Contracts, Wage Differentials and Involuntary Unemployment

Debraj Ray
Indian Statistical Institute, New Delhi.

July 1989

Note: This article will be published in a volume of essays in memory of Professor Nabendu Sen.

Postscript (2021): Alas, it never was. For some reason, the edited volume was not published, and I never sent the paper anywhere else.

1 Introduction

One of the many aspects of Nabendubabu's distinguished career as a Professor in Presidency College was his ability to communicate to his students a sense of rigour in economic analysis. He did this in economic *history*, a topic which appears to defy the most rigorous of analytical assaults on it. In this modest contribution dedicated to Nabendubabu's memory, I want to try and emulate his approach by "speaking" to the serious student of economics in rigorous yet elementary terms. My topic, however, falls in the far more tractable realm of pure economic theory. It addresses the issue of involuntary unemployment and wage differentials *in equilibrium*, from a point of view that may have several implications.

1.1 This paper and its results

I am going to study a simple model of the labour market, with one major difference from the textbook competitive framework. The basic postulate is that a labourer will *not* supply effort unless there are adequate incentives for him to do so. The firm must therefore offer a contract that *induces* each labourer to put in the specified level of effort.

One approach to this incentive problem is taken in the standard principal agent model studied by Mirrlees [1975, 1976], Holmstrom [1977], Grossman and Hart [1983] and many others. In this approach, a worker's income *depends* on the output produced. A stake in the firm's output induces worker effort to some extent.

This is an important contractual form, but it isn't the one I consider to be of fundamental interest in the present context. When a potentially large number of workers combine to produce a single output (as in a firm), it is difficult to provide adequate output-based incentives to a sizeable fraction of them. For an output-based incentive scheme to have "power", an additional unit of worker effort must be significantly related to an increase in worker income. But with a large number of workers, this cannot be done all around, unless workers can be given large *negative* payments for some realizations of output, or unless unrealistic "forcing contracts", which are not robust to the introduction of production uncertainty, are used. For further discussion, see Section 4.2.3. of this paper.

An alternative contractual form, and perhaps more realistic in the context of multi-worker firms, is *direct* supervision of a worker's effort, coupled with (a) a renewal of the contract should the worker conform to a pre-specified work standard, and (b) eviction or firing should the worker be found shirking relative to this standard. It follows right away that if there is to be any work incentive at all, *a worker's utility conditional on obtaining the contract must strictly exceed his utility conditional on being fired.*

Suppose, just for the moment, that all firms are identical. Then, in equilibrium, all firms offer the same contractual utility. It follows that the *only* situation in which this utility can exceed the utility conditional on being fired, is one where fired workers cannot find *certain* reemployment. We conclude that an equilibrium of this model *must* involve unemployment. Moreover, this unemployment must be *involuntary*, in the sense that unemployed workers *strictly* prefer to work, but cannot find a job!

This intuition is at once simple and robust. It survives numerous extensions and generalizations. In the particular model that I consider, I will permit firms to be heterogeneous, the heterogeneity arising from capital ownership of various sizes. My objective is to study not only the phenomenon of involuntary unemployment, but also the contractual terms offered by these different firms.

It turns out that an equilibrium of this model involves not only unequal utility treatment of employed and unemployed, but also unequal treatment of identical, *employed* workers across different firms. These differences spring directly from the nature of the "supervision" technology that I postulate. Specifically, I assume that apart from the fixed costs of having supervision at all, "it is more than twice as costly to supervise two workers than to supervise one". There is a simple reason for this. A worker's effort is inferred, not always through direct observation of his activity (which is often impossible), but from a number of observed "signals" that are closely correlated with his true effort. Possibly the most important signal in this regard is the final output produced. Now, with two workers *jointly* producing a *single* output, the information content of the output signal regarding the *separate* effort level of *each* individual is significantly reduced. Consequently, to achieve the same level of "supervisory accuracy" as in the one-worker case, *per worker* supervision costs must go up.

In formal terms, this assumption is expressed in the postulate that the total variable cost function of supervision is *strictly convex* in the workforce employed by the firm. For further discussion, see Section 4.2.1.

This postulate yields the following empirically testable propositions. In the same industry:

1. Larger firms pay higher wages.

2. Larger firms demand higher individual effort levels.

3. Despite the opposing pulls of items 1 and 2 on the worker, larger firms offer a higher net contractual utility.

4. On the other hand, larger firms are more capital intensive,[1] even if production functions are homothetic in capital and labour.[2]

[Above, the word "larger" describes firms with a higher capital stock.]

I also consider, informally, a number of extensions of the basic model. These yield further results that are also amenable to empirical investigation. Finally, I briefly consider some possible normative implications of the theory.

1.2 What to Read

All these results stem from the one conceptual premise, discussed above, that makes our model different from the textbook versions. This premise is novel but not new. Indeed, apart from the study of intra-occupational contract differentials, very little of what I do in this paper is original. I have already referred you to some literature on agency theory. In the present context, Calvo [1979, 1981] was possibly the first to conduct a rigorous study of incentive contracts that involve firing. You can also consult Salop [1979], and, more recently, Shapiro and Stiglitz [1984], who explicitly discuss the implications of similar models for the existence of involuntary unemployment.In the context of agricultural production, see Eswaran and Kotwal [1985]. For an integration of the theory of output-based incentive contracts and the contracts discussed here, Singh [1982], Dutta, Ray and Sengupta [1989] and the references cited in the latter paper.

[1]By capital intensity, we mean the ratio of capital to *total* effort, that is, the individual work standard multiplied by toal employment. In the light of item 2, this means that if capital intensity is defined by the capital-workforce ratio, the capital deepening effect is even more pronounced.

[2]In our model, we assume that the production function displays constant returns to scale, so that homotheticity is automatically implied.

Very broadly, the model studied here falls under the purview of efficiency wage models. Apart from the nutrition-based efficiency wage models studied by Leibenstein [1957], Bliss and Stern [1978], Dasgupta and Ray [1986] and others, there is a generalized view of wages affecting effort in a variety of ways. For a discussion of this viewpoint, see Yellen [1984] and Akerlof and Yellen [1986].

2 The Model

My objective is to construct the simplest model of a labour market that will allow me to convey the ideas discussed in the introduction. Suppose that there is a single, homogeneous product, produced using capital and labour. Firms hire labour. Their *capital* endowments are given, and I shall use this as the major variable distinguishing one firm from another.[3] A detailed specification follows.

2.1 Labourers

There are a total of $\bar{N}$ labourers. Each labourer is taken to be sufficiently small compared to the aggregate, so that I can consider their number as a continuous variable. Each labourer derives utility out of income and disutility out of work effort.[4] The simplest way of capturing this is to posit that if w is a labourer's income and x is the effort that he puts in, then his net utility is

$$w - v(x)$$

We assume

(A.1) v is an increasing, twice differentiable function defined on some domain $[0, B)$ (where $0 < B \leq \infty$), with $v(x) > 0$ and $v''(x) > 0$ for $x > 0$.[5]

[3] All the results go through even in the presence of a capital market. But then it is necessary to introduce alternative sources of heterogeneity, such as differences in productivity.

[4] Of course, the assumption of work disutility at all levels of work effort is an exaggeration and not meant to cast aspersions on the work ethics of people! The idea is simply that labour must be provided incentives to supply effort over and above some minimum, which is normalized here to zero.

[5] The assumption on v'' reflects the fact that marginal disutility of effort increases with effort. By the way, if you like to be general and wish to describe the worker's utility

Moreover, $\limsup_{x \to B} v'(x)/v''(x)x < \infty$.[6]

If a labourer is unemployed in a particular period, he obtains some utility (perhaps from self-employment) which we shall denote by r. This is his *reservation utility*, and it is exogenously given. We assume that $r > 0$.

We need a way of comparing utilities received by a labourer in different periods of *time*. This we do by simple discounting: there is a *discount factor* $\delta \in (0,1)$ such that if $\langle u_t \rangle$ is a sequence of one-period utilities received by the labourer, his total utility viewed at time zero is

$$\sum_{t=0}^{\infty} \delta^t u_t$$

2.2 Firms and Technologies

Firms produce a single commodity, which we shall use as the numeraire. The inputs are capital (K) and labour effort (L). Firms are distinguished only by their capital holdings, and we shall take capital ownership as a proxy for firm size. Denote by $F(K, L)$ the production function for output. We assume that

(A.2) F is constant returns to scale, increasing, strictly concave and twice differentiable.[7]

Our main conceptual departure from the standard model lies in the postulate that workers must be given appropriate incentives to work. In the classical principal-agent model, this is achieved, to some extent, by conditioning worker income on the output produced. With many workers producing a single output, however, this type of contract is not very powerful as an incentive device (more on this below). Here, we study a different incentive structure: the postulate of a *work standard* by each firm, coupled with supervision of workers, and the threat to fire any worker who is found not performing up to standard.

---

function as $u(w, x)$ (with the proper assumptions on derivatives, of course), you can redo the analysis this way.

[6] This is a technical assumption which is used to guarantee the existence of equilibrium. It is satisfied by many functional forms, e.g, $v(x) = hx^{\beta}$ or $v(x) = \frac{x}{1-x}$ for $x \in [0, 1)$.

[7] We also assume the familiar endpoint (Inada) conditions on each input, only to simplify the analysis.

Supervision is costly. For reasons discussed in the introduction, I am going to assume that the total variable cost of supervision increases "more than proportionately" with the workforce of a firm. Let $C(n)$ denote this cost function. I assume

(A.3) $C(n)$ is increasing, twice differentiable, with $C(0) = 0$ and $C''(n) > 0$.

In what follows, I shall not explicitly consider the fixed cost of supervision. The fixed cost is only needed to examine whether firms will produce or shut down, and this minor extension can be easily accomodated.

2.3 How Labourers Respond to a Contract

Denote by U the expected lifetime utility available to a *currently* unemployed worker. U is, of course, a discounted sum of per period expected utilities (which include the possibility of remaining unemployed), and it is ultimately an *endogenous* variable whose determination we discuss below (Sections 2.5 and 2.6).

A *contract* is a pair (w, x), where w denotes income and x is the work standard that the worker is expected to uphold in the job. If the worker demonstrably fails to meet the standard, he is fired. Otherwise, he is retained on the same terms. In addition, I postulate an exogenous probability $q > 0$ that the worker may quit or is removed from his job because of factors not explicitly modelled here. This assumption is merely a shorthand to guarantee some turnover in equilibrium. In the model, it is also assumed that the supervision technology is accurate, in the sense that shirkers are detected with probability one. This simplifies the analysis, because it effectively leaves a worker with two choices: to exactly uphold the work standard, or to shirk by setting effort level equal to zero. This latter assumption is dropped in the informal extensions discussed in Section 4.2.1.

If the worker shirks, his lifetime utility conditional on being offered the contract (w, x) today is

$$U' \equiv w + \delta U \tag{1}$$

The first term on the RHS of (1) gives the worker's current utility w, derived from no effort and the contract income of w. But he is fired for sure, and then from the *next* period onwards, he receives lifetime utility U, which, of course, is discounted by δ.

The nonshirker, on the other hand hand, enjoys an expected utility of

$$U'' \equiv [w - v(x)] + \delta(1-q)\max(U', U'') + \delta q U \tag{2}$$

The RHS of (2) has three sets of terms. The first set gives the worker's current utility, taking into account his disutility from maintaining the specified work standard x. As a nonshirker, he retains his job with probability $(1-q)$, and then continues to enjoy utility equal to the maximum of shirk/nonshirk utilities. This is discounted by δ, and weighted by its probability of occurence $(1-q)$. This explains the second term. The final term yields his expected discounted utility in the exogenous event (with probability q) that he loses or gives up his job.

The worker will not shirk if and only if $U'' \geq U'$, or equivalently if[8]

$$v(x) \leq \delta(1-q)\left[\frac{w - v(x) + \delta q U}{1 - \delta(1-q)} - U\right] \tag{3}$$

You can easily calculate (3) using (1) and (2). The RHS of (3) provides the incentive in terms of lifetime utility differences conditional on renewal and expulsion. This must outweigh the disutility of conforming to the specified work standard $(v(x))$, and this explains the inequality in (3).

2.4 How Firms Design a Contract

Each firm must design its contract keeping an eye on the constraint (3). If they do not respect (3), their employees will prefer to shirk.[9] It may help, in what follows, to retain an analogue with the standard model. There, firms take the going *wage* as given. Here, firms take the going *utility* (U) as given. Now what does this mean? The idea, simply, is this: our model will not generate a *single* equilibrium wage for labourers (as you have already

[8]We adopt the harmless convention that if a worker is indifferent between shirking and not shirking, then he will not shirk.

[9]It may be argued that firms do not know the form of the disutility function $v(x)$, so they cannot be sure of observing (3) even if they want to. This is a valid objection, but it is one that takes the formal model too literally. The model is an approximation to the more realistic scenario where firms have an imprecise notion of the form of the incentive constraint. After all, to argue that firms have *no* knowledge of it is surely far more unrealistic!

guessed from the introduction), and not even a single work standard. The "going" utility U of a currently unemployed worker is a mix of his per period reservation utility (r), and the utilities from all the contracts available on the market (conditioned on the probability of obtaining these contracts). *This* utility represents the implicit threat which a firm can impose on a worker by firing him! If the terms and conditions of a firm's offer represents the "carrot", then U represents the "stick".

Now consider a firm's contract design problem. A firm with given capital stock K seeks to maximize its profits by choosing an individual work standard (x), a wage (w), and a workforce (n). To make sure that work standards are upheld, the firm must incur a supervision cost of $C(n)$ and respect the constraint (3) in its choice of the contract. Formally, the firm solves

$$\max_{x,w,n\geq 0} F(K,nx) - wn - C(n) \tag{4}$$

subject to the constraint (3).

Presently, we shall study in detail the characteristics of the solution to this problem (see Section 4.1 and Appendix 1). For now, take it on faith that for each value of the goiing utility U and the firm's capital stock K, there exists a unique solution to (4) which we shall denote by $\{x(K,U), w(K,U), n(K,U)\}$. Observe that in the solution, workers *will* abide by the work standard $x(K,U)$ because (3) has been respected in the contract design. Hired workers will receive an expected lifetime utility from this contract which we denote by $\mu(K,U)$; clearly, this solves

$$\mu(K,U) = w(K,U) - v(x(K,U)) + \delta(1-q)\mu(K,U) + \delta q U$$

so that

$$\mu(K,U) = \frac{w(K,U) - v(x(K,U)) + \delta q U}{1-\delta(1-q)} \tag{5}$$

2.5 The Unemployed Pool

Suppose that there are m firms indexed by $j = 1, ..., m$. Suppose that firm j is offering a contract which is perceived to have expected lifetime utility U_j, and employs n_j workers. Define $N \equiv \sum_{j=1}^{m} n_j$, and suppose that $N \leq \bar{N}$. In preparation for defining the general equilibrium of the system, we shall now determine the expected lifetime utility of a worker conditional on being currently unemployed.

At any date, the pool of unemployed workers has two components:

(1) People who have become unemployed at the start of the period, by virtue of the exogenous quit rate, or because they did not conform to work standards in the firm that employed them. In our simplified model, the latter will not occur because the design of contracts respects (3), and supervision is by assumption accurate. Consequently, the total number of people in this component is qN.

(2) People continuing their state of unemployment from the previous period. They do so either through choice (i.e. by refusing an offer) or by necessity (because they have not received an offer). Call this number b. Clearly, $b = \bar{N} - N$. By our supposition, $b \geq 0$.

The total number of *vacancies* at any date is given by qN. Therefore, the probability that a currently unemployed person will obtain a job[10] is given by

$$\pi \equiv \frac{qN}{qN + b} = \frac{qN}{qN + \bar{N} - N}$$

Consider, now, a currently unemployed person. We follow a standard search-theoretic argument to determine his expected lifetime utility. With probability π this person receives an offer. Conditional on some offer being received, it will be an offer from firm j with probability n_j/N, promising a lifetime utility of U_j. The worker must use an optimal decision rule to assess those offers worthy of acceptance.[11] Denote by U the *maximum expected lifetime utility* under the optimal decision rule. The following is true:

Fact: *U is given by the unique solution to the equation*

$$\begin{aligned} U &= \zeta(U_1, \dots, U_m; n_1, \dots, n_m) \\ &\equiv \frac{\left[1 - \frac{\pi}{N}\sum_{j:U_j \geq r+\delta U} n_j\right] r + \frac{\pi}{N}\sum_{j:U_j \geq r+\delta U} n_j U_j}{1 - \delta\left[1 - \frac{\pi}{N}\sum_{j:U_j \geq r+\delta U} n_j\right]} \end{aligned} \tag{6}$$

The expression (6) follows from a dynamic programming argument, and is established rigorously in Appendix 1. It is valid for all situations with

[10] In our model, we do not allow on-the-job search.

[11] This is the basic stopping rule problem which crops up in search theory and related areas (see, e.g., McCall [1970]), and which we need to solve here.

$N \leq \bar{N}$. The expression looks complicated but is actually quite simple. *Given* that U is the maximum lifetime utility, the worker can guarantee himself $r + \delta U$ by staying unemployed today and following the optimal search strategy from the next period. Consequently, the worker will accept an offer today from firm j if and only if $U_j \geq r + \delta U$.[12] Indeed, this *is* the optimal strategy. It is now easy to see that the expected optimal lifetime utility from following this strategy is precisely the RHS of (6), and this value is U. So certainly, the optimal U satisfies (6). A little additional work shows that (6) cannot have any other solution, and so *characterizes* the optimum.

2.6 Equilibrium

We are now ready to define an equilibrium for the overall economic system. An *equilibrium* is a going utility and a collection of firm contracts, that is, a collection $\left[U^*, \{x_j^*, w_j^*, n_j^*, U_j^*\}_{j=1,\dots,m}\right]$ such that the following three sets of conditions hold

(i) *Aggregate Feasibility*:

$$\sum_{j=1}^{m} n_j^* \leq \bar{N}$$

(ii) *Firm Consistency*: For all $j = 1, \dots, m$,

$$w_j^* = w(K_j, U^*), \qquad x_j^* = x(K_j, U^*)$$

$$n_j^* = n(K_j, U^*), \qquad U_j^* = \mu(K_j, U^*)$$

(iii) *Utility Consistency*:

$$U^* = \zeta\left(U_1^*, \dots, U_m^*; n_1^*, \dots, n_m^*\right)$$

The first condition states that the equilibrium aggregate demand for labour should not exceed the available supply. The second condition states that in equilibrium, the contracts offered by each firm, and their utility values, must be consistent with profit-maximizing behaviour given the going utility U^*. The final condition states that the going utility *itself* must be consistent with the various contracts available in the system, as discussed

[12] Again, we use the harmless convention that indifference is resolved by acceptance.

in Section 2.5. Note that the going utility has a well-defined expression, given (i).

Our overall model should be consistent. It is. This is expressed in

Proposition 1 *An equilibrium exists.*

The reader interested in a rigorous proof of this proposition is referred to Appendix 3. It should be noted that our assumptions do *not* appear to guarantee the uniqueness of equilibrium.

We are now in a position to examine the qualitative features of the model.

3 Involuntary Unemployment is an Equilibrium Outcome

In this section, we show that involuntary unemployment is *always* an equilibrium outcome.

3.1 Unemployment is an Equilibrium Outcome

We start by observing that the stock of unemployed people, *not counting the flows in and out due to labour turnover at the rate* q, is always positive in equilibrium.

Proposition 2 *In an equilibrium,* $b^* \equiv \bar{N} - \sum_{j=1}^{m} n_j^* > 0$.

Proposition 2 will be proved along with Proposition 3 below. An examination of the argument yields an intuition that is simple yet striking. If there were full employment in equilibrium, the going utility would be expressible as a convex combination of the various utilities available in the market, *with per period "unemployment utility" r getting zero weight.* This means that there exists *some* firm which is hiring a positive workforce and is offering a contractual utility which does not exceed the going utility. But then it is obvious that a positive work standard cannot be upheld in this firm, which contradicts the hypothesis of profit maximization.

A feature of our model, then, is that it *always* predicts unemployment as an equilibrium outcome. This is in startling contrast to perhaps *every* other known model of involuntary unemployment. But do not be won over by (or equally, too suspicious of) this general result. It arises, in part, from our assumption that *every* productive activity in the model requires worker supervision at a cost. This may not be the case in reality. One can imagine a number of activities, such as self-employment, or production using labour with outputs attributable to each individual, where supervision costs are either absent or do not appreciably change with scale. However, you can check that even in this case, unemployment must continue to be an equilibrium outcome as long as labour turnover from the "costly supervision" firms is positive. If in addition, though, we allow for the possibility that persons can search for jobs while engaged in self employment or in activities requiring low supervision, the model becomes compatible with full employment. On the other hand, the feature of unequal treatment of identical workers persists. I remark further on this in Section 3.2.

3.2 All Unemployment is Involuntary

An unemployed person is *involuntarily unemployed* if he strictly prefers to work in one or more of the available jobs, *and* if these jobs employ individuals who are in all respects identical to him. Involuntary unemployment, in its broadest sense, deals with unequal labour-market treatment of *ex-ante* identical individuals.[13] It turns out that not only do we have unemployment as an outcome, but also that this unemployment *must* be involuntary. This is rephrased in

Proposition 3 *In any equilibrium,* $\min_{j:n_j^*>0} U_j^* > U*$*, or, all unemployment is necessarily involuntary.*

Proofs of Propositions 2 and 3: First, we claim that for all j with $n_j^* > 0$, we have $U_j^* > U^*$. Suppose not. Then for some such j, we have $U_j^* \leq U^*$. Now use the definition of U_j^*, the constraint (3), and our assumption on $v(x)$ to observe that $x_j^* = 0$ for such a firm. Moreover,

[13]This notion is, of course, not applicable in practice, given that no two people are ever exactly identical. This definition, however, can be extended: unequal treatment can be taken to refer to a discontinuity in the utility schedule as a function of individual characteristics (as in Dasgupta and Ray [1986]). For our model, however, the simpler definition will suffice.

$n_j^* > 0$, and because $r > 0$, it must be that $w_j^* > 0$. But this contradicts profit maximization for firm j, *and so proves Proposition 3.*

Now suppose that $b^* = 0$. Then $\pi^* \equiv \frac{qN^*}{qN^*+b^*} = 1$. Using this, the fact that $\min_{j;n_j^*>0} U_j^* > U^*$, and (6), we see that

$$U^* = \frac{1}{N^*}\sum_{j=1}^{m} n_j^* U_j^*$$

Because $\sum_{j=1}^{m} n_j^* = N^*$, this means that $U_j^* \leq U^*$ for some j with $n_j^* > 0$. But this contradicts Proposition 3. *Proposition 2 is therefore established*, and we are done. ∎

The intuitive basis of this result is similar to that of Proposition 2. For a firm to offer adequate incentives to its workers, it must be the case that the holding of a job confers *strictly* higher utility than the state of being unemployed. Be careful, though: from this observation, no deduction should be drawn regarding the social desirability of unemployment as a means of creating work incentives. It is a *positive* description of the equilibrium outcome of a profit-maximizing, competitive economy. And, I might add, it is a description that is sharply at variance with the essentially harmonious view of competition as envisaged by Adam Smith and his followers. See, however, Section 4.2.4 for some remarks on possible normative implications of the theory.

A further point deserves scrutiny. The analysis so far suggests that the very concept of unemployment itself is somewhat nebulous. [Of course, there is the entire issue of vagueness in *measurement*, but that's not what I mean here.] To see this, suppose in fact that the reservation utility of our model is generated, not by unemployment, but by some productive activity which is not costly to supervise, such as self-employment.[14] Now, in equilibrium, our model will still generate employment for a section of workers, with contractual utility strictly higher than that of the self-employed (the "unemployed", in the language of the model). In what sense is the basic situation unaltered by this reinterpretation of "unemployed" as "self-employed"?[15] What *does* appear to be fundamental is that there are utility

[14]These remarks are also meant to tie in with the discussion at the end of Section 3.1.

[15]In reply, one might argue that the very fact of being employed confers additional utility, and this is a genuine difference. But the argument falls apart if the whole model is reworked with this extra utility explicitly incorporated into the incentive constraints. The same qualitative structure will reemerge.

differentials among *identical* individuals that persist in equilibrium. Among these, the state we call "unemployment" involves the lowest possible utility values. These observations will be taken up further in Section 4, which goes into the question of equilibrium utility differentials.

I end this section by recalling the "classical" objection to any notion of equilibrium involving involuntary unemployment. Here it is: "An equilibrium cannot involve involuntary unemployment. If it did, we'd have a contradiction, because unemployed workers can (and will) *undercut* the employed by offering to work at a lower wage. As they are identical to their employed counterparts, firms should hire them. This means we never had an equilibrium to start with!"

Do you see why this argument breaks down in our model?

4 Contract Differentials and Firm Size

I have already observed that, in general, there will be wage and utility differentials *in equilibrium* among workers who are identical to start with. The purpose of this section is to demonstrate that such differentials are related in a systematic way to the size of the firm, and to outline a number of extensions to the basic model in this context.

4.1 Differentials

By a *larger firm* I shall mean a firm with a larger endowment of capital. A larger firm naturally wishes to engage a larger workforce, given that labour is complementary to capital. At this point, the technology of supervision becomes a critical factor. Recall our assumption that the cost of supervision is convex in the labour force. This convexity forces the larger firm to take steps other than a straightforward expansion of the workforce. Specifically, we have

Proposition 4 *In equilibrium, a larger firm (i) sets higher work standards, (ii) offers a higher wage, (iii) offers a contract with a higher utility value, (iv) hires a larger workforce, and (v)is more capital intensive in the sense of having a higher value of K/nx.*

In particular, a worker will strictly prefer to work in the largest, most capital-intensive firms.

The proof of Proposition 4 is given in Appendix 2. The main idea is straightforward. Supervision is costly, and *proportionately* costlier as a firm expands. There is, therefore, a tendency to *not* expand employment in proportion to capital ownership, even though employment does expand in absolute terms (part (iv)). To compensate for this in some measure, the work standard for each labourer is raised (part (i)). Despite these two factors which both act to increase total labour effort, it is of some interest that the *net* effect is always an increase in the capital-to-*total* effort ratio (part (v)).

The rest of the details follow without much fuss. Once work standards are raised, a larger differential between contractual utility and the going utility is needed to satisfy the worker's incentive constraint. This explains (iii). Of course, as a result of this, the wage offered must be higher, which explains (ii).

These observations might explain, at least partly, why larger and more complex organizations tend to pay higher wages for similar jobs, and why individuals prefer these jobs. Such organizations also appear to demand (and obtain!) higher work standards. Of course, the presence of these differentials is an empirical question. I hope the theory is provocative enough to induce such empirical analysis.

It must be added that this theory seeks to explain differentials *not* by taking recourse to assertions that higher-paying firms and their workers are intrinsically more "efficient". This may well be the case. However, what I am arguing is that there may be a more fundamental reason why the "law of one price" may not hold in the case of labour. The reason stems from the necessity to supervise labour effort, and from the fact that supervision is costly. Larger firms react to this by cutting employment (in relative terms), and by demanding greater effort from each employee. The price paid for this demand is a higher wage; indeed, a higher worker utility. After all, this is what gives the threat of firing its credibility.

4.2 Discussion and Possible Extensions

4.2.1 Size, complexity and hierarchies

The nature of the supervision technology is related to the "complexity" of the production activity, where this term refers to the degree of interpersonal interaction required to produce a single observable output. I have proxied this by firm *size*, and I do not believe it is a bad proxy. Firms employing more capital and labour *in the same industry* are likely to be more complex: if there is some deficiency in aggregate performance, it is that much more difficult to hold particular individuals responsible. Consequently, the precise monitoring of an individual's performance becomes a costly activity.

This discussion, and the model in general, throws some light on the question of *intra*-firm organization. These issues have been explored by Calvo and Wellisz [1981], but there is plenty of room for further research. Hierarchy implies "responsibility in layers". After a certain critical level of the workforce is reached, it may be profitable for the firm to hire a divisional manager *who takes responsibility* for a clearly identified subsection of the workforce. This manager is given a contract, just as the workers are, and is fired if something goes wrong in the activity of that subsection. As in Calvo and Wellisz (1981), one can construct realistic models to show that managers will enjoy a higher net utility than the workers he takes charge of, *even though there may be no intrinsic differences in ability.* With a larger firm, it may be profitable to hire a manager to take charge of the divisional managers, and so on.

If one takes into account these additional features and reconstructs the general equilibrium model of the preceding sections, there will be a significant enrichment of (and — no doubt! — significant differences in) the results. This is as it should be. The power of this approach lies not in the detail of individual results, but in its basic conceptual postulate and some of its remarkable implications.

4.2.2 Probabilistic supervision and uncertain detection

In my model I have assumed — quite unrealistically — that supervision is carried out in every period, *and* that shirkers are detected with probability one. Both assumptions can be dropped to achieve extensions of some interest. Consider, first, the new decision problem of the firm in a model where the first assumption is dropped. The firm chooses x, w, n as before, *and* p, the probability of supervision, to solve

$$\max_{(x,w,n)\geq 0, p\in(0,1)} F(K, nx) - wn - pC(n)$$

subject to a constraint analogous to (3).[16]

In this extension, supervision is not carried out in every period. There is an explicit *threat* of supervision (with independent probability p in every period), which the firm can credibly precommit to.[17]

Now we can go ahead and define an equilibrium just as we did before. What would we find? I do not know for sure (the results must be rigorously worked out), but I would conjecture that exactly for the same reasons advanced in Section 4.1, the probability of supervision would be lower for the larger firms. These firms will use every instrument they can to escape high supervision costs. Probabilistic supervision is one such instrument. Of course, if this conjecture is borne out, it would further reinforce the utility differentials observed in Proposition 4.

Now let us drop the second assumption, which states that shirkers are detected with probability one whenever they are supervised. One way to do this is to introduce a "probability-of-detection" for the worker, $\rho(x, x')$, which depends on x, the stated work standard, and on x', the actual work effort put in. This function must be used in a modified form of the constraint (3). With this modification, the main difference is that *in equilibrium*, shirking will exist and some proportion of workers will be fired. In my simple model, this feature is absent except via the exogenous quit component q.

4.2.3 Output-based incentive contracts

In the formal analysis, I have considered only one type of labour contract: one that provides incentives by driving a wedge between the contract utility and the "going utility", and by threatening to fire nonconformists. I have already noted that the "classical" contract studied in the principal-agent literature is different. Specifically, the income payments to a worker is related to the *output* produced.[18] There is no reason why a contract in our model should be completely devoid of output-based incentives.

However, in the particular context of the model being studied here,

[16]Supervision probabilities must now be explicitly incorporated in the constraint.

[17]Whether such precommitment can indeed be credible is an important issue, but one I do not address in this paper.

[18]In the context of agriculture, this is a commonly observed contractual form, two examples being sharecropping, and the existence of piece-rate contracts in harvesting.

one observation is quite clear. With a single, jointly produced output, output-based contracts lose their power as the workforce expands. After all, it is not possible to *simultaneously* provide a large number of workers with a significant share in the output.[19] I am therefore led to the following conjecture: in tasks that are jointly performed by a large number of workers, one would expect to observe contracts that combine explicit supervision with the threat to fire workers. Output-based incentives would occupy a secondary position here. Conversely, output-based payment schemes would acquire a greater role in tasks involving a relatively small number of workers, though firing clauses would not be absent, in general.[20] The contrasting implications for the same productive activity in large versus small firms should now be quite clear.

4.2.4 Some normative remarks

The model is a positive one that purports to explain some aspects of observed reality, and implies others that may require further empirical analysis. Are there any *normative* lessons to be learnt?

There are, in fact, some normative issues of unemployment policy that are thrown up. These are addressed in Salop [1979] and Shapiro and Stiglitz [1984], and I will avoid repetition.

On the microeconomic front, there are some implications for the design of contracts. I shall motivate these by taking up an old issue: the question of efficiency differences between private and public sector firms. I do not, by the way, treat such differences as an empirically established fact, but only as a feature that appears to be valid on the basis of casual observation. These differences coexist with the observation that private firms appear to pay more than their public sector counterparts for labour that is similar. Our model has, of course, an obvious bearing on this issue: the greater flexibility of the private sector in *firing decisions* permits a higher work standard. In order to satisfy the "no-shirking constraint", private sector

[19] With this statement, I am actually glossing over the possibility of somewhat more complicated contractual schemes. I have in mind the notion of "forcing contracts", with the firm owner as residual claimant, which would work in a model without production uncertainty (see, e.g., MacDonald [1982] for a survey). However, these contracts are not at all robust to the introduction of production uncertainty.

[20] On the coexistence of both types of clauses, see Dutta, Ray and Sengupta [1989].

firms must pay more.

The inability to fire shirkers in the public sector arises partly from a normative consideration: that stability of employment "should" be guaranteed in a labour-surplus economy.[21] Is there any ethical justification for this? This is a difficult question, given that the very same stability restricts job opportunities for the *un*employed. However, let us grant for the sake of argument that, at least for low-income, unskilled jobs, employment stability should be a consideration. In that case we must be prepared to bear its inevitable consequence: that such workers will exhibit dramatically lower productivity relative to their private sector counterparts.

What about middle and higher-income employees? Here, the ethical consideration of stability must surely be weaker, if not nonexistent. Yet a comparable degree of insulation persists at these levels too. We cannot *simultaneously* bemoan the fact that there is low productivity, *and* uphold the norm of unconditional employment stability. To gain the one is to sacrifice the other, unless we are prepared to believe that public sector employees are somehow imbued with the highest degree of social consciousness.

5 Summary

In this paper I have outlined a simple model of intraoccupational wage differentials and involuntary unemployment. The analysis is based on the postulate that each worker in a firm must be supplied with an appropriate reward/punishment incentive contract in order to put in effort. This contractual structure is achieved by offering the worker a utility which strictly *exceeds* the "going" utility conditional on unemployment. The offer is backed up by rewarding nonshirkers with contract renewal, and shirkers with expulsion. It is shown that an economy based on this postulate *must* exhibit involuntary unemployment in equilibrium and will generally display inter-firm wage differentials. Specifically, firms with larger capital ownership will pay a higher wage, and set a more demanding work standard. The combination of these factors is a contract which offers higher net utility. On the other hand, the model predicts that larger firms tend to be more capital-intensive, *even if* the production function is homothetic in

[21] There may be other considerations, such as the possible abuse of the power to fire by individuals who possess such power.

capital and labour. These outcomes are consequences of the technology of supervision, and how this technology is affected by the size and complexity of the organization.

A number of extensions and some normative implications are explored.

6 Appendix 1: Determination of U

Proof of the fact in Section 2.5: Let U be the maximum expected utility from following an optimal strategy. Then, at the current date, it is optimal to accept an offer iff

$$U_j \geq r + \delta U$$

The expected utility of this strategy must be precisely U. But then,

$$U = \frac{\pi}{N} \sum_{j:U_j \geq r+\delta U} n_j U_j + \left[1 - \frac{\pi}{N} \sum_{j:U_j \geq r+\delta U} n_j\right] (r + \delta U)$$

Rearranging this yields (6). It remains to show that the solution to (6) must be unique. An equivalent form of (6) is

$$(1-\delta)U = r + \frac{\pi}{N} \sum_{j:U_j \geq r+\delta U} n_j \left[U_j - (r + \delta U)\right]$$

Suppose that this admits of two distinct solutions, U and U' with $U' > U$. We then have

$$(1-\delta)(U'-U) = \frac{\pi}{N}\left\{ \sum_{j:U_j \geq r+\delta U'} n_j \left[U_j - (r + \delta U')\right] - \sum_{j:U_j \geq r+\delta U} n_j \left[U_j - (r + \delta U)\right]\right\}$$

But if $U' > U$, the RHS of the above equation must be nonpositive, while the LHS is strictly positive, which is a contradiction. ∎

7 Appendix 2: Partial Equilibrium Analysis

Consider the problem set out in the text:

$$\max_{x,w,n \geq 0} F(K, nx) - wn - C(n)$$

subject to the constraint (3). This constraint may be equivalently rewritten as

$$\delta(1-q)\left[w-(1-\delta)U\right] \geq v(x) \tag{7}$$

Set up the Lagrangean

$$\mathcal{L} \equiv F(K, nx) - wn - C(n) + \lambda\{\delta(1-q)\left[w-(1-\delta)U\right] - v(x)\} \tag{8}$$

You should verify that the interior first-order conditions I shall now write down are *necessary* and *sufficient* for the optimal solution. You can do this by noting, first, that we will have an interior solution in (x, w, n), so that the first-order conditions do hold with equality. Next, use (17) below to check the second-order conditions.

Differentiating in the order (x, n, w, λ) yields the following first order conditions:

$$nF_L - \lambda v'(x) = 0 \tag{9}$$

$$xF_L - w - C'(n) = 0 \tag{10}$$

$$-n + \lambda\delta(1-q) = 0 \tag{11}$$

$$\delta(1-q)\{w-(1-\delta)U\} - v(x) = 0 \tag{12}$$

Using (10) and (11) to eliminate the variables w and λ from the system, we obtain

$$\delta(1-q)F_L - v'(x) = 0 \tag{13}$$

$$\delta(1-q)\{xF_L - C'(n) - (1-\delta)U\} - v(x) = 0 \tag{14}$$

We will totally differentiate (13) and (14). For our purposes, it will suffice to consider parametric changes in K and U only. We obtain, using (13) once in what follows, that

$$\left[\delta n(1-q)F_{LL} - v''(x)\right]dx + \delta(1-q)xF_{LL}dn = -\delta(1-q)F_{LK}dK \tag{15}$$

$$nxF_{LL}dx + \left[x^2F_{LL} - C''(n)\right]dn = -xF_{LK}dK + (1-\delta)dU \tag{16}$$

The determinant of the Jacobian in this linear system in (dx, dn) is

$$\mathcal{D} \equiv \begin{vmatrix} \delta(1-q)nF_{LL} - v''(x) & \delta(1-q)xF_{LL} \\ nxF_{LL} & x^2F_{LL} - C''(n) \end{vmatrix} \tag{17}$$

Using our assumptions and routine computation, it is easy to see that $\mathcal{D} > 0$.

Further, Cramer's Rule and our assumptions yield:

$$\begin{aligned}\frac{dx}{dK} &= \frac{1}{\mathcal{D}}\begin{vmatrix} -\delta(1-q)F_{LK} & \delta(1-q)xF_{LL} \\ -xF_{LK} & x^2F_{LL} - C''(n) \end{vmatrix} \\ &= \frac{\delta(1-q)F_{LK}C''(n)}{\mathcal{D}} \quad > \quad 0 \end{aligned} \tag{18}$$

[You should check that our assumption (A.1) on F guarantees $F_{LK} > 0$.]

This verifies part (i) of Proposition 4, if one treats this exercise as being carried out at the equilibrium of the system. Now inspect (13) and apply (18). We see that F_L is related positively to K. By a well-known property of linear homogeneous functions, it follows that K/nx is related positively to K. *This verifies part (v) of the proposition.* Next, observe that

$$\begin{aligned}\frac{dn}{dK} &= \frac{1}{\mathcal{D}}\begin{vmatrix} \delta(1-q)nF_{LL} - v''(x) & -\delta(1-q)F_{LK} \\ xnF_{LL} & -xF_{LK} \end{vmatrix} \\ &= \frac{xF_{LK}v''(x)}{\mathcal{D}} \quad > \quad 0 \end{aligned} \tag{19}$$

The expression (20) verifies part (iv) of the proposition.

Return to (18) and apply it to the constraint (3). We immediately see that $\mu(K, U)$ is an increasing function of K. *This verifies (iii) of the proposition.* Finally, apply (18) to the first order condition (12) to see straight away that $\frac{dw}{dK} > 0$. *This verifies part (ii) of the proposition, and so completes the proof of Proposition 4.* ∎

8 Appendix 3: Existence of Equilibrium

Let $\tilde{K}$ denote the vector $(K_1, ..., K_m)$. For all $U \geq 0$, define a map $\phi(U)$ by

$$(1-\delta)\phi(U) \equiv r + \frac{\pi(\tilde{K}, U)}{N(\tilde{K}, U)} \sum_{j=1}^{m} n_j(K_j, U) \max\{\mu(K_j, U) - r - \delta U, 0\} \tag{20}$$

where

$$N(\tilde{K}, U) \equiv \sum_{j=1}^{m} n_j(K_j, U) \tag{21}$$

and

$$\pi(\tilde{K}, U) \equiv \frac{qN(\tilde{K}, U)}{qN(\tilde{K}, U) + \max\{\bar{N} - N(\tilde{K}, U), 0\}} \tag{22}$$

I claim the following: *if U^* satisfies $\phi(U^*) = U^*$, and if $\{x_j^*, w_j^*, n_j^*, U_j^*\}_{j=1,\ldots,m}$ are defined from U^* by the firm consistency condition, the resulting collection is an equilibrium.*

To prove this, we observe first that if $\phi(U^*) = U^*$, then $N(\tilde{K}, U^*) \leq \bar{N}$. Suppose not. Then, using (22), we have $\pi(\tilde{K}, U^*) = 1$. Using the constraint (3), and an argument similar to that in the proofs of Propositions 2 and 3, we see that $\mu(K_j, U^*) > U^*$ for all j with $n_j(K_j, U^*) > 0$. Also, by (20), $U^* \geq r + \delta U^*$. Consequently, using (20) once again, we have

$$\phi(U^*) = \frac{1}{N(\tilde{K}, U^*)} \sum_{j=1}^{m} n_j(K_j, U^*)\mu(K_j, U^*) \tag{23}$$

We already know that $\mu(K_j, U^*) > U^*$ for all j with $n_j(K_j, U^*) > 0$. But using this in (23), we contradict our supposition that $\phi(U^*) = U^*$.

Given this, it is easy to see that the utility consistency requirement is also satisfied. This proves the claim.

It remains to show that there exists U^* such that $\phi(U^*) = U^*$. To see this, observe that because each firm's optimal choice function (as expressed in the firm consistency condition) is single-valued (Appendix 2) and because the optimal choice correspondence is upper hemicontinuous (the "maximum theorem"), the optimal choice function is continuous. Consequently, using (20), we see that

(I) $\phi(U)$ is a continuous map.

Next, observe that because $r > 0$, we have from (20) that

(II) $\phi(0) > 0$.

We will need some information from Appendix 2. Using the constraint (3) and the fact that it holds as an equality in the firm's optimization problem, we see that for each $j = 1, \ldots, m$, we have (dropping the subscript j on x for simplicity):

$$\frac{d\left[\mu(K_j, U) - r - \delta U\right]}{dU} = \frac{v'(x)}{\delta(1-q)} \frac{dx}{dU} + (1 - \delta) \tag{24}$$

From Appendix 2, it is easy to see that

$$\begin{aligned} \frac{dx}{dU} &= \frac{-\delta(1-q)(1-\delta)xF_{LL}}{\mathcal{D}} \\ &\leq \frac{\delta(1-q)(1-\delta)}{v''(x)x} \end{aligned} \tag{25}$$

Combining (24) and (25), we see that

$$\frac{d\left[\mu(K_j, U) - r - \delta U\right]}{dU} \leq \frac{(1-\delta)v'(x)}{v''(x)x} + (1-\delta) \tag{26}$$

Now, as $U \to \infty$, we have either that $\mu(K_j, U) - r - \delta U$ is bounded, or that $\mu(K_j, U) - r - \delta U$ goes to ∞ along a subsequence of U. In the latter case, it is easy to see, using (3) (with equality) that $x \to B$, where B is given in (A.1). Consequently, using (A.1) again and (26), we see that in *both* cases there exist $0 \leq D, E < \infty$ such that for all $U \geq 0$ and all $j = 1, ..., m$,

$$\mu(K_j, U) - r - \delta U \leq D + EU \tag{27}$$

Using this in (20), we see that

$$(1-\delta)\phi(U) \leq r + \pi(\tilde{K}, U)\{D + EU\} \tag{28}$$

Now, note that as $U \to \infty$, we have $\pi(\tilde{K}, U) \to 0$. This can be verified using the profit maximization condition, and I omit a formal argument for the sake of brevity. Using this, we see that there exists $\bar{U}$ such that $U \geq \bar{U}$ implies that $\pi(\tilde{K}, U) \leq \beta$, where $\beta \leq \frac{1-\delta}{2E}$. For all such U, we have, using (28), that

$$\phi(U) \leq \frac{r + \beta D}{1-\delta} + \frac{1}{2}U \tag{29}$$

But (29) means that

(III) For all U sufficiently large, $\phi(U) < U$.

Combining (I), (II) and (III), we are done. ∎

References

G. Akerlof and J. Yellen (1986), *Efficiency Wage Models of the Labour Market*, Cambridge University Press.

C. Bliss and N. Stern (1978), Productivity, wages and nutrition. I. The theory, *Journal of Development Economics* **5**, 331–362.

J. Mc Call (1970), Economics of information and job search, *Quarterly Journal of Economics* **84**, 113–126.

G. Calvo (1979), Quasi-Walrasian theories of unemployment, *American Economic Review* **69**, 102–106.

G. Calvo (1981), On the inefficiency of unemployment , mimeo., Columbia University.

G. Calvo and S. Wellisz (1979), Hierarchy, ability and income distribution, *Journal of Political Economy* **87**, 991–1010.

P. Dasgupta and D. Ray (1986), Inequality as a determinant of malnutrition and unemployment:theory, *Economic Journal* **96**, 1011–1034.

B. Dutta, D. Ray and K. Sengupta (1989), Contracts with eviction in infinitely repeated principal agent relationships , in P. Bardhan (ed.), *The Economic Theory of Agrarian Institutions*, Oxford University Press, 93–121.

M. Eswaran and A. Kotwal (1985), A theory of two-tier labour markets in agrarian economies, *American Economic Review* **75**, 162–177.

S. Grossman and O. Hart (1983), An analysis of the principal agent problem, *Econometrica* **51**, 7–45.

B. Holmstrom (1977), Moral hazard and observability, *Bell Journal of Economics* **10**, 74–91.

H. Leibenstein (1957), *Economic Backwardness and Economic Growth*, Wiley (New York).

G. Mc Donald (1984), New directions in the economic theory of agency, *Canadian Journal of Economics* **17**, 415–440.

J. Mirrlees (1975), Moral hazard and unobservable behaviour , mimeo., Nuffield College, Oxford.

J. Mirrlees (1976), The optimal structure of authority and incentives within an organization, *Bell Journal of Economics* **7**, 105–131.

S. Salop (1979), A model of the natural rate of unemployment, *American Economic Review* **69**, 117–125.

C. Shapiro and J. Stiglitz (1984), Equilibrium unemployment as a worker discipline device, *American Economic Review* **74**, 433–444.

N. SINGH (1982), The possibility of nonrenewal of a contract as an incentive device in multiperiod principal agent models , mimeo., Department of Economics, University of California, Santa Cruz.

J. YELLEN (1984), Efficiency wage models of unemployment, *American Economic Review* **74**, 200–205.