for unskilled workers (e.g. Berman, Bound and Grilliches, 1992, Katz and Murphy, 1992). Our theory suggests that whatever the reason for the fall in the wages of production workers, the corporate structure designed to control and motivate them has to change substantially. Therefore, in our theory, the increase in the ratio of managers and some of the other organizational changes may be the result of firms' efforts to restore worker incentives eroded by falling wages. Once more our theory also predicts that as the amount of resources spent in monitoring increase, la-

5

bor productivity would be lower, which is also consistent with recent U.S. trends. Finally, the extension of our model to two types of workers in Section 4 will predict that a reduction in the demand for production workers will not only reduce their wages and increase monitoring, but may also increase the salaries of college graduate workers who are more heavily used in monitoring activities.

Our work is clearly related to the "efficiency wage" literature of a decade ago (Foster and Wan, 1984; Bulow and Summers, 1985; see Katz, 1987, and Weiss, 1990, for surveys), especially to the work of Shapiro and Stiglitz (1984). The main difference of our analysis is that we endogenize the monitoring technology and try to understand some cross-country patterns and temporal developments in the organization of firms through this lens. To our knowledge, ours is the only model that analyzes the impact of labor demand and regulations on corporate structure and economic performance.[7]

Our model in Section 2 is a static one which only focuses on the *ex-ante utility* effect we mentioned above but nevertheless illustrates the basic mechanism at work. In Section 3, we generalize our model to a dynamic setting which also incorporates the *ex-post utility* and *cost-of-monitoring* effects discussed above. The setup is based on Shapiro and Stiglitz's model, but nests their model as well as our model of section 2 as special cases. We show that in the original Shapiro-Stiglitz model, labor market regulations that increase wages will have the same effect as in our static model, but a change in labor demand conditions could leave the degree of monitoring unchanged. This is because the trade-off between wages and monitoring is such that firms prefer to increase wages but leave monitoring unaffected in response to a tightening labor market. We demonstrate, however, that this result is not robust. For example, if firms can have contractual arrangements with their workers (as in our static model), or if the cost of monitoring is endogenized so that it changes with the state of labor demand, then a tighter labor market will lead to less monitoring, as our basic and more tractable model of section 2 shows.

The paper concludes with some extensions of our basic framework. First of all, to the extent that more educated workers are engaged in monitoring,

---

[7]Calvo and Wellisz (1979) also endogenize monitoring in an efficiency wage model but without our focus on the determinants of corporate structure. Finally, Gordon (1996) has also pointed out some of the same differences between the U.S. and other economies' corporate organizations, but sought to explain these differences by arguing that corporate bureaucracies have a tendency to expand, and they have been allowed to do so in the U.S.

an increase in the amount of monitoring activities will increase their earnings relative to the wages of production workers, thus linking wage inequality to corporate structure. Secondly, if one of the roles of information technology is to provide more information about worker behavior and thus facilitate monitoring, the changes in labor market conditions that we discuss will change the demand for information technology. Finally we briefly discuss the possible responses of firms' use of long-term contracts to labor market changes. The main conclusion here is that the value of these contracts depends on prevailing wages and monitoring costs, and so, like other aspects of organization, their use will be governed by the labor market equilibrium.

2   A Static Model

We start with a one-period model which illustrates the basic ideas in the simplest environment. In particular, it focuses on the *ex-ante utility* effect, and abstracts from the other two effects discussed in the introduction. Those will be incorporated in the dynamic model of the next section and shown not to affect the basic qualitative conclusions reached with the static model.

2.1   Basics

Consider a one-period economy consisting of a continuum of measure N of workers and a continuum of measure 1 of firm owners who are different from the workers.[8] Each firm i has the production function $AF(L_i)$ where L_i is the number of workers it hires who choose to exert effort. Workers who shirk (do not exert effort) are not productive. In a world without moral hazard,

---

[8]We are implicitly assuming that firms are agents who own some sort of capital (human or physical) for which the market is imperfect. Workers are then agents who are not initially endowed with such capital. This prevents free entry which would compete away all firm profits (which are really just returns to this capital). The crucial "limited liability" assumption we make below simultaneously helps to explain why workers cannot post incentive bonds (which would obviate the need for monitoring), why they have imperfect access to capital and cannot therefore form their own firms, and why there is a covariance of power and level of compensation. We abstract from further consideration of the capital market to keep things tractable. For a model which explicitly treats the role of capital market imperfections in determining the type and efficiency of organizational form, see Legros and Newman (1996).

7

firms would simply contract with the workers to exert effort, but in our economy this is impossible because firms do not directly observe whether their employees have exerted effort or not. We also assume that firms are large, so that the output of an individual worker is not observable and therefore not contractible.

A firm can use other information to give the correct incentives to its workers. A worker's actions affect the probability distribution of some observable signal on the basis of which the firm compensates him (e.g. Holmstrom, 1979). Specifically, when the worker exerts effort, this signal takes the value 1. When he shirks, this signal is equal to 1 with probability $1 - q_i$ and 0 with probability q_i. The worker, like the firm, is risk-neutral and maximizes income minus effort cost which is denoted by e.

The probability of detecting low effort by the worker, q_i, is determined by a host of factors including the production technology used by the firm, the numbers of supervisors, managers and accountants, and more generally the information technology of the firm (e.g. computers and cameras). Our analysis turns on the fact that firms are able to choose many of these attributes of organization; thus q_i will be a key decision variable in our analysis. We assume that $q_i = q(m_i)$ where m_i is the degree of monitoring per worker by firm i; the cost of monitoring for firm i which hires L_i workers is sm_iL_i. For example, we can think of m_i as the number of managers per production worker and s as the salary of managers. For now, s is fixed and exogenous. In the next section, we will endogenize s as an equilibrium outcome. We assume that q is increasing, concave and differentiable with $q(0) = 0$ and $q(m) < 1$ for all m. The choice of q_i in our model will be the crucial aspect of organizational form.[9]

Since there is a limited liability constraint, workers cannot be paid a negative wage, and the worst thing that can happen to a worker is to receive an income of zero. Since all agents are risk-neutral, without loss of generality we can restrict attention to the case where workers are paid zero when caught shirking. Therefore, the incentive compatibility constraint of a worker employed in firm i can be written as:

$$w_i - e \geq (1 - q_i)w_i$$

---

[9]In fact, throught most of the paper q will be the only endogenous aspect of organizational form; in section 5 we will discuss some other dimensions of corporate structure.

If the worker exerts effort, he gets utility $w_i - e$, which gives the left hand side of the expression. If he chooses to shirk, he gets caught with probability q_i and receives zero. If he is not caught, he gets w_i without suffering the cost of effort. This gives the right hand side of the expression.

Firm i's maximization problem can be written as:

$$\max_{w_i, L_i, q_i} \Pi = AF(L_i) - w_i L_i - s m_i L_i \tag{1}$$

subject to:

$$w_i \geq \frac{e}{q(m_i)} \tag{2}$$

$$w_i - e \geq \underline{u} \tag{3}$$

The first constraint is the incentive compatibility condition rearranged, and the second is the participation constraint where $\underline{u}$ is the *ex ante* reservation utility (outside option) of the worker.

As we pointed out in the Introduction, it is important to bear in mind the difference between the *ex ante* and *ex post* outside options. These play distinct roles in the worker's incentive problem, and are affected differently by market conditions. Specifically, if the worker gets fired for shirking, he does *not* receive $\underline{u}$ but instead gets 0, *ex post* outside option (recall there is no more production after the first period). On the other hand, the firm takes the *ex ante* reservation utility $\underline{u}$ as given: constraint (3) reflects the fact that it is not enough for a firm to convince the worker to exert effort once he has joined the firm, but it also has to convince them to join the firm in the first place.

Observe that the problem (1) has a recursive structure: m and w can be determined first without reference to L by minimizing the cost of a worker $w + sm$ subject to (2) and (3); then, once this cost is determined, the profit maximizing level of employment can be found.[10] Each subproblem is strictly convex, so the solution is uniquely determined, and all firms will make the same choices: $m_i = m$, $w_i = w$ and $L_i = L$. In other words, the equilibrium will be symmetric.

Another useful observation is:

---

[10]The recursiveness of this problem is similar to that in Calvo and Wellisz (1979).

Lemma 1 *In equilibrium, the incentive compatibility constraint, (2), always binds.*

To see why this is true, note that if it were not, the firm could lower q, and increase profits without affecting anything else. This differs from the simplest moral hazard problem with fixed q in which the incentive compatibility constraint (2) could be slack.

By contrast, the participation constraint (3) may or may not bind. The comparative statics of the solution have a very different character depending on whether it does. The two situations are sketched in Figures 2 and 3. When (3) does not bind, the solution is characterized by the tangency of the (2) with the per-worker cost $w + sm$ (Figure 2). Call this solution (w^*, m^*), where:

$$\frac{eq'(m^*)}{(q(m^*))^2} = s \tag{4}$$

$$w^* = \frac{e}{q(m^*)}.$$

In this case, because the participation constraint (3) does not bind, w and m are given by (4) and small changes in $\underline{u}$ leave these variables unchanged. In contrast, if (3) binds, w is determined directly from this constraint as equal to $\underline{u} + e$, and an increase in $\underline{u}$ causes the firm to raise this wage, and since (2) holds, the firm will also reduce the amount of information gathering, m.

What determines whether (3) binds? Let $\hat{w}$ and $\hat{m}$ be the per-worker cost minimizing wage and monitoring levels (which are not equal to w^* and m^* when (3) binds). Then, labor demand of a representative firm solves:

$$F'(\hat{L}) = \hat{w} + s\hat{m}. \tag{5}$$

Using labor demand, we can determine $\underline{u}$, workers' *ex ante* reservation utility from market equilibrium. It depends on how many jobs there are. If aggregate demand $\hat{L}$ is less than N, then a worker who turns down a job is not sure to get another; in this case, $\underline{u} = \frac{\hat{L}}{N}(\hat{w} - e) + (1 - \frac{\hat{L}}{N})z$, where z is an unemployment benefit that a worker who cannot find a job receives.[11] The

---

[11] Here we have assumed for simplicity that a worker who gets fired from his job does not receive z.
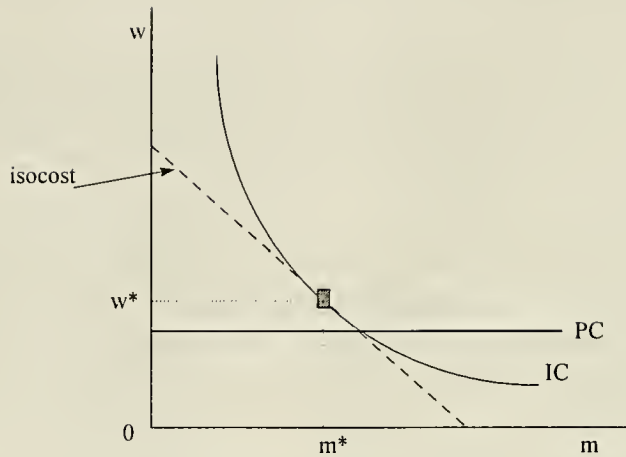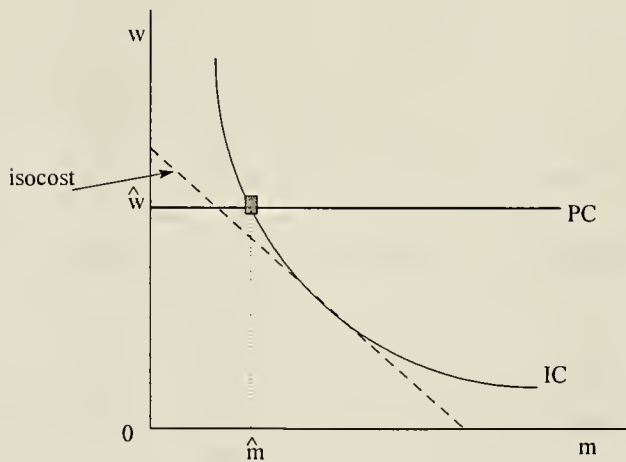
Figure 2: Participation Constraint is Slack.

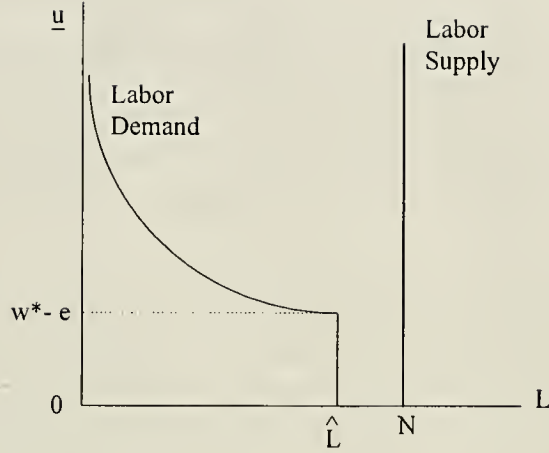Figure 3: Participation Constraint is Binding.

Figure 4: Participation Constraint is Slack.

unemployment benefit z will be useful for some of our comparative statics below, and we always assume that z is not large enough to shut down the economy.

When $\hat{L} = N$, there are always firms who want to hire an unemployed worker at the beginning of the period, and thus $\underline{u} = \hat{w} - e$. If there is excess supply of workers, i.e. $\hat{L} < N$, then firms can set the wage as low as they want, and so they will choose the profit maximizing wage level w^* as given by (4). In contrast, with full employment, firms have to pay a wage equal to $\underline{u} + e$ which will generically exceed the (unconstrained) profit maximizing wage rate w^*. Therefore, we can think of labor demand as a function of $\underline{u}$, the reservation utility of workers: firms are "utility-takers" rather than price-takers. Figures 4 and 5 show the two cases; the outcome depends on the state of labor demand. More importantly, the comparative statics of organization are very different in the two cases.

An equilibrium in this economy is then a vector $(\underline{u}, \hat{w}, \hat{m}, \hat{L})$ such that (i) given $\underline{u}$, $(\hat{w}, \hat{m}, \hat{L})$ are chosen to maximize (1) subject to (2) and (3); (ii) $\hat{L} \leq N$; and (iii) $\underline{u} = z + \min\left\{1, \frac{\hat{L}}{N}\right\}(\hat{w} - e - z)$. Note that workers' reservation utility level, $\underline{u}$, plays the role of a price in equilibrating the market.

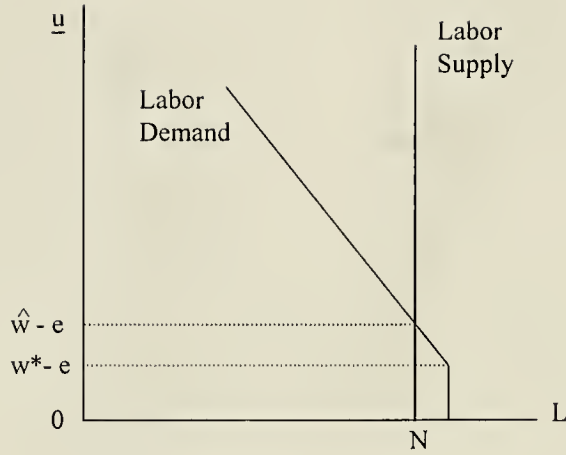**Proposition 1** *An equilibrium exists, is unique and takes one of two forms.*

12

Figure 5: Participation Constraint is Binding.

1. *Full Employment Equilibrium (FEE) in which (3) holds as an equality, thus $\underline{u} = \hat{w} - e$.*

2. *Unemployment Equilibrium (UE) in which (3) is slack and thus $\underline{u} < \hat{w} - e$ and $\hat{w} = w^*$ and $\hat{m} = m^*$ as given by (4).*

The proof of this result is straightforward and is omitted; inspection of Figures 4 and 5 should suffice to make it plausible. In FEE, the participation constraint (3) binds, $\hat{w} > w^*$ and $\hat{m} < m^*$. In this case, in order to attract workers, the market forces firms to pay wages higher than their unconstrained optimum w^*; as a consequence, they engage in less than their privately optimal level of monitoring m^*. By contrast, in UE when (3) is slack, there is an "excess supply" of workers, and firms choose (w^*, m^*). Cutting wages below w^* would still attract workers, but would be unprofitable for firms because in order to ensure incentives, they would have to increase m above their optimum m^*.

13

2.2 Comparative Statics

Now, we will carry out three comparative static exercises. First, we will look at the impact of changing A (or increasing N) which shifts labor demand.[12] Second, we will analyze the impact of imposing a wage floor $\underline{w}$. Finally, we will analyze the implications of a change in unemployment benefits, z. In all three cases, it will matter whether or not the participation constraint, (3), binds.

First, consider a small increase A and suppose that (3) is slack. The tangency between (2) and the per worker cost, shown in Figure 2, is unaffected. Therefore, neither w nor m change. Instead, the demand for labor in Figure 4 shifts to the right and firms hire more workers. As long as (3) is slack (that is as long as the vertical portion of labor demand remains to the left of N in Figure 4), firms will continue to choose their (market) unconstrained optimum, (w^*, m^*), which is independent of the marginal product of labor: as a result, changes in labor demand do not affect the organizational form of the firm.

If instead (3) holds as an equality, comparative static results will be different. In this case, (2), (5), and $L = N$ jointly determine $\hat{q}$ and $\hat{w}$. An increase in A induces firms to demand more labor, increasing $\hat{w}$. Since (2) holds, this reduces $\hat{q}$ as can be seen by shifting the PC curve up in Figure 5. Therefore, when (3) holds, an improvement in the state of labor demand reduces monitoring. The intuition is closely related to the fact that workers are subject to limited liability. When workers cannot be paid negative amounts, the level of their wages is directly related to the *power* of the incentives. The higher are their wages, the more they have to lose by being fired and thus the less willing they are to shirk.

There is another channel which also links the state of the economy to organizational form. Suppose that when more workers are hired, *total* resources spent on monitoring increase[13]; the cost of monitoring s may increase. This would induce firms to make less use of monitoring. In fact, the same result

---

[12]More generally, all our results would hold with a production of the form $F(K, L)$ with constant returns to scale. Here, increasing the supply of workers, N, should be thought as reducing the capital-labor ratio in this economy. If capital could adjust immediately in response to a change in N so that capital labor ratio was unchanged, then there would be no impact on the equilibrium.

[13]Recall that total resources spent on monitoring are mL, so this is possible even if average resources m are decreasing.

would obtain whenever monitoring uses any factor which has a price covarying with the state of labor demand. These issues will be discussed in more detail in the next section when we endogenize s.

Next, suppose that government introduces a wage floor $\underline{w}$ above the equilibrium wage (or alternatively, unions demand a higher wage than would have prevailed in the non-unionized economy). It is straightforward to see that Lemma 1 still holds so that the incentive compatibility constraint (2) will never be slack. Therefore, a higher wage will simply move firms along the IC curve in Figure 3 and reduce m. However, this will also increase total cost of hiring a worker, reducing employment.

Finally, suppose that z changes and that $\underline{u} > z$. Then it is easy once again to verify that if (3) is slack, a small change in z affects neither m nor w. On the other hand, when $\underline{u} = z$, a rise in z increases w and reduces m. Also it is important to note that both high z and high $\underline{w}$ make an unemployment equilibrium more likely than a full employment equilibrium, whereas a high level of labor productivity A makes a full employment equilibrium more likely.

These comparative static results suggest a way of thinking about the corporate structure in the U.S. and Europe. First, European economies, characterized by high minimum wages and unemployment benefits,[14] are more likely to be in unemployment equilibrium, and thus our model suggests that they should have less monitoring, lower m, and thus a lower ratio of managerial to production workers. This is the pattern that emerges from the ILO data reported in Figure 1. The comparative statics with respect to A suggest that a change in the demand for production workers (say due to technical change or international trade) should have a very different impact in an economy in full employment equilibrium as compared to an economy in unemployment equilibrium. Once again, if the U.S. is thought to be in the full employment regime and the more regulated European economies in the unemployment regime, our simple model predicts that in response to a falling demand for unskilled and semi-skilled workers, wages should fall in the U.S. and the degree monitoring should increase. In contrast, in Europe,

---

[14]Another labor market regulation that is common in Europe is severance pay (firing costs). These are not as straightforward to incorporate into our model. At one level, they would act similar to an increase to z, but they would also make firing less desirable for firms and perhaps make the threat of firing less credible. This will tend to weaken worker incentives, which will tend to raise monitoring levels and be detrimental to aggregate performance.

only unemployment rates should increase. Therefore, even though many important effects are left out by these simple comparative statics, the overall effects resemble the actual trends.

2.3 Welfare

Consider the aggregate surplus Y generated by the economy:

$$Y = AF(L) - smL - eL, \tag{6}$$

where $AF(L)$ is total output, and eL and smL are the (social) input costs.

In this economy, the equilibrium is constrained Pareto efficient: subject to the informational constraints, a social planner could not increase the utility of workers without hurting the owners. But total surplus Y will never be maximized in laissez-faire equilibrium:

Proposition 2 *The decentralized equilibrium never maximizes Y. Subsidizing w and taxing profits would increase Y.*

This result follows from noting that if we can reduce q without changing L, then Y increases. A tax on profits used to subsidize w relaxes the incentive constraint (2) and allows a reduction in monitoring.[15] Indeed, the second-best allocation which maximizes Y subject to (2) would set wages as high as possible subject to zero profits for firms. Suppose that the second-best optimal level of employment is $\tilde{L}$, then we have:[16]

$$\tilde{w} + sq^{-1}\left(\frac{e}{\tilde{w}}\right) = \frac{AF(\tilde{L})}{\tilde{L}} \tag{7}$$

In this allocation, all firms would be making zero-profits; since in the decentralized allocation, due to decreasing returns, they are always making positive profits, the two will never coincide.

---

[15]However, recall that w is the wage that workers receive only when they are not caught shirking. Subsidizing wages irrespective of whether workers are caught shirking or not would not affect the incentive compatibility constraint and would not have this beneficial effect.

[16]Since the social cost of a worker is $e + sm$, as long as m and s cannot be made equal to zero, the optimal level of employment will be lower than it would be under the full-information first-best. However, since the planner is minimizing this social cost, he will always want to employ at least as many workers as the decentralized equilibrium would.

A different intuition for why the decentralized equilibrium fails to maximize net output is as follows: part of the expenditure on monitoring, smL, can be interpreted as "rent-seeking" by firms. That is, firms are expending resources to reduce wages — they are trying to minimize the private cost of a worker $w + sm$ — which is to a first-order approximation, a pure transfer from workers to firms. A social planner who cares only about the size of the national product wants to minimize $e + sm$, and therefore would spend less on monitoring, increasing net output.

Figure 6 draws the equilibrium, first-best and second-best surpluses as a function of the supply of labor N for a parametric case. Over the range of N depicted, the first-best, which prevails when moral hazard is absent, is simply given by $F(N) - Ne$. The second-best adopts the wage rule (7) and also chooses full employment in this range. Finally, the equilibrium is the outcome characterized in Proposition 1. Observe that the equilibrium surplus is decreasing in N. This is because high levels of N reduce wages through the usual supply effect and thus induce firms to increase m in order to ensure incentive compatibility. This suggests that over a certain range labor market policies which increase wages and reduce employment will actually increase surplus, or at least have only a small effect on aggregate welfare.[17]

This discussion has implications for the contrast of the performance of European and American labor markets. Our earlier interpretation of Europe as in an unemployment equilibrium and the U.S. in full employment is not uncommon (though as noted above, we are not aware of any other work deriving implications for organizational form from this contrast). However, the conventional wisdom is that the labor market regulations, unions and

---

[17]In fact, we can show that the equilibrium surplus must be declining in a neighborhood of N^{FE}, which is the maximum labor force size compatible with full employment equilibrium. To see this, observe that at N^{FE}, the equilibrium wage is w^* and monitoring level is m^*. Now suppose that the labor force is reduced a bit. This results in an increase in the equilibrium wage, and an associated decrease in the level of montitoring. But by the envelope theorem, at their unconstrained optimum, firms suffer no increase in the cost of a worker, since they were at the otpimium w^*. So they will not decrease their demand for workers, and gross aggregate output is unaffected. But since the level of monitoring falls, net output increases, as we claimed. Indeed, a straightforward calculation shows that along the equilibrium surplus curve, $\frac{dY}{dN} = w - e - sN\frac{dm}{dN}$ for $N < N^{FE}$; with $\frac{dm}{dN} = \frac{AF''}{s - \frac{eq'}{q^2}} > 0$ in this range, $\frac{dY}{dN} \to -\infty$ as $N \uparrow N^{FE}$. Thus the maximal laissez-faire surplus is achieved below N^{FE}.
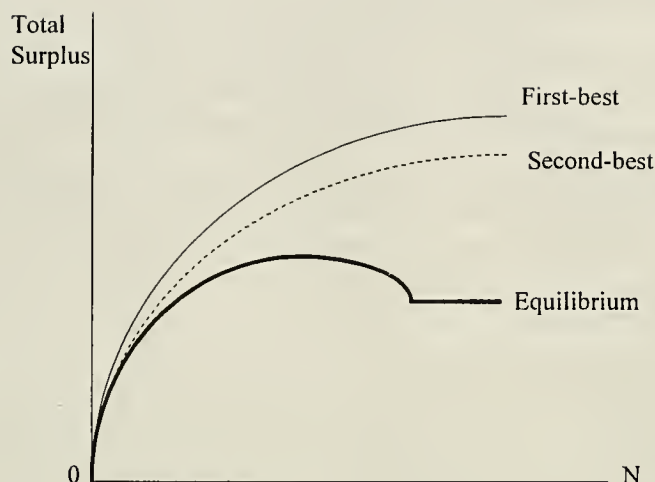
Figure 6: Plot of the total surplus in first-best, second-best and equilibrium.

minimum wages in European labor markets are purely distortionary and reduce output (and net national product). They certainly appear to have led to much lower employment in Europe, but there is no evidence that output growth or labor productivity have suffered in Europe as compared to the U.S. Our model suggests that Europe may be at a different point of the trade-off between wages and monitoring than the U.S. In other words, it is possible that the U.S. has chosen to increase employment, which in the logic of our model implies a high level of m. In contrast, Europe may have chosen relatively low employment, high wages and low m. The rough numbers reported in Figure 1 suggest that this is not totally implausible. And Figure 6 suggests that welfare (total surplus) may be higher in a high m or low m environment. Thus it is perfectly possible for aggregate performance (in terms of output and/or growth) in an economy with high wages and unemployment to equal or even exceed that of a full employment economy. According to this interpretation, the U.S. and Europe may have chosen allocations that differ radically in terms of the (functional) distribution of income, but not much in terms of total output or efficiency.

18

3 A Dynamic Model

We now analyze a dynamic model which will generalize the main results of the previous section and also incorporate some of the effects discussed in the Introduction which were missing from the static model. This will enable us to show that most of the results of the static case generalize to this dynamic environment, and additionally, even in the equivalent of "unemployment regime", an increase in the productivity of production workers will lead to higher wages and less monitoring, thus to a different corporate structure.

3.1 The Environment

As before, there is a measure N of workers and a unit measure of firms. Time is continuous. All agents are risk-neutral, infinitely-lived, and discount the future at the rate r. Workers can be employed either to produce output or to supervise other workers. If a supervisor (manager) is monitoring $1/m$ workers, then a shirking worker is detected with probability $q(m)$. As before, workers are never mistakenly detected shirking. The cost of effort both to production workers and managers is equal to e, and this cost is not incurred if they shirk. Owners undertake the monitoring of managers. If an owner employing L production workers and mL managers exerts effort amL, then each manager is caught shirking with probability $p(a)$. We assume that $q(m)$ and $p(a)$ are smooth, increasing, and concave, with $q(0) = p(0) = 0$ and $q(.)$, $p(.) < 1$. We denote per period wages of production workers by w and the salaries of managers by s. Observe that to focus on our main interest, we are making the extreme assumption that managers are not directly productive: their only role is to gather information and monitor production workers.

This model is closely related to the one studied by Shapiro and Stiglitz (1984). The main differences from their analysis are that (i) q is endogenous; (ii) some workers are employed as supervisors; (iii) there is greater scope for of contracting. On this last point recall that Shapiro and Stiglitz assume that if a worker is caught shirking, he suffers no monetary penalty but instead is just fired. In contrast, in our previous analysis, we assumed a worker only gets his wage if he is not caught shirking. Reality presumably lies somewhere in between. It is difficult to retain wages for work already performed, but workers lose their bonuses, their chances of promotion and their pensions when they are fired. We shall model this in a simple reduced form way by

19

supposing that a worker (or manager) is caught shirking can be made to suffer a financial loss of αw (or αs), where $\alpha \geq 0$.

3.2 Characterization of Equilibrium

Since there are no adjustment costs, every period firm i maximizes:

$$\Pi_i = AF(L_i) - w_i L_i - s_i m_i L_i - a_i m_i L_i$$

by choosing L_i, s_i, w_i, m_i, and a_i subject to a participation constraints and incentive compatibility constraints for each occupation.

To write the incentive compatibility constraints, we need to work with Bellman equations. Let us define V_U, V_E^P, V_E^M, V_S^P, V_S^M respectively as the expected present discounted values of unemployment; employment as a production worker and exerting effort; employment as a supervisor and exerting effort; employment as a production worker and shirking; and employment as a supervisor and shirking. We will use i as an additional argument to indicate when these values are in principle different across firms. Following Shapiro and Stiglitz (1984), we will concentrate on steady states and thus impose that the time derivatives for all these value functions are equal to zero.

Using standard arguments, we can write:

$$rV_U = z + x\left[\mu V_E^M + (1-\mu)V_E^P - V_U\right] \tag{8}$$

where z is utility when unemployed, x is the probability (flow rate) of getting a job and μ is the fraction of jobs that are managerial. Intuitively, an unemployed worker gets a job with probability x; with probability μ, this is a management job, and with probability $1 - \mu$, he becomes a production worker. In both cases he gains the expected present value of the relevant job and loses the present value of unemployment.

All firms take the value of unemployment V^U as given by the market, but through their choice of wages and corporate structure affect all other values, hence they are indexed by i. For firm i, we have:

$$
\begin{aligned}
rV^P(i) &= w_i - e + b\left[V_U - V^P(i)\right] \tag{9}\\
rV_E^M(i) &= s_i - e + b\left[V_U - V_E^M(i)\right],
\end{aligned}
$$

where w_i is the worker's wage and s_i the manager's salary in firm i, and b is the exogenous flow rate at which jobs dissolve. In equilibrium since all workers exert effort there are no firings for shirking. However, the value of shirking will be important in determining the incentive compatibility conditions. These are written as:

$$rV^P(i) = w_i - q_i\alpha w_i + (b + q_i)\left[V_U - V^P(i)\right] \qquad (10)$$
$$rV_S^M(i) = s_i - p_i\alpha s_i + (b + p_i)\left[V_U - V_S^M(i)\right]$$

The main difference between (9) and (10) is that in (10), there is no cost of effort e, but the relation comes to an end faster as shirking employees are caught (at the rates q and p). When they are caught, they also lose a proportion α of their wages.

The two incentive compatibility constraints are:

$$V_E^P(i) \geq V_S^P(i) \qquad (11)$$
$$V_E^M(i) \geq V_S^M(i)$$

Simple algebra enables us to write (11) as:

$$w_i \geq \frac{\frac{e}{q_i}(r + b) + z + e + rV_U}{(r + b)\alpha + 1} \qquad (12)$$
$$s_i \geq \frac{\frac{e}{p_i}(r + b) + z + e + rV_U}{(r + b)\alpha + 1}$$

As in the previous section, both incentive compatibility constraints will bind (otherwise m or a could be reduced).

Since there are two different occupations, there are also two participation constraints:

$$V_E^P(i) = \frac{w_i - e + bV_U}{r + b} \geq V_U \qquad (13)$$
$$V_E^M(i) = \frac{s_i - e + bV_U}{r + b} \geq V_U$$

21

Although it is somewhat obscured by the notation and restriction to steady state analysis, it is important to note that V_U, the value of unemployment, is playing a dual role here. First, it is very similar to $\underline{u}$ in Section 2, the *ex ante* reservation utility. But here V_U also enters into the *ex post* reservation utility: a worker who is caught shirking receives not 0 as he did in Section 2, but $(1 - \alpha)w + V_U$. In fact, in steady state with $\dot{V}_U = 0$, if $\alpha = 0$, these two concepts will coincide at all points.

Now, the problem of firm i can be written as:

$$\max_{L_i, m_i, a_i, w_i, s_i} AF(L_i) - w_i L - m_i (s_i + a_i) L_i \tag{14}$$

subject to (12) and (13). This problem is more complicated than (1) in Section 2, but it is still straightforward to establish that it consists of a recursive set of strictly convex optimization problems and therefore that the solution is unique. Thus, we will have $w_i = w$, $s_i = s$, $m_i = m$, $a_i = a$ and $L_i = L$ for all i. In particular, we can first determine s and a, then w and q and then finally, L, which will once more simplify the analysis.

A steady state equilibrium is then a vector $\left(\hat{a}, \hat{s}, \hat{m}, \hat{w}, \hat{L}, \hat{V}_U \right)$ in which (i) the sub-vector $\left(\hat{a}, \hat{s}, \hat{m}, \hat{w}, \hat{L} \right)$ maximizes (14) subject to (12) and (13) given $\hat{V}_U$; (ii) $\hat{L}(1 + \hat{m}) \leq N$; and (iii) $\hat{V}_U$ solves (8) with $x = \frac{b(1 + \hat{m})\hat{L}}{N - (1 + \hat{m})\hat{L}}$ and $\mu = \frac{\hat{m}}{1 + \hat{m}}$.[18] Intuitively, an equilibrium requires that given the reservation utility of workers, firms choose the optimal wage, salary and organizational forms and then the reservation utility of workers be determined consistently in general equilibrium.

We first have:

Proposition 3 *A steady state equilibrium* $\left(\hat{a}, \hat{s}, \hat{m}, \hat{w}, \hat{L}, \hat{V}_U \right)$ *always exists.*

The proof employs standard arguments which are sketched in the appendix. In contrast to Proposition 1 uniqueness is no longer guaranteed because of the general equilibrium interactions determining the value of unemployment, V_U.

It is straightforward to see that the equilibrium takes one of three forms, depending on which of the two participation constraints bind:

---

[18] Provided that $\hat{L}(1 + \hat{m}) < N$; if $\hat{L}(1 + \hat{m}) = N$, (8) becomes $V_U = \mu V_E^M + (1 - \mu)V_E^P$.

1. Full Employment Equilibrium (FEE) where both participation constraints in (13) hold as equality. In this equilibrium $\hat{s} = \hat{w} = \hat{V}_U + e$, $\hat{L} = N$ and $\hat{m}$ and $\hat{s}$ solve given (12).

2. Unemployment Equilibrium (UE) where both participation constraints in (13) are slack and $\hat{L} < N$. In this regime $\left(\hat{a}, \hat{s}, \hat{m}, \hat{w}, \hat{L} \right)$ maximizes (14) subject to (12) only.

3. Semi-Constrained Equilibrium (SCE) where one of the participation constraints in (13) hold and the other is slack.

The Semi-Constrained Equilibrium can have either the participation constraint of workers or managers bind, but we think of the case where that of the managers hold, so that $s > w$, as more relevant. The recursive structure of the problem once again helps a lot in the analysis. In the full employment equilibrium, the market dictates what wages must be paid, and thus $\hat{w} = \hat{s} = \hat{V}_U + e$. Once the wages are determined, then the firm minimizes its costs by minimizing monitoring which entails setting $\hat{m}$ and $\hat{s}$ to solve (12). This has an obvious similarity to the full-employment regime of the static model. In contrast, in the Unemployment Equilibrium, both participation constraints are slack, thus the firm is unconstrained by the market and can choose the wage and monitoring levels that maximize profits[19]: $\hat{w} = w^*$, $\hat{s} = s^*$, $\hat{m} = m^*$ and $\hat{a} = a^*$. In other words, as in section 2, when the ex ante reservation utility, $\hat{V}_U$, is sufficiently low that the firm does not have to compete with other firms to obtain workers, it can attain its "market-unconstrained" optimum. In contrast, in the FEE, $\hat{V}_U$ was sufficiently high that the firm was forced to pay $\hat{s} > s^*$ and $\hat{w} > w^*$ and choose $\hat{q} < q^*$ and $\hat{a} < a^*$.

An important point to note is that when $\alpha = 0$, there is no possibility to contract on the wage of the worker when he is caught shirking, so that he receives exactly the same payment as when he is not caught shirking. In this case, a Full Employment Equilibrium is not possible. To see this, note that if the participation constraint binds, then $w = rV_U + e$. Substituting this into (12) and setting $\alpha = 0$ gives a contradiction. This is the case considered by Shapiro and Stiglitz (1984), albeit without endogenous monitoring, and in fact, in their model, equilibrium always entails some positive level of unemployment. In contrast, the same exercise shows that when $\alpha > 0$, there

---

[19]That is the firm is maximizing (14) subject to (12) alone.

will exist a sufficiently high level of V_U such that (12) can be satisfied with $w = rV_U + e$, thus giving a FEE.

3.3 Comparative Statics

Let us start the comparative statics with the Full Employment Equilibrium (which, recall, is only possible when $\alpha > 0$). The following proposition is proved in the appendix:

Proposition 4 *In the FEE,* $\frac{dm}{dA} < 0, \frac{dw}{dA} > 0, \frac{ds}{dA} > 0, \frac{da}{dA} < 0.$

The intuition is exactly the same as in the static model. In the FEE, an improvement in A increases wages (and salaries) and thus makes workers incentives more *powerful*. This moves firms along both the incentive compatibility constraints of workers and managers, and both types of employees are monitored less.

Next, let us turn to the Unemployment Equilibrium. Here, in contrast to the Full Employment Equilibrium, multiple equilibria are possible, and we have to make sure that we are doing comparative statics on the right equilibria. As is well-known in models of multiple equilibria, it is most sensible to look at the extremal equilibria, here defined as those with the highest or lowest value of unemployment, V_U. Then we can state (proof in the appendix):

Proposition 5 *Consider extremal UE. Then* $\frac{dm}{dA} < 0, \frac{dw}{dA} > 0, \frac{ds}{dA} > 0$ *and* $\frac{da}{dA} = 0.$

The intuitive reason for this result is that when A goes up, there is more demand for labor and therefore, wages, and together with wages, salaries increase. One may conjecture that as in the static model, q would remain unchanged because the participation constraints are not binding. However, this conjecture is incorrect due to the cost-of-monitoring effect: the salaries paid to managers are part of the cost of monitoring, and the *cost* of monitoring is higher due to the higher managerial salaries dictated by the market. When monitoring is more costly, firms will want to use less of it, and once again, a more buoyant labor market leads to less monitoring and more discretion for production workers. Similar arguments can also be developed for the case of the Semi-Constrained Equilibrium, and we omit this case. It

24

can be noted at this point that if we were to endogenize monitoring in the exact equivalent of Shapiro and Stiglitz's (1984) set-up with $\alpha = 0$ and no cost-of-monitoring effect, then we would have $\frac{dm}{dA} = \frac{da}{dA} = 0$, that is corporate structure would not respond to changes in the state of labor demand.

Next, it is also straightforward to see that, in this dynamic economy, a binding wage floor due labor market regulations or wage setting by unions will work exactly as before. It will push up wages, and therefore induce firms to reduce monitoring. Therefore, the dynamic model also predicts that European economies characterized with more wage push should have less monitoring. We state this as a result and omit the proof:

Proposition 6 *Suppose that $\underline{w}$ is a wage floor imposed by the government. Then, in any steady state equilibrium, $\frac{dm}{d\underline{w}} \leq 0$.*

3.4 Welfare

Once again, net surplus (or net output) is:

$$Y = AF(L) - (1 + m)Le - mLa$$

where total production is given by the number of production workers, and total effort number of workers in employment is $(1 + m)L$ and they incur the effort cost e and finally, owners incur the monitoring cost a for each monitor, thus a total of mLa.

Proposition 7 *The decentralized equilibrium never maximizes net surplus.*

This proposition again follows by noting that the planner would increase wages and salaries in order to reduce monitoring until there are zero-profits, but in the decentralized equilibrium firms are making positive profits. Taxing profits and subsidizing s and w increases total production as more workers can become producers rather than supervisors.[20]

---

[20]Note that in this case, there are additional issues because the lower of unemployment induces workers to shirk more, thus creating a negative externality on firms. However, as in the original Shapiro and Stiglitz (1984) model, this effect is always dominated.

4 Discussion and Extensions

This paper has developed an approach to the macroeconomics of organization. In our model, organizational forms are designed to provide incentives to workers. When workers cannot be contractually punished arbitrarily severely, low wages naturally imply weak incentives, and firms are induced to choose organizational structures that increase monitoring. Wages may be high either because of labor demand variations or because of labor market regulations. In particular, when labor demand increases, firms reduce monitoring for two reasons: (i) workers are paid higher wages and have better incentives; (ii) monitor's salaries also increase and thus monitoring becomes more expensive. Counteracting these two forces, when labor demand is higher, unemployment is low and does not act as an effective discipline device, but we show that this effect is always dominated by (i) and (ii). We argue that these effects help to explain why organizations differ across countries and over time. The model also shows that the organizational differences can have significant implications for macroeconomic performance.

We now consider some further implications and extensions of our framework.

4.1 Income Distribution

The distribution of income is tied to corporate structure because corporate structure determines both the earnings of production workers, those of managers, and also what fraction of workers become managers.. Also, given that cross-country differences in corporate structure appear to be correlated with wage inequality patterns (i.e. the U.S., the U.K. and Canada have experienced sharper increases in wage inequality than other countries in our sample, e.g. Katz, et al., 1995), it is important to investigate the links between the evolution of corporate structure and income distribution. To address this question, we consider a variant of the model in which there are two types of workers.

4.1.1 The Environment

The two types of workers are capable of doing different kinds of jobs. N "unskilled" workers can only work in production. H "skilled" workers can

either work as managers and monitor workers or they can work as engineers. Total output from production workers is equal to $F(L)$ as before; the measure of firms is still 1. Engineering output is given by $\Phi(E)$, where E is the total number of engineers; we assume that there are no incentive problems for workers in the engineering sector. As before, monitors are not directly productive. We make the standard assumptions on both production functions: F and Φ are increasing and strictly concave and they satisfy Inada type conditions. We assume that entry into the engineering sector is free and each engineer is paid his marginal product. Also, college graduates do not increase their probability of getting into managerial jobs by being unemployed: they can equally well work as engineers and still receive offers of management jobs.[21]

As in the previous section, the flow rate of detecting a worker who shirks is $q(m)$ where $1/m$ is the number of workers monitor by one manager and the flow rate of detecting a shirking manager is $p(a)$; p and q are both increasing and strictly concave. As before, all agents are risk-neutral, infinitely lived and discount the future at the rate r.

4.1.2 Characterization of Steady State Equilibrium

Firm i once more maximizes:

$$AF(L_i) - w_i L_i - s_i m_i L_i - a_i m_i L_i \tag{15}$$

where s_i is salary for the monitors and w_i is the wage rate of the workers. Let us define, V_E^P, V_E^M, V_S^P, V_S^M as the value functions of working and shirking managers and workers. Also differently from the previous section, we need two reservation utilities: V_U, the value of unemployment for unskilled workers, and V_C value of working in the engineering sector for college graduates, which will act as the *ex ante* and *ex post* reservation utility for college graduates since they can always choose this option.

Now we have the equations (9) and (10) determining the value functions as before with the only change that for production workers, the reservation utility is V_U and for managers it is V_C. Combining these two equations, we can write the incentive compatibility constraints in this case as:

---

[21]Thus, there will be no "unemployment" of college graduates; instead there may be equilibria in which engineers would strictly prefer to be managers.

$$w_i \geq \frac{\frac{e}{q(m_i)}(r+b) + z + e + bV_U}{(r+b)\alpha + 1}$$

$$s_i \geq \frac{\frac{e}{p(a_i)}(r+b) + z + e + bV_C}{(r+b)\alpha + 1}$$

$$(16)$$

And the two participation constraints are:

$$V_E^P(i) \geq V_U$$
$$V_E^M(i) \geq V_C$$

$$(17)$$

Once more, the maximization problem of firm i (15) subject to (16) and (17) is strictly concave, thus has a unique solution. Therefore, we have $w_i = w$, $s_i = s$, $m_i = m$, $a_i = a$ and $L_i = L$.

Now, the Bellman equation that determines the reservation utility of unskilled workers is:

$$rV_U = z^P + x^P \left[V_E^P - V_U \right]$$

$$(18)$$

where z^P is the unemployment benefit for production workers, and x^P is their job-finding rate, which in steady state is equal to $x^P = \frac{bL}{N-L}$.

The reservation utility of college graduates can be written as follows:

$$rV_C = \Phi'(E) + x^m \left[V_E^m - V_C \right]$$

$$(19)$$

where $\Phi'(E)$ is the wage they receive in the engineering sector and $x^m = \frac{bmL}{E}$ is the rate at which engineers get managerial job offers, and market clearing for college graduates implies: $E = H - mL$.

Then, an equilibrium is a vector $\left(\hat{a}, \hat{s}, \hat{m}, \hat{w}, \hat{L}, \hat{V}_U, \hat{V}_C \right)$ such that $(\hat{a}, \hat{s}, \hat{m}, \hat{w}, \hat{L})$ maximizes (15) subject to (16) and (17), and $\hat{V}_U$ and $\hat{V}_C$ are given by (18) and (19) with $x^P = \frac{b\hat{L}}{N-\hat{L}}$, $x^m = \frac{b\hat{m}\hat{L}}{\hat{E}}$ and $\hat{E} = H - \hat{m}\hat{L}$.

This model is quite similar to the one-type dynamic model. The next result establishes the existence of a steady state equilibrium.

Proposition 8 *A steady state equilibrium exists and takes one of the following forms:*

1. *Full Employment Equilibrium (FEE) where $\hat{L} = N$, $\hat{w} = r\hat{V}_U + e$, $\hat{s} = \Phi'(H - \hat{m}\hat{L}) + e$ and $\hat{m}$ and $\hat{a}$ are given by (16).*

2. *Unemployment Equilibrium with Managerial Constraint (UEMC) where $\hat{L} < N$, $\hat{w} > r\hat{V}_U + e$, $\hat{s} = \Phi'(H - \hat{m}\hat{L})$*

3. *Unemployment Equilibrium (UE) where $\hat{L} < N$, $\hat{w} > r\hat{V}_U + e$, $\hat{s} > \Phi'(H - \hat{m}\hat{L})$ and $(\hat{a}, \hat{s}, \hat{m}, \hat{w}, \hat{L})$ maximize (15) subject to (16) only.*

4. *Managerial Constraint Equilibrium (MCE) where $\hat{L} = N$, $\hat{w} = r\hat{V}_U + e$, and $\hat{s} > \Phi'(H - \hat{m}\hat{L})$.*

The proof of this result is similar to that of Proposition 3 and is omitted.

4.1.3   Comparative Statics

The comparative static results are very similar to those in section 3. In particular in the FEE, an increase in A (labor demand) leads to higher wages and to lower $m(v)$, that is to less monitoring. What is different, however, is that this increase in A will increase E, the number of college graduates who go into engineering, and thus reduce Φ', and therefore reduce s. Hence, the prediction of the two-type model is that starting from a full employment equilibrium, a reduction in the productivity of production workers will reduce their wages, increase the extent of monitoring but also increase the salaries of managers. Therefore, in the context of our model, some of the trends of the U.S. economy over the past twenty years can be explained quite simply by a reduced demand for production workers. In particular, our model predicts that in response to changes in the demand for production workers, we should observe increasing management ratios and relative managerial salaries, which otherwise have to explained by increasing productivity of managers relative to production workers.

Next consider UEMC. Again as in section 3, focusing on extremal equilibria, an increase in A increases labor demand at given $m(V)$, and this leads to a larger number of skilled workers employed as managers (i.e. mL increases). As a result E falls, this *increases* s. When s increases, as in section 3, the privately optimal amount of monitoring, m, falls (immediately from the first-order condition of the firm with respect to m). Therefore, in UEMC, higher

productivity of workers leads to higher wages both for managers and workers and to less monitoring.

The contrast between this regime and FEE is interesting. In particular, it implies that a reduction in labor demand will reduce wages of all types of labor in the unemployment equilibrium whereas in the full employment equilibrium, it will reduce production workers' wages but increase managerial wages and thus inequality. Once again, this stylized result gives a different way of interpreting the differential trends in the labor markets of Europe and the U.S.

A testable implication is that a binding minimum wage in the FEE not only increases wages, but it also *reduces* monitoring and the wages of managers. This is because as monitoring decreases, there is less demand for college graduates from managerial jobs, and more of them become engineers, and the marginal product of engineers and managerial salaries decline.

4.2 Information Technology

Many of the changes in workplace organization and the structure of wages have been attributed to the changes in information technology. Computers presumably increase the productivity of workers in many tasks, but surely information technology is also very helpful for information gathering and monitoring. Our framework suggests that part of the increased use of computers may be endogenous to changes in labor market conditions, but it also allows for analysis of the effects of exogenous improvements in computers on the labor market.

To incorporate computers into our setup most simply, let us make the extreme assumption that information gathering is the *only* place in which computers are useful. Consider the static model of Section 2 and assume that $q = q(m, c)$, where c is the computer input per worker into the monitoring process, and q is smooth, supermodular, strictly concave and strictly increasing. Let the cost of computers be γ. Then the maximization problem of firms can be written as:

$$\max_{L,m,c,w} AF(L) - wL - smL - \gamma cL$$

$$s.t. w \geq \frac{e}{q} \tag{20}$$

30

$$w - e \geq \underline{u} \qquad (21)$$

$$q = q(m, c) \qquad (22)$$

Once again, this problem has a recursive structure: the firm wants to minimize the per-worker cost subject to the three constraints. This it will do by ensuring that (20) binds. Then whatever the optimal wage $\hat{w}$, the firm will minimize $sm + \gamma c$ subject to $q(m, c) = e/\hat{w}$. The last is a well-behaved convex problem leading to solutions $c(w)$ and $m(w)$ which are increasing in w. The solution to the problem $\min_{w} w + \gamma c(w) + sm(w)$ s.t. $w - e \geq \underline{u}$ must be nondecreasiong in $\underline{u}$ (increasing when (21) binds), so that c and m are nondecreasing as well.

If there is a reduction in labor demand under full employment, say due to a fall in A, there will be an increase not only in m but also in c. Therefore, when firms want to increase monitoring, they will make more use of computers. If the increased demand for monitoring also leads to a higher salaries for managers, as in Sections 3 and 4, then this version of the model predicts an increase in the relative wages of employees working with computers (monitors), which is observed in the data (e.g. Krueger, 1993). It would be naive to try to explain changes in the wage structure solely by this mechanism, but it is important to note from this discussion that an increase in the wages of workers using computers does not necessarily mean that these workers have become more productive.

In contrast to this endogenous change in the use of information technology, exogenous changes in the efficiency of information technology can be captured by a reduction in γ. This would increase the desired monitoring level of firms, and in the unemployment equilibrium, would tend to reduce wages of production workers. In contrast, in the full employment case, firms would not increase monitoring because they cannot reduce wages, and instead would tend to reduce the their demand for managers. This has some affinity to the recent developments in the American workplace where some functions of middle managers are being replaced by computers.

4.3 Long-Term Contracts

Another dimension of corporate structure is the form of the contract between the firm and the employees. For example, Japanese firms are often argued to have much longer-term relations with their employees than U.S. firms (e.g. Hashimoto, 1994). It has long been recognized that long-term contracts (LTCs) and similar institutions such as rising tenure-earnings profiles, promotions, and pensions may be powerful incentive devices (see Lazear, 1995). Intuitively, because compensation is deferred, workers have more to lose under the LTC than under a short-term contract (STC) if they shirk. What is perhaps less recognized is that the value of long term contracts depends on the cost of information gathering: LTCs are useful because they enable a firm to save on monitoring.

In this subsection, we analyze a two-period version of the model of Section 2 which allows for long-term contracts. We make the same assumptions as we made there except that now everyone lives for two periods and there is no discounting.

First consider the case of the spot market transactions where the firm hires labor with a STC in each period. The optimal one period contract is identical to the one we characterized in Section 2, which implies that the firm has to satisfy (2). Denote the level of information gathering activity that satisfies (2) by $q(m_S) = \frac{e}{w}$ where w is the equilibrium one-period wage rate. To simplify the exposition, we will suppose throughout that we are in a regime where the participation constraint (3) binds, so the firm effectively takes w as given. Over the two periods, the short term contract will cost the firm $2sm_S + 2w$ per worker.

The alternative organizational form that the firm could adopt is to hire the worker for two periods using a LTC. In this case, the firm can defer some or all of the payment of the worker to the end of the second period. For simplicity consider the most extreme form of this whereby the worker receives zero in the first period and $2w$ at the end of the second period if he is not caught shirking in either period; if he is caught in either period, he gets zero in both periods.

Start with the incentive compatibility constraint in the second period:

$$2w - e \geq (1 - q(m_2))2w, \tag{23}$$

where m_2 is the period-2 monitoring level; if the worker is caught shirking

then, he loses $2w$. Given that he is incentive compatible in the second period, the worker obtains $2w - e$ in the first period if he shirks and is not caught; if he works in the first period he gets $2w - 2e$. Therefore, the incentive compatibility constraint for period 1 of the LTC is:

$$2w - 2e \geq (1 - q(m_1))(2w - e) \tag{24}$$

By the same reasoning as before, these constraints will bind, thus:

$$q(m_1) = \frac{e}{2w - e}$$
$$q(m_2) = \frac{e}{2w}$$

Monitoring and wage costs associated with this contract are then $sm_1 + sm_2 + 2w$. Since $w > e$, both $q(m_1)$ and $q(m_2)$ are less than $q(m_S)$; thus there are fewer resources devoted to monitoring under the LTC than under the STC. The benefit of the LTC is therefore always positive and is equal to

$$s[2m_S - (m_1 + m_2)]. \tag{25}$$

One result is immediate: if the cost s of monitoring declines, the benefit of the LTC falls and, presuming that the cost of using the LTC (in terms of establishing reputation or loss of flexibility to market shocks) to be unchanged, we should expect to see fewer of them.

Secondly, the benefit of the LTC also depends on the level of wages, though here the relationship is more ambiguous. First, all else equal, for very high wages the benefit is negligible, because the level of monitoring is small even for the STC.[22] For wages close to e, the benefit will tend to be high. However, it is possible to show that the maximal benefit is attained at some wage greater than e; a precise characterization depends on the form of $q(\cdot)$. Thus, it is possible that as wages become more dispersed, as they have in the U.S. over the last twenty years, there would be an overall decline in the use of LTCs. The more general point is that as wages and monitoring costs change with labor market conditions, firms will alter their use of LTCs just as they alter other aspects of their organizational form.

---

[22]However, this point does not imply that in practice, high wage jobs should not have long-term contracts, because high wage jobs (e.g. airline pilot, manager, physician) are also very costly to monitor, so they may tend to be governed by LTCs.

5   Appendix: Proofs of Proposition 3, 4, 5

Here we sketch the proofs of the three propositions in section 3. The remaining results in the text have very similar proofs which are not repeated here.

Proof of Proposition 3: The maximization problem of each firm (14) subject to (12) and (13) for given $V_U \equiv V$ defines: $a(V), s(V), m(V), w(V)$ and $L(V)$. As established in the text, these are all functions; by the maximum theorem, they are continuous in V. It is straightforward to establish that $(1 + m(V))L(V)$ is a decreasing function of V, and therefore, the equation $(1 + m(V))L(V) = N$ has a unique solution, which we denote by $\bar{V}$. It also follows immediately from the same monotonicity that $(1 + m(V))L(V) < N$ if and only if $V > \bar{V}$.

Substituting the two value functions in (9) into the right-hand side of (8) we define, for $V > \bar{V}$:

$$G_1(V) \equiv \frac{1}{r} \frac{(r + b)z + x(V)\left[\frac{m(V)s(V)+w(V)}{1+m(V)} - e\right]}{(r + b) + x(V)}, \qquad (26)$$

where $x(V) = \frac{b(1+m(V))L(V)}{N-(1+m(V))L(V)}$. As $V \downarrow \bar{V}$, $G_1(V) \to \frac{1}{r}\left[\frac{m(V)s(V)+w(V)}{1+m(V)} - e\right] \equiv G_2(V)$. Therefore an equilibrium, by construction, corresponds to a fixed point $\hat{V}$ of

$$G(V) = \begin{cases} G_1(V), V > \bar{V} \\ G_2(V), V \leq \bar{V} \end{cases}, \qquad (27)$$

provided $L(\hat{V}) \leq N$. We will now prove that (27) has a fixed-point that satisfies this property.

First observe that $G(V)$ is continuous. Next, we show that $G(V)$ is bounded by showing that both of its components are. To start with, since G_2 is continuous on the compact domain $[0, \bar{V}]$, it is bounded above. Next, write G_1 as

$$\frac{1}{r} \frac{(r + b)z(N - (1 + m(V))L(V)) + b(1 + m(V))L(V)\left[\frac{m(V)s(V)+w(V)}{1+m(V)} - e\right]}{(r + b)(N - (1 + m(V))L(V)) + b(1 + m(V))L(V)}.$$

Because $N > (1 + m(V))L(V) \geq 0$ on $V \in (\bar{V}, \infty)$, the denominator is bounded by bNr and $(r + b)Nr$; the numerator is bounded below by 0. Also because

$(1+m(V))L(V)\left[\frac{m(V)s(V)+w(V)}{1+m(V)}-e\right] < L(V)[m(V)s(V)+w(V)] < AF(L(V)) <$ $AF(N)$ (the second inequality because maximized profit is always nonnegative), the numerator is bounded above by $(r+b)zN+bAF(N)$. Thus G_1 is bounded above and below, proving that $G(V)$ is a bounded function.

Now consider the continuous, bounded function $H(V) = G(V) + \max\left\{\min\left\{1, \frac{(1+m(V))L(V)}{N}-1\right\}, 0\right\}$. Since H continuously maps a compact domain onto itself, it has a fixed point $\hat{V}$. We claim that this is an equilibrium.

If $(1+m(\hat{V}))L(\hat{V}) \leq N$, $\hat{V}$ is also a fixed point of G and is an equilibrium by construction. To complete the proof we need to show that $(1+m(\hat{V}))L(\hat{V}) \leq N$. Suppose $(1+m(\hat{V}))L(\hat{V}) > N$: then $\hat{V} < \bar{V}$ and so $\hat{V} = H(\hat{V}) > G_2(\hat{V}) = \frac{1}{r}\left[\frac{m(\hat{V})s(\hat{V})+w(\hat{V})}{1+m(\hat{V})}-e\right]$. But adding the two participation constraints (13) together gives $\hat{V} \leq \frac{1}{r}\left[\frac{m(\hat{V})s(\hat{V})+w(\hat{V})}{1+m(\hat{V})}-e\right]$, which is a contradiction. This establishes the claim. ∎

Proof of Proposition 4: Let $r\hat{V}_U + e = v$. For full employment we have that:

$$AF'\left(\frac{N}{1+m(v)}\right) = v + m(v)(v+a(v)) \tag{28}$$

where $a(v)$ and $m(v)$ solve (12). $N/(1+m(v))$ is the number of production workers that need to be employed when the monitoring level is given by $m(v)$ in order to ensure full-employment. It is straightforward to see that, since p and q are concave, $a(v)$ and $m(v)$ are decreasing functions of v.

Next note that the firm is actually choosing s and a subject to the constraint that $s \geq v$, thus $s = v$ if and only if $\frac{\partial\Pi(s=v)}{\partial s} < 0$. Thus, we have $1 + a'(v) > 0$. By the same argument regarding the choice of w and m, we have $1 + m'(v)(v+a) > 0$. Therefore, the right-hand side of (28) is increasing in v. In contrast, the left-hand side is decreasing in v, since $m'(v) < 0$ and $F'' < 0$. Therefore, a full employment equilibrium, when it exists, is uniquely defined. Now an increase in A raises the left-hand side, thus requires an increase in the right-hand side, hence an increase in v. $\frac{dw}{dA} > 0$, $\frac{ds}{dA} > 0$, $\frac{dm}{dA} < 0$ and $\frac{da}{dA}$ immediately follow from $\frac{dv}{dA} > 0$. ∎

35

Proof of Proposition 5: UE is characterized by:

$$AF'(L^*) - w^* - m^*(s^* + a^*) = 0 \qquad (29)$$

$$-\frac{(r+b)}{(r+b)\alpha + 1}\frac{e}{Q(m^*)^2}Q'(m^*) + s^* = 0$$

$$-\frac{(r+b)}{(r+b)\alpha + 1}\frac{e}{P(a^*)^2}P'(a^*) + 1 = 0$$

and also (12) and (8). It is then straightforward to see that a^* is fixed, but all other variables vary with $V_U \equiv V$, thus we have $s^*(V)$, $m^*(V)$ and $w^*(V)$ with s^* and w^* as increasing functions of V and m^* as a decreasing function of V. Then substituting into (8), we obtain:

$$V = G(V) \equiv \frac{(r+b)z + x(V)\left[\frac{m^*(V)s^*(V)+w^*(V)}{1+m^*(V)} - e\right]}{(r+b)r + bx(V)} \qquad (30)$$

where $x(V) = \frac{b(1+m^*(V))L^*(V)}{N-(1+m^*(V))L^*(V)}$. Since the right-hand side of (30), $G(V)$, is a non-linear function, we cannot establish uniqueness of Unemployment Equilibrium. But it is clear that $G(0) > 0$ and also $\lim_{V\to\infty}G(V) < \infty$. Thus, the extremal equilibria always have $G(V)$ cutting the 45^0 line from above. Next, note that an increase in A for given V only affects L^*, thus x. In particular, x increases when A goes up. Hence, a higher A shifts $G(V)$ up, therefore, at extremal equilibria: $\frac{dV}{dA} > 0$. This immediately implies that at extremal equilibria, $\frac{dw}{dA}, \frac{ds}{dA} > 0$ and $\frac{ds}{dA} < 0$, but a remains at a^*. ∎

References

[1] Appelbaum Eileen and Rosemary Batt (1994); *The New American Workplace* ILR Press, New York.

[2] Berman, Eli, John Bound and Zvi Griliches (1994); "Changes in the Demand for Skilled Labor Within U.S. Manufacturing Industries" *Quarterly Journal of Economics*, Vol 109, 367-398.

[3] Bulow, Jeremy and Lawrence Summers (1986); "A Theory of Dual Labor Markets with Application to Industrial Policy, Discrimination and Keynesian Unemployment" *Journal of Labor Economics,* Vol 4, pp376-415.

[4] Calvo Guillermo, and Stanislaw Wellisz (1979), "Hierarchy, Ability and Income Distribution," *Journal of Political Economy,* Vol. 87, 991-1010.

[5] Chandler, Alfred E. (1977), *The Visible Hand: The Managerial Revolution in American Business,* Cambridge, MA: Harvard University Press.

[6] Foster, J. E. and H. Y. Wan (1984); "Involuntary Underemployment as a Principal-Agent Equilibrium" *American Economic Review,* Vol 74, pp 476-484.

[7] Freeman, Richard (1995); "Are Your Wages Set in Beijing?" *Journal of Economic Perspectives* Vol 9, 15-32.

[8] Freeman, Richard and Edward Lazear (1994); "An Economic Analysis of Works Councils" in Joel Rogers and Wolfgang Streeck, editors, *Works Councils: Consultation, Representation and Cooperation in Industrial Relations,* Chicago,University of Chicago Press.

[9] Gordon, David (1996); *Fat and Mean: The Corporate Squeeze of Working Americans and the Myth of Corporate Downsizing,* Martin Kessler Books, New York.

[10] Holmstrom, Bengt (1979), "Moral hazard and Observability" *Bell Journal of Economics,* Vol. 13, 324-340.

[11] Houseman, Susan (1995);"Job Growth and the Quality of Jobs in the U.S. Economy" *Labour,* pp S93-S124.

[12] Jovanovic, Boyan (1993), "The Diversification of Production," *Brookings Papers on Economic Activity: Microeconomics.*

[13] Katz, Lawrence (1987);"Efficiency Wage Theories: A Partial Evaluation" *NBER Macroeconomic Annual,* pp 235-176.

[14] Katz Lawrence and Kevin Murphy (1992);"Changes in Relative Wages: Supply and Demand Factors" *Quarterly Journal of Economics.,* Vol 107, pp35-78.

[15] Katz Lawrence, Gary W. Loveman and David G. Blanchflower (1995);"A Comparison of Changes in the Structure of Wages in Four OECD Countries" in Richard Freeman and Lawrence Katz (editors *Differences and Changes in Wage Structures* National Bureau of Economic Research, University of Chicago Press.

[16] Krueger, Alan (1993);"How Computers Have Changed the Wage Structure: Evidence From Microdata" *Quarterly Journal of Economics,* pp33-60

[17] Lazear, Edward (1995); *Personnel Economics,* MIT Press, Cambridge.

[18] Legros, Patrick and Andrew F. Newman (1996); "Wealth Effects, Distribution and the Theory of Organization," *Journal of Economic Theory,*

[19] OECD (1994); *OECD Job Study; Evidence and Explanations,*Volumes 1 and 2

[20] Shapiro, Carl and Joseph Stiglitz (1984); "Unemployment as a Worker Discipline Device" *American Economic Review,* Vol 74, pp 433-444.

[21] Weiss, Andrew (1990); *Efficiency Wages: Models of Unemployment, Layoffs and Wage Dispersion,* Princeton University Press, Princeton.