Contributions to the Theory of Labour Contracts

Åsa Rosén

Submitted for the Degree of Phd in Economics

September 1991

London School of Economics and Political Science

Abstract

The thesis consists of thee parts. Part one considers firm-union bargain over wages and working conditions (effort). It studies in a partial equilibrium setting, the theoretical relationship between the scope of firm-union bargains and the outcome in terms of wages, employment and effort. In particular, the outcome of a bargain over wages and effort is contrasted with the outcome of a pure wage bargain. A main result is that both effort and wages are lower when effort requirements are negotiable (rather than being determined by the employer). The analysis yields implications for the impact of unions on productivity, and gives an explanation to cross-industry differences in union mark-up on wages.

In the second part I study labour contracts under temporarily asymmetric information. Under the assumption that workers are more heavily credit rationed than firms, the standard model of testing and self-selection in the labour market is extended in several directions. First, it is shown that ex post inefficient termination may be used as a self-selection device. This is a new explanation for up-or-out contracts in occupations where workers' productivity is revealed slowly. Second, when risk neutral workers can be of more than two different productivities, only the best worker should be overpaid. Finally, when productivity is non-verifiable, large firms may have an advantage in hiring more able workers.

The issue of discrimination in the labour market is adressed in the last part. A model in which firms have incomplete information about workers at the hiring stage is shown to entail discrimination as the unique stable equilibrium outcome, even if no agents have a taste for discrimination. Discriminated groups (e.g., blacks, women) earn lower wages, endure longer unemployment spells, and must satisfy stricter requirements in order to obtain work. The model also offers a new explanation for duration dependence in exit rates from unemployment.

Contents Page

Acknowledgments

Writing this thesis I have benefited from the intellectually stimulating environment at L.S.E.. Thus, I am very grateful to those who made may stay financially possible. These are Göteborgs Handelshögskolefonder, Helge Ax:son Johnsons Stiftelse, Torsten och Ragnar Söderbergs Stiftelse, Tore Browadhs Stiftels, Svenska Institutet, Bertil Ohlins fond and Stiftelsen Blanceflor Boncompagni Ludovise född Bildt. A special thanks than goes to Lennart Hjalmarsson for his keen encouragement when I decided to go abroad.

Here at L.S.E. I have benefitted from discussions with several people. In particular, I would like to mention participants at John Moore's lunch seminars for Phd students.

Beyond being my supervisor Chrisopher Pissarides provided invaluable moral support during the early (and frustrating) stages of writing this thesis. Alan Manning, who supervised me while Pissarides were on leave read the various drafts both thoroughly and speedily. His kind advice prompted a number of improvements.

Last, but not least I want to thank Tore Ellingsen, who suffered through endless detailed discussions about all of the chapters, and also gave suggestions for expositional improvements.

Introduction

The traditional theory of labour markets implies that a worker's wage should be equal to the value of his marginal product. However, this theory clearly fails to account for a number of important empirical phenomena. The awkward fact which has been of most concern to economists is that the theory cannot account for involuntary unemployment. But there are also other observations which have suggested the inadequacy of the theory, and it is some of these which motivate the research here.

First, the traditional theory, resting on the assumption that workers' behavior is atomistic, is obviously silent on the issue of trade unions. The analysis of unions, and their impact on wages, working conditions and employment, requires a different approach. In recent years, non-cooperative bargaining theory has been developed to a point where it can guide our thinking about wage negotiations. In formal models, it has been common to assume that bargaining occurs either over wages or wages and employment. Empirically, however, working conditions constitute an important aspect of such negotiations - much more so than does employment. The first chapter in the thesis considers this issue. There, I study how the outcome depends on the scope of the bargain. A main result is that both effort and wages are lower when effort requirements are negotiable, (rather than being determined by the employer). The model might explain the low union mark-up on wages in the public sector and also why the mark-up on wages differs significantly across industries. The analysis also casts some doubt on the traditional definition of union power as being the mark-up on the competitive wage. A strong union may well agree to low wages if working conditions are sufficiently luxurious. Thus, the mark-up on wages may overstate or understate the returns to unionization when effort is taken in to consideration.

Another phenomenon for which the traditional theory offers no explanation is the fact that firms are concerned over the recruitment of workers. According to the traditional theory, an employer should be indifferent with respect to the skills of different workers, as differences in skill would merely be reflected in the wages. One potential explanation is that information is not symmetric, i.e. the worker has information which the employer does not have. Then the employer have to make hiring decisions before he knows the exact abilities of the workers. However, it is easily shown that if workers are risk neutral, all contracts enforceable, and capital markets are perfect, the firm can simply defer the payment until productivity is revealed and pay accordingly, in which case the traditional conclusion applies; the employer is indifferent whom to hire. This was one of the basic insights of the early literature on self-selection in the labour market. Chapter 2 therefore explores the impact on hiring and firing policies when one or more of these assumptions do not hold. E.g., it is (realistically) assumed that workers' ability to lend money against future income is limited. Thus, wage payments cannot be costlessly postponed, and the use of other self-selection policies may therefore be justified. In particular, I identify conditions under which up-or-out contracts, i.e. terminations of productive relationships, are optimal. (It is noteworthy that, since self-selection is costly in this model, it could provide a rationale for careful selection of workers. Although this is not emphasized in the paper, it is obvious that pre-screening would reduce the need for self-selection. Furthermore, if application is costly to the worker, the pre-screening would also serve as a self-selection device.) Another regularity at which the analysis may throw some light is the observation that large firms usually employ high productivity workers.

A third observation with which the traditional theory is incompatible is the prevalence of discrimination - the unequal treatment of (ex ante) identical groups of workers. Indeed, equal treatment of equals is among the most robust predictions of the standard marginalist analysis of market equilibrium, unaffected by e.g. the introduction of market power. Empirically, it is well established that black men earns less than white men and women less than men. Typically, the black/white wage ratio is around 0.6 to 0.8. The same is true for the women/men wage ratio. The variance in estimates is mainly due to whether one includes variables such as education, experience and occupation. In chapter 3, I have constructed a model of discrimination in the labour market. The model exhibits multiple equilibria, but only discriminating equilibria are stable. Like in chapter 2, asymmetric information plays an important role in generating this result. However, I also depart from the Walrasian framework by couching the analysis in a search and matching context. Briefly, the discrimination occurs because a discriminated group will apply for jobs in which they are less well matched, thus fulfilling the belief held by the employer that this group of worker has a lower expected productivity.

While the three chapters break with the traditional model of labour markets, they are all linked firmly to different strands of the "new orthodoxy" being developed over the last 30 years. The development of new analytical tools has allowed the analytical consideration of bargaining, contracting under incomplete information, and search. These are essential institutional features of the labour market, and it is not surprising that we can gain improved insight into the workings of the labour market when these features are explicitly incorporated into our analysis.

Chapter 1: BARGAINING OVER EFFORT[1]

1.Introduction

In the theoretical literature on union behavior it is common to compare wages and employment in two bargaining situations. Either the union bargains with the firm over employment and wages, or the bargaining issue is wages only, and the firm decides employment (see e.g. Mc Donald and Solow (1981) and the recent survey by Farber (1987)). In reality, however, bargains between unions and firms rarely involve employment directly (see e.g. Oswald (1987)). On the other hand, bargaining over such conditions as hours and work intensity (in our terminology effort) is widespread. Two questions immediately arising from this observation are, first, what determines the scope of firm-union contracts and, second, how does this scope influence the actual outcome, i.e. wages, effort, profits, employment etc.

Here I leave aside the challenging first question, just taking as my starting point that wages and effort are in fact the variables explicitly agreed upon by the two parties.[2] The basic aim of the analysis is to study, within this more "realistic" framework, the relationship between wages, effort and employment levels. In particular I want to compare the outcome of effort bargaining with the outcome from having the firm set effort requirements unilaterally.

A problem associated with making such comparison is that it is not always reasonable to treat all bargaining variables symmetrically. The disagreement outcomes may be very different for effort compared to fiscal remuneration; e.g there may be a

[1] Valuable comments and suggestions have been received from James Albrecht, Tore Ellingsen, Steinar Holden, Michael Hoel, Richard Layard, Stephen Nickel, Andrew Oswald and Christopher Pissarides.

[2] Some documentation of the extent of effort bargaining is given in Millward and Stevens (1986) Table 9.19 and 9.20. They report that 1984, 76% of the establishments did bargain over physical working conditions and 55% over staffing/manning levels. Clark and Oswald (1989) find that 91 % of the unions bargain over working practices such as size of crews, demarcation, or the way in which new technology can be used.

strict set of working rules that can be imposed by the union, or the employer may be free to dictate all relevant effort variables. Indeed, in these matters there are large differences both across industries and across countries. Because the results are sensitive to this assumption, two situations will be portrayed. In the first situation, the employer holds the right to dictate effort, but is willing to negotiate it with the union. In the second scenario, effort, like wages, is freely negotiable, with none of the parties having special privileges with respect to any of the bargaining variables. The two cases will be referred to as voluntary and forced effort bargaining respectively. Each of these situations are then compared to the case where it is not feasible to negotiate effort.

Of course, voluntary effort bargaining entails a range of strictly Pareto-superior contracts compared to the pure wage bargain. These contracts all specify a lower level of effort and a lower wage than would have otherwise been the case. Another feature of the Pareto-dominating contracts is that they imply a higher level of employment than do the pure wage contracts. It is demonstrated that this is not due to any "work-sharing effects" from lower effort, an argument which is frequently put forward in support of shortening the working week. On the contrary, for a given wage, employment increases with effort under our specification of the model. Hence, the effect on employment is solely due to the lower wage. For the case of forced effort bargaining, the effect on employment is no longer clear, whereas the implications for wage and effort are the same.

Comparing effort levels in unionized and non-unionized firms, the model yields clear-cut predictions only with respect to effort variables which are explicitly accounted for in the firm-union negotiations. Such working conditions are expected to be better in unionized firms.

The model might explain the low union mark-up on wages in the public sector and also why the mark-up on wages differs significantly across industries. Here, these phenomena are due to differences in the scope of bargaining. The analysis also casts some doubt on the traditional definition of union power as being the mark-up on the competitive wage. A strong union may well agree to low wages if working conditions are sufficiently luxurious. Thus, the mark-up on wages may overstate or understate the returns to unionization when effort is not taken into consideration.

Other papers that have studied bargaining over working conditions are Johnson (1990) and Clark (1991). Their main focus is whether bargaining over wages and effort or over wages and capital/labour ratios can mimic bargaining over wages and employment. In the model by Clark (1991) working conditions per se does not enter workers´ utility functions. They only influence the union´s preferences through the employment outcome. Johnson (1990) does consider the case where workers care about effort directly. However, he does not compare the outcome of a pure wage bargain with the outcome of bargaining over effort as well as wages, which is my main focus here.

The paper is organized as follows. In section 2 I present necessary definitions and a characterization of the payoff functions for the firm and the union. Section 3 is an analysis of three different cases where effort is specified in the labour contract. In the benchmark case there is no union, and the firm sets both wages and effort. The second case assumes that the firm determines the effort and that there is wage bargaining afterwards. Finally, in case 3, there is simultaneous bargaining over effort and wages. In all cases the firm unilaterally chooses employment after the bargaining outcome becomes known. The endogenous variables are the wage rate, effort, profits, union utility, employment, production and product price. Section 4 contains a comparison of the outcomes under each of the model specifications. Section 5 gives a summary and discussion of the results.

2.The model

The list of working conditions that firms and workers can potentially bargain over is endless. Some examples are working hours, breaks, holidays, shifts, work assignment, work intensity (speed of assembly line, say), supervision etc. In order to make the model as general as possible I simply treat effort as an arbitrary input. Thus, effort can be interpreted as the quality as well as the quantity of labour supplied by a single worker. When using this approach, I am careful not to define a productivity-dependent wage, such as a wage per unit of output or per hour. Consequently, my wage rate is the total payment to each worker per unit time employed.

2.1 The firm

The firm chooses the level of employment to maximize profit. It faces a constant elasticity of demand function:

$$Y = AP^{-\lambda} \tag{1}$$

where λ is the demand elasticity ($\lambda>1$) and A is a positive constant. Perfect competition in the product market is described by (1) when $\lambda \rightarrow \infty$. Thus, the model specification is quite robust to the nature of competition in the product market. Production is given by:

$$Y = (Ne)^{\alpha} \tag{2}$$

where N is the number of employees, e is effort, and α describes returns to scale.

Profit in excess of fixed cost payments is given by

$$\Pi = PY - wN. \tag{3}$$

If fixed costs are strictly positive, this formulation implies increasing returns to scale.

Using equations (1), (2) and (3), profit maximization for a given e and w, gives price, production, labour demand and profit as functions of the wage and effort:

$$P(w,e) = A^{1/\lambda}B^{-\alpha/\lambda}\left(\frac{e}{w}\right)^{-\varepsilon\alpha/\lambda} \tag{4}$$

$$Y(w,e) = B^{\alpha}\left(\frac{e}{w}\right)^{\varepsilon\alpha} \tag{5}$$

$$N(w,e) = Bw^{-\varepsilon}e^{\varepsilon-1} \tag{6}$$

$$\Pi(w,e) = B\left(\frac{e}{w}\right)^{\varepsilon-1}\left(\frac{1}{\varepsilon-1}\right) \tag{7}$$

where $\varepsilon = \lambda/(\alpha+\lambda-\alpha\lambda)$ and $B=[A^{1/\lambda}(\varepsilon-1)/\varepsilon]^{\varepsilon}$. The second order conditions imply that $\varepsilon>1$.

We see from these expressions that the price depends positively on the wage and negatively on effort. Production, labour demand (which is equal to employment) and profit depend negatively on the wage and positively on effort. The positive relationship between effort and production is trivial in the sense that it generalizes to all reasonable demand functions: For a given wage increased effort means a downward shift in the firm's marginal costs. Hence, if the marginal revenue function is non-increasing, production will go up. The positive relationship between effort and employment is not equally general. With higher effort it is possible to increase production without hiring more labour. Therefore, one can construct demand and production functions for which higher productivity of labour leads to less employment.

2.2 The Union

The union has D identical members. N of these members obtain a job in the firm. The other D-N members get a utility of v. The utility enjoyed by each employed union member is assumed to be separable in wages and effort.

$$c = u(w) - g(e) \tag{8}$$

I also assume that utility increases with the wage at a decreasing rate and that the utility decreases with effort at an increasing rate, i.e. $u'(w)>0$, $u''(w)<0$, $g'(e)>0$ and $g''(e)>0$. Treating v as exogenous, the union maximizes the total expected utility of its members, Z.

$$Z(w,e) = N(w,e)(u(w) - g(e)) + (D\text{-}N(w,e))v \tag{9}$$

We can rewrite equation (9) as:

$$Z(w,e) = N(w,e)(u(w) - g(e) - v) + vD \tag{10}$$

To find the shape of the union utility function we first differentiate equation (10) with respect to w and e.

$$\frac{\partial Z}{\partial w} = N(u'(w) - \frac{\varepsilon r}{w}) \tag{11}$$

$$\frac{\partial Z}{\partial e} = N((\varepsilon\text{-}1)\frac{r}{e} - g'(e)) \tag{12}$$

where $r = u(w) - g(e) - v$. Equations (11) and (12) imply:

$$\frac{dw}{de}\Big|_Z = \frac{g'(e) - (\varepsilon - 1)r/e}{u'(w) - \varepsilon r/w} \tag{13}$$

In Appendix 1, I show formally that the resulting union indifference curves in the w-e

plane are ellipsoides, as drawn in figure 1.

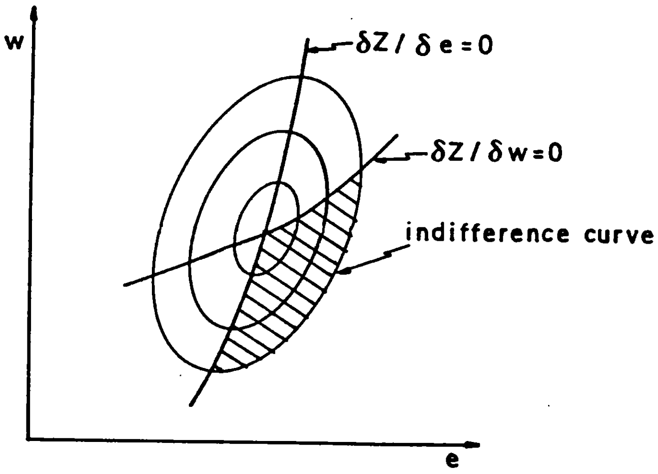

figure 1

To the right of the $\frac{\partial Z}{\partial w}$=0 line, the sign of $\frac{\partial Z}{\partial w}$ is positive, and to the left it is negative. Similarly, the $\frac{\partial Z}{\partial e}$ is negative to the right of the $\frac{\partial Z}{\partial e}$ =0 line, and positive to the left. Thus, since $\frac{\partial \Pi}{\partial w}$ <0 and $\frac{\partial \Pi}{\partial e}$ > 0, the region for possible outcomes is the shaded area in figure 1. If we compare the utility of two indifference curves, the one that lies inside the other has the higher utility. The maximum utility is of course obtained when $\frac{\partial Z}{\partial w}$=0 and $\frac{\partial Z}{\partial e}$=0, i.e. the crossing between the two lines in the figure.

3.Some scenarios

To evaluate the impact of effort bargaining, we need some standard of comparison. The two candidates are that effort is fixed unilaterally either by the firm or by the union. My reason for selecting the first alternative is that the second seems unrealistic. A firm, almost by definition, has some influence over working conditions.

As a benchmark case, I also include the situation where the firm controls both variables (i.e., there is no union).

3.1 Case 1:No Union

If there is no union to secure the workers a share of the rent, the firm maximizes profit subject to the constraint that each worker's utility, $c=u(w) - g(e)$, is not less than a minimum level, c^1. The function $u(w) - g(e)=c^1$ is convex and upwards-sloping in the *w-e* plane.

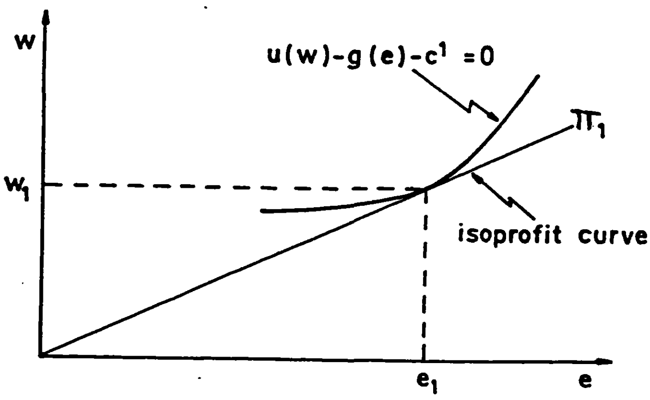

figure 2

Total differentiation of the profit function (equation (7)) gives the isoprofit curves

$$\frac{dw}{de}\Big|_{\Pi}=\frac{w}{e} \tag{14}$$

Since $\frac{d^2w}{de^2}\,|_{\Pi} = 0$, the isoprofit curves are straight lines passing through the origin. Profit increases when w/e decreases, i.e. when the slope of the isoprofit curves in figure 2 decreases. The firm chooses the solution on the function $u(w) - g(e) - c^1=0$ that gives highest profit, i.e. the tangency point (w_1,e_1).

3.2 Case 2: Bargaining over wages

Here I assume that the firm chooses effort, $e=e_2$, and that the firm and the union bargain over the wages thereafter. The firm then makes its hiring decision, conditioning on the bargaining outcome. Effort requirements are assumed to be fixed in the contract period. In particular they can not be altered without the workers' consent[3]. The bargain over wages is described by the Nash bargaining solution

$$\max_{w} X = (\Pi-\Pi_0)^{\beta}(Z-Z_0) \qquad \text{st. } e=e_2, N(w,e) \tag{15}$$

where Z_0 is the union's fallback level, and Π_0 is the firms fallback level (again without deducting fixed cost). In other words, Z_0 and Π_0 are the outcomes if there is a delay in reaching an agreement. The parameter β describes the relative bargaining power of the two parties.

Differentiating the Nash maximand with respect to the wage we get:

$$\frac{\partial X}{\partial w} = \left(\frac{\varepsilon-1}{w}\right)\Pi(\Pi-\Pi_0)^{\beta-1}[(u'(w) - \frac{\varepsilon r}{w})(\Pi-\Pi_0) - \beta(Z-Z_0)] \tag{16}$$

$$\frac{\partial X}{\partial w} = 0 \Rightarrow (u'(w) - \frac{\varepsilon r}{w})(\Pi-\Pi_0) - \beta(Z-Z_0) = 0 \tag{17}$$

Now, total differentiation of equation (17) yields:

$$\frac{dw}{de}\bigg|_{\frac{\partial X}{\partial w}=0} = \frac{\varepsilon g'(e)(\Pi-\Pi_0) + \frac{\varepsilon-1}{e}\Pi\left(wu'(w) - r(\varepsilon+\beta\varepsilon-\beta) + \beta e g'(e)\right)}{\left(-wu''(w) - \frac{\varepsilon r}{w} + \varepsilon u'(w)\right)(\Pi-\Pi_0) + (1+\beta)(\varepsilon-1)(u'(w) - \frac{\varepsilon r}{w})\Pi} \tag{18}$$

[3]If the effort requirement can be altered after the wage negotiation, the union cannot influence the workers utility, because a firm will then always choose effort so that $u(w)-g(e) = c^1$.

As already mentioned, the solution has to be where $\frac{\partial Z}{\partial w} > 0$ and $\frac{\partial Z}{\partial e} < 0$. It can be shown that $\frac{dw}{de}\Big|_{\frac{\partial X}{\partial w}=0}$ is positive in this region. This means that the higher effort the firm chooses, the higher will be the wage reached by the following wage negotiation. The firm's problem is therefore to choose effort in such a way that profit is maximized given that the solution has to be on the function defined by $\frac{\partial X}{\partial w} = 0$. Recall that $\frac{dw}{de}\Big|_{\Pi} = \frac{w}{e}$.

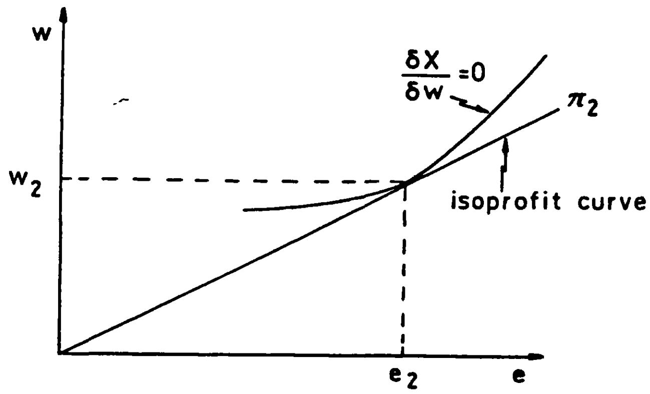

figure 3

Under the assumption that $u'''(w) > 0$ and $\varepsilon > 2$ (sufficient but not necessary conditions) it can be shown that the solution is characterized by $\frac{w}{e} = \frac{dw}{de}\Big|_{\frac{\partial X}{\partial w}=0}$, i.e. where an isoprofit curve is tangential to the function defined by $\frac{dX}{dw}=0$, (see Appendix 2).[4] This is illustrated in figure 3, where the solution is denoted (e_2, w_2). Setting equation (18) equal to $\frac{w}{e}$ we obtain;

$$[\varepsilon(\Pi-\Pi_0) + \beta(\varepsilon-1)\Pi][wu'(w) - r - eg'(e)] - w^2u''(w)(\Pi-\Pi_0)=0 \tag{19}$$

Equations (17) and (19) then give the algebraic expressions for w_2 and e_2.

[4]The necessary and sufficient condition is: $w^3(1/e^2)u'''(w)(\Pi-\Pi_0) + \varepsilon g''(e)[(\Pi-\Pi_0) + \beta(\varepsilon-1)\Pi] - w^2u''(w)[(\varepsilon-2)(\Pi-\Pi_0) + \beta(\varepsilon-1)\Pi] > 0$.

3.3 Case 3: Bargaining over effort and wages.

In this case, the firm and the union bargain over wages and effort simultaneously. The firm then chooses the level of employment. Clearly, if effort is negotiated symmetrically to wages - with no party having the right to invoke minimum requirements or maximum limits - the Nash Bargaining solution can be derived as before. (This is done in the next section.) However, in order to accommodate the case where one party (here, the firm) can constrain effort, I will also derive the full contract curve (i.e. all Pareto optimal solutions).

One condition for Pareto efficiency is that $\frac{dw}{de}\big|_{\bar{\Pi}} = \frac{dw}{de}\big|_{Z}$, i.e. that an isoprofit curve is tangential to a union indifference curve. If we set $\frac{dw}{de}\big|_{\bar{\Pi}} = \frac{dw}{de}\big|_{Z}$, i.e. set equation (14) equal to equation (13) we get

$$wu'(w) - r - eg'(e) = 0 \tag{20}$$

Total differentiation of (20) gives

$$\frac{dw}{de}\Big|_{\frac{dw}{de}|_{\bar{\Pi}} = \frac{dw}{de}|_{Z}} = \frac{eg''(e)}{wu''(w)} < 0 \tag{21}$$

Thus, equation (20) is downward sloping in the w-e plane, as illustrated in figure 4.

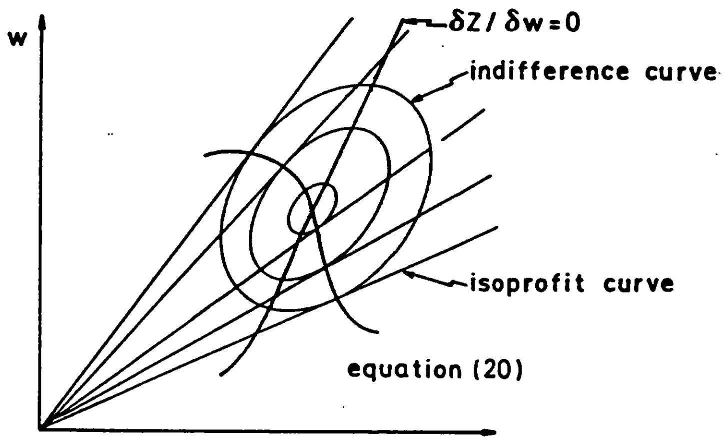

figure 4

In the special case where the marginal disutility of effort is constant ($g''(e)=0$) the contract curve will be horizontal in the w-e plane, i.e. the level of the wage will be unrelated to the bargaining power and fall back levels of the two parties and the three parameters (λ, α and A). Thus, when $g''(e)=0$ we have a sticky wage result. In the special case where the marginal utility of wage is constant the contract curve is vertical in the w-e plane and we get a sticky effort result. As figure 4 shows, only points on the curve defined by equation (20) where $\frac{\partial Z}{\partial w}>0$ are Pareto efficient. Thus, the contract curve is jointly defined by equation (20) and the condition that $\frac{\partial Z}{\partial w}>0$. The profit increases and the union's utility decreases along the curve when e increases.

4. A comparison of the cases:

I now compare the three cases in order to see the effect of each bargaining structure on wages, effort, firm's profit, union's utility, employment, production and price. The solutions to Case 1, Case 2 and Case 3 are denoted with subscripts 1,2 and 3 respectively.

4.1 Comparison of cases 2 and 3

As already emphasized, it is not entirely clear what we should mean by a comparison of the cases 2 and 3. On the one hand we could think of fallback levels and bargaining power as variables invariant to the bargaining issues. This approach is consistent e.g. with the common interpretation of the fallback levels and the bargaining power as indices of payoff during a strike and of impatience respectively. Neither of these factors are likely to be dependent on the bargaining issues. On the other hand, since in the second case the firm is given power unilaterally to fix the effort level, the most interesting question might be how case 3 differs from case 2 when the firm can refuse to bargain over effort. Within an industry the first comparison can be seen as studying the consequences of depriving the firm of the right to determine the effort level, whereas the second assumes that the firm voluntarily trades its privilege to determine effort. If on the other hand the comparison is across industries, there may be technological factors determining what variables are negotiable, and "forced" effort bargaining may be the more interesting case. Below I report both.

Recall that equation (17) and (19) give the solution for case 2.

$$[\varepsilon(\Pi-\Pi_0) + \beta(\varepsilon-1)\Pi][wu'(w) - r - eg'(e)] - w^2u''(w)(\Pi-\Pi_0)=0 \qquad (19)$$

Since $u''(w)<0$ and $\Pi-\Pi_0>0$, $u'(w) - r - eg'(e)$ must be less than zero for equation (19) to hold. This expression is less than zero above the contract curve and greater than zero below. Thus, the solution to case 2 must be above the contract curve. Only in the case where the marginal utility of income is constant ($u''(w)=0$) will the solution to case 2 be on the contract curve.

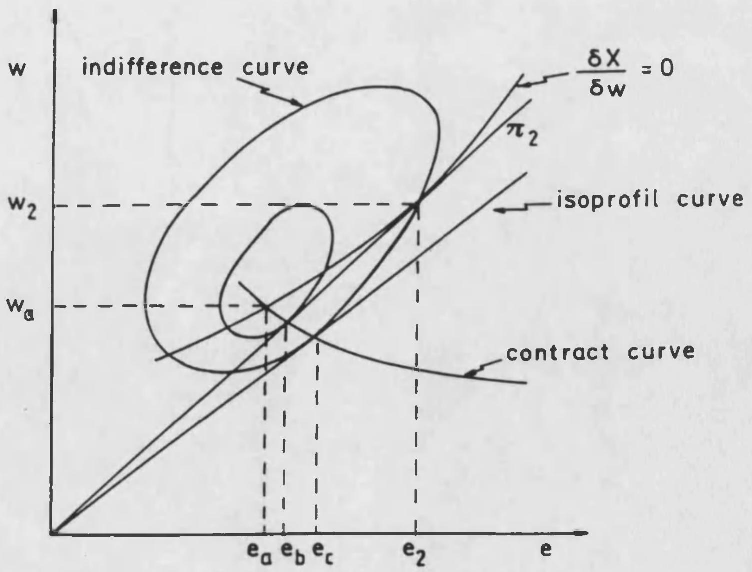

figure 5

Comparison of cases 2 and 3 with "forced" effort bargaining

When the firm is denied the right to determine effort, with given fallback levels and bargaining power, we need to find the Nash bargaining solution for both the parameters w and e. If the Nash maximand is again denoted X, the solution must have the property that $\frac{\partial X}{\partial w} = \frac{\partial X}{\partial e}$=0. Rather than deriving the explicit solution, we can now use diagram 5. The line $\frac{\partial X}{\partial w}$ =0 is already there, and since the bargain solution is Pareto-optimal, it must be described by the intersection of this line and the contract curve, i.e. the point (w_a,e_a). It can be seen from figure 5 that the wage and effort will then be lower when the two parties bargain over both wages and effort. Note that the lower effort is desirable for the union not because it raises employment -on the contrary, for a given wage it lowers employment, -but because it raises the utility of employed workers. It can also be seen that profit will be lower and the union's utility higher in case 3 compared to case 2.

Without making further assumptions, we cannot show whether employment will be higher or lower in case 3 compared to case 2. Since profit is lower in case 3 we know that $(\frac{w}{e})_3 > (\frac{w}{e})_2$, and so production is lower and price higher in case 3 compared to case 2.

Comparison of cases 2 and 3 with voluntary effort bargaining

When the firm trades its privilege to determine effort voluntarily, the wage-effort bargain will Pareto dominate the pure wage bargain. The Pareto superior solutions lie on the contract curve between e_b and e_c in figure 5. Here, employment is unambiguously higher than in case 2. From equation (7) we have that $N=B(\frac{e}{w})^{\varepsilon}(\frac{1}{e})$, and since $\Pi_3>\Pi_2$ we have that $(\frac{e}{w})_3>(\frac{e}{w})_2$. Since also $e_3<e_2$ this implies $N_3>N_2$. Because here $\Pi_3>\Pi_2$, production will be higher and price lower in case 3 compared to case 2.

4.2 Comparison of the non-union case with the union cases

In figure 6 we have drawn the solutions to cases 1,2 and 3. If the outside opportunity, v, is equal to c^1, the $u(w)$-$g(e)$-c^1=0 function intersects the contract curve where $w=w^1$ and $e=e^1$. This is the case drawn in figure 6. It can then be seen from figure 6 that the wage will be higher and effort lower in case 3 compared to case 1. The latter of these two results (lower effort) was also noted by Johnson (1990). This implies that employment and production will be lower and prices higher in the union case with effort bargaining than in the non-union case.[5]

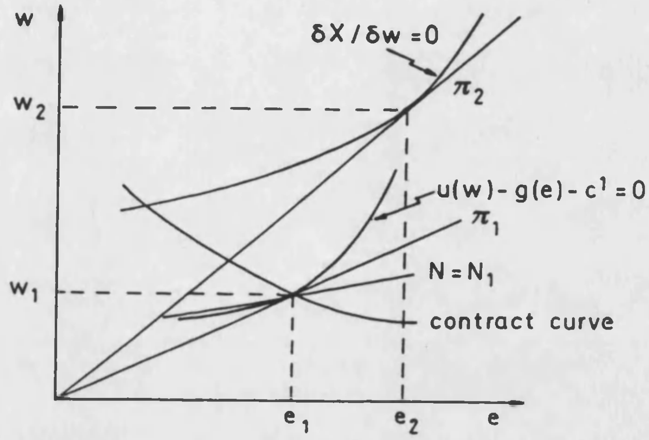

figure 6

Apart from the level of effort, the qualitative differences between the two union cases and the non-union case are the same. When there is bargaining over wages only, effort can be higher in the unionized firm than in the non-unionized firm. Whether it will be higher or lower depends on how well the union can compensate itself for higher effort by getting higher wages in the wage negotiation. One important factor is the concavity of the utility function, $-u''(w)$. The higher $-u''(w)$, the less the union is able to compensate its members. This is because the higher $-u''(w)$, the less the unions utility will increase when the wage goes up, whereas profit decreases in the same proportion regardless of worker utility. The wage will thus respond less to an increase

[5]These results hold true also if $v>c^1$.

in effort, and effort will be correspondingly higher. Wages will be higher in case 2 than in case 1 since the solution to case 2 lies above the contract curve. Employment will be lower in case 2 than in case 1. This can be seen by looking at the constant employment line where $N=N_1$, drawn in figure 6. The $N=N_1$ line has a smaller slope than the iso-profit curves, $\frac{dw}{de}\big|_N = \left(\frac{\varepsilon-1}{\varepsilon}\right)\left(\frac{w}{e}\right)$. To the left of the $N=N_1$ line, employment is lower than N_1, to the right it is higher. Production will be lower and price higher in case 2 compared to case 1, since profit is lower in case 2 than in case 1.

5. Summary and discussion

The main results in the paper are summarized in table 1.

Table 1: Summary of the results

Comparison of Case 1 with Case 2 and Case 3		Comparison of Case 2 with Case 3	
		forced effort bargain	voluntary effort bargain
$\Pi_1 > \Pi_2$	$\Pi_1 > \Pi_3$	$\Pi_2 > \Pi_3$	$\Pi_2 < \Pi_3$
$Z_1 < Z_2$	$Z_1 > Z_3$	$Z_2 < Z_3$	$Z_2 < Z_3$
$w_1 < w_2$	$w_1 < w_3$	$w_2 < w_3$	$w_2 > w_3$
$e_1 \gtrless e_2$	$e_1 > e_3$	$e_2 > e_3$	$w_2 > w_3$
$N_1 > N_2$	$N_1 > N_3$	$N_2 < N_3$	$N_2 \gtrless N_3$
$Y_1 > Y_2$	$Y_1 > Y_3$	$Y_2 < Y_3$	$Y_2 > Y_3$
$P_1 < P_2$	$P_1 < P_3$	$P_2 > P_3$	$P_2 < P_3$

Case 1: Non-union case, Case 2: Bargain over effort only, Case 3: Bargain over wages and effort, Π=profit, Z=union utility, w=wage, e=effort, N=employment, Y=production, P=product price.

If the firm is deprived of the right to determine effort and instead has to bargain over effort and wages, the firm's profit is reduced and the union's utility increased. In this case, wages, effort and production are lower and price higher than when bargaining is only over wages. Whether employment is lower or higher when the bargain includes effort is not possible to say without further assumptions.

If the firm has the right to determine the effort level when negotiations with the union fail, wages, effort and price are lower and employment and production higher when bargaining is over both effort and wages and not over wages only.

Freeman and Medoff (1984) report some empirical comparisons of productivity levels in unionized and non-unionized firms (Table 11.1). They find that the effect of unionization on productivity is not conclusive, though mainly positive. Metcalf (1988) finds that unions are associated with lower productivity. Our results suggest that the inconclusiveness may be due to variations in bargaining issues across firms (and unions). In firms where there is bargaining over productivity (effort) variables, effort is lower than in the non-unionized firms. On the other hand, our model does not yield predictions about the relative productivity levels when the bargain is over wages only. Without knowing the scope of firm-union negotiations it is therefore not possible to say whether effort (productivity) will go down or up in the case of unionization.

The union mark-up on wages is generally found to be lower in public sectors than in private.[6] Millward and Stevens (1986) present evidence that bargaining over effort is more common in the public sector than in the private.[7] Thus, the lower union mark-up in the public sector could be explained by a broader scope of bargaining in the public sector. The model might similarly help to explain the big differences in union mark-up on wages across industries, as documented by Stewart (1983,1987) and Lewis (1986). More generally, the analysis of unions based on the assumption that unions ignore questions of effort is suspect. The mark-up on the competitive wage is

[6]Lewis (1990) and Ehrenberg and Schwartz (1986) survey different studies. It should be mentioned that both Lewis and Ehrenberg and Schwartz are critical of many of the studies, and suspect that they underestimate the union mark-up on wages in the public sector.

[7]Table 19.19 and 19.20.

frequently used as a measure of union power, (see e.g. Layard and Nickell (1986)). This measure ignores that a union may use some of its power to secure good working conditions. Although in our model, the wage is a monotonically increasing function of union bargaining power, inter-industry differences in the scope of bargaining, in technology (i.e. the costliness of providing good working conditions) and taste (caused eg by differences in workers´wealth) may destroy this relationship. Hence, the mark up on the competitive wage is not an entirely reliable measure of union power. It may overstate the returns to unionization if effort is higher in unionized firms and understate the returns if effort is lower.

Appendix 1.

I will here show that the union's indifference curves are ellipsoids in the *w-e* plane.

Since the union utility, Z, is a function of *e* and *w*, $\frac{dw}{de}|_Z = -\frac{\partial Z}{\partial e} / \frac{\partial Z}{\partial w}$.

Thus,

$\frac{dw}{de}|_Z > 0$ if $\frac{\partial Z}{\partial w}>0$ and $\frac{\partial Z}{\partial e}<0$ or if $\frac{\partial Z}{\partial w}<0$ and $\frac{\partial Z}{\partial e}>0$

$\frac{dw}{de}|_Z < 0$ if $\frac{\partial Z}{\partial w}>0$ and $\frac{\partial Z}{\partial e}>0$ or if $\frac{\partial Z}{\partial w}<0$ and $\frac{\partial Z}{\partial e}<0$

$\frac{dw}{de}|_Z = \infty$ if $\frac{\partial Z}{\partial w}=0$ and $\frac{dw}{de}|_Z = 0$ if $\frac{\partial Z}{\partial e}=0$

The $\frac{\partial Z}{\partial w}=0$ function intersects the $\frac{\partial Z}{\partial e}=0$ function only once, and for small values of *e* the *w* that solves $\frac{\partial Z}{\partial w}=0$ is higher than the *w* that solves $\frac{\partial Z}{\partial e}=0$. Also, for high values of *e*, the *w* that solves $\frac{\partial Z}{\partial w}=0$ is lower than the *w* that solves $\frac{\partial Z}{\partial e}=0$ (Appendix 1a). It can also be shown that $\frac{d^2w}{de^2} > 0$ when $\frac{\partial Z}{\partial w}>0$ and $\frac{\partial Z}{\partial e}<0$ and $\frac{d^2w}{de^2} < 0$ when $\frac{\partial Z}{\partial w}<0$ and $\frac{\partial Z}{\partial e}>0$ (Appendix 1b). Hence, the indifference curves are ellipsoids.

Appendix 1a

$$Z = Bw^{-\varepsilon}e^{\varepsilon-1}(u(w) - g(e) - v) + vD \qquad \text{(A1.1)}$$

$$\frac{\partial Z}{\partial w} = 0 \iff u'(w) - \left(\frac{\varepsilon}{w}\right)(u(w)-g(e)-v) = 0 \qquad \text{(A1.2)}$$

$$\frac{\partial Z}{\partial e} = 0 \iff -g'(e) + \left(\frac{\varepsilon-1}{e}\right)(u(w)-g(e)-v) = 0 \qquad \text{(A1.3)}$$

$$\frac{dw}{de}\Big|_{\frac{\partial Z}{\partial w}=0} = \frac{\varepsilon g'(e)}{(\varepsilon - 1)u'(w) - wu''(w)} \qquad \text{(A1.4)}$$

$$\frac{dw}{de}\Big|_{\frac{\partial Z}{\partial e}=0} = \frac{\varepsilon g'(e) + eg''(e)}{(\varepsilon - 1)u'(w)} \tag{A1.5}$$

Since $u'(w) > 0$, $g'(e) > 0$, $u''(w) < 0$ and $g''(e) > 0$ $\frac{dw}{de}\Big|_{\frac{\partial Z}{\partial w}=0} < \frac{dw}{de}\Big|_{\frac{\partial Z}{\partial e}=0}$ for a given w and e. Thus, the $\frac{\partial Z}{\partial w}$=0 curve does only intersect the $\frac{\partial Z}{\partial e}$ =0 curve once, and for small values of e, the w that solves $\frac{\partial Z}{\partial w}$=0 is higher than the w that solves $\frac{\partial Z}{\partial e}$ =0. Conversely, for high e the w that solves $\frac{\partial Z}{\partial w}$ is lower than that for $\frac{\partial Z}{\partial e}$.

Appendix 1b

I will here show here that $\frac{d^2w}{de^2}\Big|_Z \gtrless 0$ when $\frac{\partial Z}{\partial w} \gtrless 0$ and $\frac{\partial Z}{\partial e} \lesseqgtr 0$.

$$\frac{dw}{de}\Big|_Z = \frac{g'(e) - (\varepsilon - 1)r/e}{u'(w) - \varepsilon r/w} \tag{A1.6}$$

$r = u(w) - g(e) - v$

$$\frac{d^2w}{de^2}\Big|_Z = \frac{\partial\left(\frac{dw}{de}\Big|_z\right)}{\partial e} + \frac{\partial\left(\frac{dw}{de}\Big|_z\right)}{\partial w}\frac{dw}{de}\Big|_Z \tag{A1.7}$$

$$\frac{\partial\left(\frac{dw}{de}\Big|_z\right)}{\partial e} = \frac{g''(e) + (\varepsilon-1)\frac{r}{e^2}\left(u'(w) - \varepsilon\frac{r}{w}\right) + (\varepsilon-1)\frac{1}{e}g'(e)u'(w) - \frac{\varepsilon}{w(g'(e))^2}}{\left(u'(w) - \varepsilon\frac{r}{w}\right)^2} \tag{A1.8}$$

$$\frac{\partial\left(\frac{dw}{de}\Big|_z\right)}{\partial w}\frac{dw}{de}\Big|_Z =$$

$$\frac{\left(g'(e) - \frac{(\varepsilon-1)r}{e}\right)\left(-(\varepsilon-1)\frac{(u'(w))^2}{e} - g'(e)\left(u''(w) + \frac{\varepsilon r}{w^2} - \frac{\varepsilon u'(w)}{w}\right) + \frac{r(\varepsilon-1)}{e}\left(u''(w) + \frac{\varepsilon r}{w^2}\right)\right)}{\left(u'(w) - \varepsilon\frac{r}{w}\right)^3}$$

(A1.9)

I will look at the sign of $\frac{d^2w}{de^2}|_Z$ along the curve defined by $\frac{dw}{de}|_\Pi = \frac{dw}{de}|_Z$ (equation (20) in the text)

$$\frac{dw}{de}|_\Pi = \frac{dw}{de}|_Z \Rightarrow g'(e) = \frac{wu'(w) - r}{e} \tag{A1.10}$$

By inserting equation (A1.10) in (A1.8) and (A1.9) we get

$$\frac{\partial\left(\frac{dw}{de}|_Z\right)}{\partial e} = \frac{\left(\frac{w}{e^2}\right)\left(u'(w) - \frac{\varepsilon r}{w}\right)\left(\frac{e^2g''(e)}{w} + \frac{\varepsilon r}{w} - u'(w)\right)}{\left(u'(w) - \varepsilon\frac{r}{w}\right)^2} \tag{A1.11}$$

$$\frac{\partial\left(\frac{dw}{de}|_Z\right)}{\partial w}\frac{dw}{de}|_Z = \frac{\left(\frac{w}{e^2}\right)\left(u'(w) - \frac{\varepsilon r}{w}\right)\left(u'(w) - \frac{\varepsilon r}{w} - wu''(w)\right)}{\left(u'(w) - \varepsilon\frac{r}{w}\right)^2} \tag{A1.12}$$

$$\frac{\partial\left(\frac{dw}{de}|_Z\right)}{\partial e} + \frac{\partial\left(\frac{dw}{de}|_Z\right)}{\partial w}\frac{dw}{de}|_Z = \frac{e^2g''(e) - w^2u''(w)}{e^2\left(u'(w) - \varepsilon\frac{r}{w}\right)} \tag{A1.13}$$

since $g''(e) > 0$, $u''(w) < 0$ and $\frac{\partial Z}{\partial w} = N(u'(w) - \varepsilon r/w)$, we have shown that $\frac{d^2w}{de^2}|_Z \gtrless 0$ when $\frac{\partial Z}{\partial w} \gtrless 0$ and $\frac{\partial Z}{\partial e} \lessgtr 0$ along the curve defined by equation (20) in the text. Since equation (20) is monotonic (for a given w there is only one e and for a given e there is only one w), and since $\frac{dw}{de}|_Z > 0$ when $\frac{\partial Z}{\partial w} > 0$ and $\frac{\partial Z}{\partial e} < 0$, we know that $\frac{d^2w}{de^2}|_Z \gtrless 0$ if $\frac{\partial Z}{\partial w} \gtrless 0$ and $\frac{\partial Z}{\partial e} \lessgtr 0$

Appendix 2.

I will here show that when $\frac{dw}{de}\big|_{\frac{\partial X}{\partial w}=0}=\frac{w}{e}$, $\frac{d^2w}{de^2}\big|_{\frac{\partial X}{\partial w}=0}>0$, which is needed for case 2 to have an interior solution.

$$\frac{dw}{de}\Big|_{\frac{\partial X}{\partial w}=0}=\frac{\varepsilon g'(e)(\Pi-\Pi_0)+\frac{\varepsilon-1}{e}\Pi\big(wu'(w)-r(\varepsilon+\beta\varepsilon-\beta)+\beta e g'(e)\big)}{\left(-wu''(w)-\frac{\varepsilon r}{w}+\varepsilon u'(w)\right)(\Pi-\Pi_0)+(1+\beta)(\varepsilon-1)(u'(w)-\frac{\varepsilon r}{w})\Pi}\qquad\text{(A2.1)}$$

When $\frac{dw}{de}\big|_{\frac{\partial X}{\partial w}=0}=\frac{w}{e}$ we get (after a lot of algebra)

$$\frac{d^2w}{de^2}=\left(\frac{1}{q}\right)\{w^3\left(\frac{1}{e}\right)^2u'''(w)(\Pi\text{-}\Pi_0)+\varepsilon g''(e)[(\Pi\text{-}\Pi_0)\ +\beta(\varepsilon\text{-}1)(\Pi\text{-}\Pi_0)\ ]$$

$$-w^2u''(w)[(\varepsilon\text{-}2)(\Pi\text{-}\Pi_0)\ +\beta(\varepsilon\text{-}1)\Pi]\}\qquad\text{(A2.2)}$$

where $q=(-wu''(w)-\frac{\varepsilon r}{w}+\varepsilon u'(w))(\Pi\text{-}\Pi_0)\ +(1+\beta)(\varepsilon\text{-}1)(u'(w)-\frac{\varepsilon r}{w})\Pi>0$

Under the assumption that $u'''(w)>0$ and $\varepsilon>2$, all the terms are positive.

References

Clark, Andrew, (1991): Efficient Bargain and the McDonald-Solow Conjecture, *Journal of Labor Economics*, 8, October, 502-528.

Clark, Andrew and Andrew J. Oswald, (1989): An Empirical Study of Union Preferences, London School of Economics, Centre for Labour Economics, Discussion Paper no. 352, 1989.

Ehrenberg, Ronald G. and Joshua L. Schwartz, (1986): Public-Sector Labor Markets, In *Handbook of Labor Economics* (ed. O.Ashenfelter and R. Layard), 2, chapter 22, North Holland, Amsterdam.

Farber, Henry S., (1986): The Analysis of Union Behavour, In *Handbook of Labor Economics* (ed. O.Ashenfelter and R. Layard), 2, chapter 18, North Holland, Amsterdam.

Johnson, George E., (1990): Work Rules, Featherbedding, and Pareto-optimal Union-Management Bargain, *Journal of Labor Economics*, 8, January, Part 2, 237-259.

Freeman, Richard B. and James L. Medoff, (1984): *What do Unions Do?* Basic Books, New York.

Layard, Richard and Stephen Nickell, (1986): Unemployment in Britain, *Economica*, 53, Supplement, 121-169.

Lewis, H. Gregg, (1990): Union/Nonunion Wage Gaps in the Public Sector, *Journal of Labor Economics*, 8, January, Part 2, 260-328.

Lewis, H. Gregg, (1986): Union Relative Wage Effect. In *Handbook of Labor Economics* (ed. O.Ashenfelter and R. Layard), 2, chapter 20, North Holland, Amsterdam.

McDonald, Ian M. and Robert M. Solow, (1981): Wage Bargaining and Employment, *American Economic Review*, 71, December, 896-908.

Metcalf, David, (1988): Trade Unions and Economic Performance: The British Evidence, London School of Economics, Centre for Labour Economics, Discussion Paper no. 320.

Millward, Neil and Mark Stevens, (1986): *British Workplace Industrial Relations* 1980-1984. The DE/ESRC/PSI/ACAS Surveys, Gower, Aldershot.

Oswald, Andrew J., (1987): Efficient Contracts are on the Labour Demand Curve: Theory and Facts. London School of Economics, Centre for Labour Economics, Discussion Paper no. 284.

Stewart, Mark B., (1987) : Collective Bargaining Arrangements, Closed Shops and Relative Pay, *The Economic Journal*, 97, March, 140-156.

Stewart, Mark B., (1983): Relative Earnings and Individual Union Membership in the United Kingdom, *Economica*, 50, May, 111-125.

Chapter 2: TEMPORARILY ASYMMETRIC INFORMATION AND LABOUR CONTRACTS[8]

1. Introduction

In the literature on adverse selection in the labour market it has been common to assume either (i) that the firm can never learn anything about the abilities of individual workers, or (ii) that learning is instantaneous. The first assumption is made by e.g. Weiss (1980), whereas the second is associated with the works of Guasch and Weiss (1980,1981,1982). Arguably, the intermediate case is more realistic; when a worker first applies for a job, the firm's estimate of the worker's productivity is imperfect and less precise than that held by the worker himself. However, as time goes by, the employer's estimate is improved - possibly converging to the true productivity. This paper is concerned with the optimal labour contracts in such an environment.

In a world with no capital market imperfections, this intermediate case does not differ in any important way from the situation where the employer learns instantaneously, because payment can just be deferred until productivity is revealed. But the capital markets are empirically known to be imperfect; a substantial number of workers being severely credit rationed.[9] In this situation, the firm will impose a cost on the worker if the labour contract specifies a very low wage early in tenure. The present model therefore studies the consequences for optimal labour contracts of workers being more severely credit rationed than firms.

In order to understand the features of the model, it is useful to recall the analysis

[8] Valuable comments and suggestions have been received from Tore Ellingsen, Eric Hansen, Alan Manning, Dilip Mookherjee, John Moore, Tore Nilssen, Christopher Pissarides and Bengt Rosén.

[9] Jappelli (1990) estimates that 20% of the U.S. families are credit rationed. He also reports that important variables are; low age, no established credit history and low income. It should also be emphasized that Jappelli defined a person as credit rationed only if he was unable to obtain any credit at all. In another recent paper, Mayfield (1990) finds that 43% of households under the age of 30 are credit constrained and that 37% of households under the age of 50 face significant constraints.

of Guasch and Weiss (1980, 1981, 1982). They show that if applicants can be subjected to a test, where the test score is related to their expected productivity (known to the worker), the firm can sort the applicants by requiring that each worker pays an application fee and offer them contracts which are contingent on the test scores. In equilibrium, only highly productive workers will apply. Thus, the firm has achieved its aim; to attract able workers without having to pay less able workers the same wage. This result remains valid within the present model as well. But other results do not survive. First, in Guasch and Weiss there will be no inefficient terminations under the assumption of competition in the labour market. In the present paper conditions are developed under which such terminations are optimal. Secondly, Guasch and Weiss conclude that a substantial number of workers (everyone who pass a certain test) will be overpaid - at least late in their career.[10] Here, I show that this is an artefact due to the assumption that there are only two possible productivity levels. With risk neutral workers, typically only the very best worker will receive a payment in excess of the value of his marginal product. In addition to these two results, the present framework also allows me to address some new questions. E.g., large firms may have a comparative advantage in employing workers of high productivity, and that top wages will vary less over the business cycle than do other wages.

As indicated, the model is largely driven by the following two assumptions. First, instead of only two possible productivity levels, I allow for a continuum. (Actually, for much of the analysis three levels would suffice.) Secondly, capital markets are not perfect. A third important assumption is that productivity levels are not verifiable. The non-verifiability of production levels means that the worker cannot be fined for bad performance ex post (if he could, consumption smoothing is not a problem). In the absence of reputation building by firms, it may also exclude as non-credible any wage schedule depending on actual productivity.

Under these assumptions it may be profitable to use inefficient terminations as a

[10] If the test is taken to be an initial employment period, and capital markets are perfect, this is a valid interpretation of their analysis.

self-selection device, because it is costly to induce self-selection by paying a low wage to credit constrained workers. The workers need to finance consumption from wage payments, and will prefer a flatter wage schedule to a steeper one. The longer it takes to learn the worker's productivity, the more important is the capital market imperfections, and the more willing is the firm to use termination threats as a self selection device.

As in the model of Guasch and Weiss (1981), the firm must compensate for a low initial wage by promising later to pay an expected wage which is higher than the workers' expected productivity. This of course begs the question of credibility. Here, there is an additional credibility problem associated with the threat of firing productive workers.[11] The common way to defend such deviations from sequential rationality is that the firm lives for many periods and can establish a reputation for sticking to the ex ante optimal contract. But this explanation does not look entirely convincing in the present context, for the following reason. The ex ante optimal contract specifies that only the very highest possible productivity will earn a wage premium.[12] Presumably, this productivity level will occur so rarely, and require such a large premium, that even concerns for reputation cannot keep the firm from cheating after the fact (remember; productivity levels are not verifiable). A natural way to get around this problem is for the firm to condition its remuneration on relative rather than absolute performance. Arguably, a rank-order contract is easily verifiable. At the very least it seems a more convincing scheme around which to build a reputation.

Here, it is shown that a rank-order contract will perform "almost" as well as the

[11] In the present model, it may be credible to terminate productive workers. Suppose that abilities are not firm-specific, and that firms can observe each others employment contracts and wage offers. Then it becomes impossible to pay employees less than their marginal revenue product in period 2: here, the wage offer by the original employer reveals the employee's productivity, and competitors will bid the wage up to the point where the firm earns no rent on the worker. Knowing that this will happen, the firm might as well fire the worker without revealing her type through a wage offer. Separation is therefore ex post costless for the firm.

[12] This entails a technical problem as well: If there is an interval of possible productivities is unbounded from above, the problem has no solution.

ex ante optimal contract. A feature of the contract is that a prize is given to the most productive worker. A similar result has earlier been derived in a moral hazard framework by Rosen (1986).

Interestingly, if the firms are restricted to rank-order contracts the model suggests that self-selection can be obtained at a smaller cost in a large firm. In a small firm, even a less able worker has a realistic chance of being the the most productive, and the wage profile therefore has to be steeper to deter such workers from applying. Consequently, we may expect large firms to offer contracts inducing the more able workers to apply, whereas small firms employ workers who are less optimistic with respect to future productivity.

The paper is organized as follows. Section 2 describes the basic model. The optimal contract when firms do not cheat are derived in 3. In section 4, I describe the optimal rank-order contract. The relation to other literature is spelled out in section 5.

2. The model

Employees can be of two types, α and β. The type is private information to each worker. Type α has a higher expected productivity in the industry than type β. Employees live for two periods. Productivity in period 1 is observed only ex post, just prior to period 2.

Productivity, x, in period 2 is equal to actual productivity in period 1. Wage in period 1 is denoted w_1 and $w_2(x)$ denotes wage in period two given that the productivity in period 1 was x.

Productivity is a random variable. Each type is described by a density function $f_\iota(x)$, ($\iota \in \alpha,\beta$) with support $X=[l,m]$. I will assume that $f_\alpha(x)/f_\beta(x)$ is increasing in x, i.e. the Monotone Likelihood Ratio Property (MLRP). Let E_ι denote the expected value of the marginal productivity of workers of type ι. The number of

α-types and β-types in the economy is given and μ is the proportion of α-types. Firms and employees are risk neutral. Firms have a zero discount rate, but employees are assumed to have a positive discount rate r. This is a simple way to capture the fact that workers' access to credit is worse than that of the firm.[13]

Expected utility for a worker of type ι joining the industry is

$$U_\iota = w_{1\iota} + \frac{\int_{x_\alpha}^{m} w_{2\iota}(x) f_\iota(x)dx + F_\iota(x_\alpha)\overline{w}}{1+r} \qquad \iota \in \{\alpha,\beta\} \qquad (1)$$

where $\overline{w}$ is the wage in another industry and x_c is the minimum productivity level required to stay in the firm in period 2. I.e., expected utility is equal to wage in period 1 plus expected wage payments in period 2 divided by $1+r$. The criterion for a ι-worker to join the industry (the industry participation constraint) is

$$U_\iota \geq \overline{w} + \frac{\overline{w}}{1+r} \qquad \iota \in \{\alpha,\beta\} \qquad \text{(PC)}$$

I assume that (PC) is satisfied for both α-types and β-types. All the firms in the industry have an identical constant returns to scale technology. Each firm maximizes profit and takes the product price, the workers' reservation utility and the employment contracts of other firms as given. Without loss of generality the product price is normalized to 1.

Firms live forever. There is free entry into the industry, so that in equilibrium it is not possible for a firm to enter and earn positive profits. Since we have a constant returns to scale technology, the equilibrium number of firms is indeterminate.

[13] A more rigorous approach would be to define workers' utility over consumption in the two periods, and let preferences be convex. Since workers are assumed to be risk neutral, there is no room for an insurance market. Let interest rates be zero. A risk neutral worker would wish to equate marginal utility of consumption in period 1 with the marginal utility of expected consumption in period 2. Of course, if the wage in period 1 is lower than the wage in period 2, this would require borrowing. If the worker face an absolute credit constraint, or would have to pay a larger interest than the employer, a higher wage in period 1, ceteris paribus, yields a greater benefit to the worker than it is costly to the firm.

Employees are equally productive in each firm.[14] Further, I assume that firms can observe each others' contracts, but not a particular worker's productivity. This implies that for each group of workers who earn the same wage, the wage has to be greater than or equal to their average expected productivity. Formally;

$$w_{2\iota}(x') \geq E_\iota[x \mid w_{2\iota}(x')] \qquad \text{for all } x' \in X,\ \iota \in \{\alpha,\beta\}$$

The fact that only contracts are observable implies that any two workers earning the same wage look the same to an outside firm, as do all workers who are fired. This assumption provides a link between this article and the literature on visible and invisible workers, as developed by Waldman (1984) and Milgrom and Oster (1987) and the work by Geenwald (1986). There too, promotion, wage rises and firing decisions are signals to the market about a specific workers quality, and outside firms have less information than the does the present employer. I will assume that the expected productivity given that a worker is fired is less than the alternative wage, i.e. $E_\iota[x \mid x \leq x_{c\iota}] < \overline{w}$. This implies that workers who are fired will move to another industry. This is only a simplification and do not affect any of the main results.

Clearly, termination of productive relationships occurs if $x_{c\iota} > \overline{w}$. Note that if outside firms could distinguish between the workers who are fired, $x_{c\iota}$ can not be above $\overline{w}$. Workers with productivities above $\overline{w}$ would then immediately be hired by another firm and due to competition their wage would be equal to their productivity. By implication, firing would not be an effective threat in the first place. Thus, my assumption of a continuum of possible performance levels as opposed to only two, as e.g. in Guasch and Weiss (1980, 1981, 1982), is instrumental in bringing about ex post inefficient firing when there is more than one firm in the industry.[15] Finally, I

[14]As long as productivity in one firm is positively correlated with productivity in another firm a version of the results go through.

[15]To be precise, what is needed is more than one performance level of those who are fired, and that the former employer have more information about the workers than a potential employer.

make the assumption that $E_\alpha < \overline{w}$, where E_α is the average productivity of an α-type in period 1. This assumption ensures that my set-up where workers prefer wage payments in period 1 to wage payments in period 2 is consistent.[16]

3.Equilibrium reward schemes

As a benchmark case, I now derive the equilibrium wage contract when firms can distinguish α-types from β-types ex ante. Then I go on to characterize the equilibrium second-best contracts in the case with ex ante asymmetric information.

3.1 Benchmark case: Full information

Consider the expected profit to firm j from employing an ι-type worker

$$\Pi_{j\iota} = E_\iota - w_{1\iota} + \int_{x_c}^{m} (x - w_{2\iota}(x))f_\iota(x)\, dx \tag{3}$$

Equilibrium is defined by; (i) there is no other contract that give workers higher utility and firms non-negative profit, given that the other firms pursue the equilibrium strategy, and (ii) for each group of workers earning the same wage in period 2, the wage has to be greater than or equal to their average expected productivity.

To find the optimal wage contracts i.e. $w_{1\iota}$ and $w_{2\iota}(x)$ we solve the following maximization problem;

[16] If I had assumed that the firm had to pay a training cost, k, to make use of a new worker the weaker assumption would have been $E_\alpha - k < \overline{w}$.

$$\max_{w_{1\iota},\, w_{2\iota}(x)} U_\iota \qquad \text{(MP.1)}$$

st. (1.a) $\Pi_{j\iota} \geq 0$

(1.b) $w_{2\iota}(x') \geq E_\iota[x \mid w_{2\iota}(x')]$ for all $x' \in X$, $\iota \in \{\alpha, \beta\}$

This yields,

Proposition 1: If the workers′ types are observable ex ante, the equilibrium contract has the following properties;

$$w_{1\iota} = E_\iota,$$
$$w_{2\iota}(x') = E_\iota[x \mid w_{2\iota}(x')]$$

i.e., each wage group is paid its average expected productivity.

Proof: (1.a) will hold with equality since a worker's utility can always be increased by lowering profit. (1.b) will hold with equality since workers prefer wage payments in period 1 to wage payments in period 2. ♦

This result is intuitive. The firm offers the the highest possible period 1 wage subject to the requirement that workers in each wage category in period 2 are paid at least their average expected marginal product. Note that because workers are risk neutral, the wage schedule in period 2 is indeterminate. An example is $w_{2\iota}(x) = x$ for all x.

Obviously, in this case inefficient terminations would serve no purpose, so $x_{c\iota} = \overline{w}$. Formally, this can be checked by substituting the optimal wage schedules given in proposition 1 into U_ι, (equation (1));

$$U_\iota = E_\iota + \frac{\int_{x_{c\iota}}^{m} x f_\iota(x)dx + F_\iota(x_{c\iota})\overline{w}}{1+r} \tag{4}$$

and differentiate with respect to x_{ci}, yielding the first-order condition

$$-f_\iota(x_{c\iota})x_{c\iota} + f_\iota(x_{c\iota})\overline{w} = 0 \tag{5}$$

Thus, of course, workers will stay in the firm if and only if their productivity in the firm is higher than their productivity outside.

3.2 Asymmetric information

When workers know more about their productivities than do firms, the first-best contract can no longer be implemented. The reason is that low productivity workers would like to masquerade as high productivity workers. Formally,

Observation 1: Workers of type β prefer an α-contract to an β-contract:

Proof: To prove this, it is sufficient to show that

$$\int_{\overline{w}}^{m} w_{2\alpha}(x) f_\beta(x)dx \geq \int_{\overline{w}}^{m} w_{2\beta}(x) f_\beta(x)dx \text{ since } w_{1\alpha} > w_{1\beta}.$$

From Proposition 1 we know that;

$$\int_{\underline{w}}^{m} w_{2\alpha}(x) f_\beta(x)dx = \int_{\underline{w}}^{m} E_\alpha[x \mid w_{2\alpha}(x)]\, f_\beta(x)dx$$

and

$$\int_{\underline{w}}^{m} w_{2\beta}(x) f_\beta(x)dx = \int_{\underline{w}}^{m} E_\beta[x \mid w_{2\beta}(x)]\, f_\beta(x)dx$$

Denote the expected productivity of β-type who in an α-contract is paid $w_{2\alpha}(x)$ as $E_\beta[x \mid w_{2\alpha}(x)]$. A β-type has the same expected productivity in period 2, regardless of whether he works in an α-contract or a β-contract, thus,

$$\int_{\underline{w}}^{m} E_\beta[x \mid w_{2\beta}(x)]\, f_\beta(x)dx = \int_{\underline{w}}^{m} E_\beta[x \mid w_{2\alpha}(x)]\, f_\beta(x)dx.$$

In Appendix 1 it is shown that MLRP implies that $E_\alpha[x \mid w_{2\alpha}(x)] \geq E_\beta[x \mid w_{2\alpha}(x)]$ completing the proof. ♦

With the first best wage schedule, an α-worker will have a higher period 1 wage than does a β-worker, and a β-worker will (in expectation) earn at least as much in the second period if he opts for an α-contract as if he opts for an β-contract.

To avoid low productivity workers selecting a contract aimed at high productivity workers, the firm can alter the contract so as to induce self-selection. I will now characterize the separating equilibrium. In Appendix 2 I show that no pooling equilibrium in pure strategies exists and derive conditions under which the separating equilibria exist.

I will (in common with most papers on self-selection) assume that once a worker have signed a contract, the firm can credibly commit not to renegotiate, due to reputation effects. Also, the worker can not leave the firm immediately after he has signed a contract. For simplicity, assume that the worker can not leave until beginning of period 2.

A separating equilibrium is defined by (i) there exists no other contract that workers prefer and give non-negative profit, given that the other firms pursue the equilibrium strategy, (ii) for each group of workers earning the same period 2 wage, the wage has to be greater than or equal to their average expected productivity, and (iii) a β-worker does not prefer an α-contract to a β-contract, and vice versa.

Now the maximization problem becomes

$$\max_{w_{1\iota},\, w_{2\iota}(x)} U_\iota = w_{1\iota} + \frac{\int_{x_{c\iota}}^{m} w_{2\iota}(x) f_\iota(x)dx + F_\iota(x_{c\iota})\overline{w}}{1+r} \qquad \text{(MP.2)}$$

$$\text{st.(2.a)} \quad \Pi_{j\iota} = E_\iota - w_{1\iota} + \int_{x_{c\iota}}^{m} (x - w_{2\iota}(x)) f_\iota \, dx \geq$$

$$\text{(2.b)} \quad w_{2\iota}(x') \geq E_\iota[x \mid w_{2\iota}(x')] \qquad \text{for all } x' \in X$$

$$\text{(2.c)} \quad U_\beta^* \geq w_{1\alpha} + \frac{\int_{x_{c\alpha}}^{m} w_{2\alpha}(x) f_\beta(x)dx + F_\beta(x_{c\alpha})\overline{w}}{1+r}$$

$$\text{(2.d)} \quad U_\alpha^* \geq w_{1\beta} + \frac{\int_{x_{c\beta}}^{m} w_{2\beta}(x) f_\alpha(x)dx + F_\alpha(x_{c\beta})\overline{w}}{1+r}$$

Constraints (2.a) and (2.b) are the same as in the symmetric information case.

Constraints (2.c) and (2.d) are the incentive compatibility constraints, and U_ι^* is the maximum utility an ι-type can get given a separating contact. Clearly, (2.a) will be binding for α-workers since their utility can always be increased without affecting incentive constraints, by lowering profit. (2.c) also binds with respect to the α-workers. Note that if it did not, their utility (for given profit) could be increased by transferring wage payments from period 2 to period 1 and/or lowering the cut-off productivity $x_{c\alpha}$.

The best contract for β-types is the same as under full information, this contract is feasible and no other feasible contract gives β-workers higher utility. It will be straightforward to show that α-workers strictly prefer the the α-contract to the β-contract, i.e. (2.d) will not be binding.

Proposition 2: In a separating equilibrium,

$$w_{1\alpha} = \min\left[E_\alpha,\ \frac{(1+r)E_\beta - E_\alpha \dfrac{f_\beta(m)}{f_\alpha(m)} + \displaystyle\int_{\overline{w}}^{x_{c\alpha}} (x-\overline{w}) f_\beta(x)dx}{1+r-\dfrac{f_\beta(m)}{f_\alpha(m)}} \right]$$

$$w_{2\alpha}(x) = x \text{ for all } x < m$$

$$w_{2\alpha}(m) = m + \frac{E_\alpha - w_{1\alpha}}{f_\alpha(m)} \quad .$$

Proof: (i) If $w_{2\alpha}(x')=w_{2\alpha}(x'')$, $x'\neq x''$ then there exists another contract where $w_{2\alpha}(x')=x'$ and $w_{2\alpha}(x'')=x''$ which gives the same profit to the firm and the same utility to α-workers, but due to the MLRP, will lower β-types utility from an α-contract, contradicting the assumption that (2.c) is binding. The fact that $w_{2\alpha}(x') \neq w_{2\alpha}(x'')$ if $x'\neq x''$, together with (2.b), imply that $w_{2\alpha}(x) \geq x$.

Let $w_{2\alpha}(m)=m'$. If $w_{2\alpha}(x')>x'$ for $x'<m$, then there exists another contract where $w_{2\alpha}(x')=x'$ and $w_{2\alpha}(m)>m'$ which gives the same profit and utility to α-workers, but, because of MLRP, will lower β-types utility from an α-contract, contradicting that (2.c) is binding. This implies that $w_{2\alpha}(x)=x$ if $x<m$ and $w_{2\alpha}(m)\geq m$.

(ii) Using the now established properties that $w_{2\alpha}(x)=x$ if $x<m$, $w_{2\alpha}(m)\geq m$, $U_\beta^* = E_\beta + \left[\int_{w}^{m} x f_\beta(x)dx + F_\beta(\overline{w})\overline{w}\right] / [1+r]$ and that (2.a) and (2.c) are binding, we get the wage schedule described in Proposition 2. $w_{1\alpha}$ can't be larger than E_α since $w_{2\alpha}(x) \geq x$, and thus it would violate the non-negative profit condition. It is straightforward to show that α-workers strictly prefer the α-contract to a β-contract, i.e. (2.d) is not binding.♦

In previous literature, the feature that only one performance level is rewarded by a wage exceeding the value of the performance has not been emphasized. The reason is presumably that in these models there are only two performance levels; pass or fail. Thus, the impression they give that a large proportion of retained workers should be overpaid is misleading. The reason why only performance level m is overpaid, if any, is that by MLRP, $f_\alpha(x)/f_\beta(x)$ is maximized at m. Thus, the proportion α-types is at its highest at m and therefore only m-performers are awarded a "prize".

There are two ways to deter low productivity workers from choosing a contract intended for the high productivity workers; either by a high cut-off, $x_{c\alpha}$, or by a low initial wage, being compensated by a higher expected wage in period 2. The latter (upwards sloping wage schedule) is the self-selection device identified by Guasch and Weiss. If self-selection is obtained solely by a high cut-off level, then $w_{1\alpha}=E_\alpha$ and $w_{2\alpha}=x$ for all x. An explicit expression for the cut-off level in this case can be found by inserting this wage schedule into (2.c) and require it to hold with equality.

This is not to say that the two self-selection devices are mutually exclusive. It is not guaranteed when $w_{2\alpha}(m) > m$, that $x_{c\alpha} = \overline{w}$. Instead, there is a trade-off between the steepness of the wage schedule and the productivity at which the workers are fired. The higher is the cut-off productivity $x_{c\alpha}$, the higher is $w_{1\alpha}$, and the lower is $w_{2\alpha}(m)$. Thus, having $x_{c\alpha}$ above $\overline{w}$ will give α-workers flatter wage schedules than if $x_{c\alpha}=\overline{w}$. If $x_{c\alpha}$ is sufficiently above $\overline{w}$, we can even have that $w_{1\alpha}$ is greater than $w_{1\beta}$, which is equal to E_β.

It is interesting to see what determines the optimal mix of these two self-selection devices. This can be done most simply by characterizing the optimal cut-off productivity, $x_{c\alpha}$. To do this, substitute the wage schedules given in Proposition 2, when $w_{1\alpha}<E_\alpha$, into U_α

$$U_\alpha = \frac{rw_{1\alpha} + \int_{x_{c\alpha}}^{m} xf_\alpha(x)dx + E_\alpha + F_\alpha(x_{c\alpha})\overline{w}}{1+r} \tag{6}$$

Differentiating, we have

$$\frac{dU_\alpha}{dx_{c\alpha}} = \frac{(x_{c\alpha} - \overline{w})\left[rf_\beta(x_{c\alpha}) - f_\alpha(x_{c\alpha})(1 + r - \dfrac{f_\beta(m)}{f_\alpha(m)})\right]}{(1+r)(1+r - \dfrac{f_\beta(m)}{f_\alpha(m)})} \tag{7}$$

Observation 2: Ex post inefficient terminations are optimal if and only if

$$\lim_{x_{\alpha} \to \overline{w}+} \frac{dU_\alpha}{dx_{c\alpha}} > 0$$

This condition is easily translated into

Proposition 3: The α-contract given in Proposition 2 entails inefficient terminations iff

$$r\left[\frac{f_\beta(\overline{w})}{f_\alpha(\overline{w})} - 1\right] > 1 - \frac{f_\beta(m)}{f_\alpha(m)} \qquad \text{(IT)}$$

In other words, $x_{c\alpha}$ will be above $\overline{w}$ if and only if condition (IT) holds. Thus, a necessary condition for involuntary separations is that $f_\beta(\overline{w})/f_\alpha(\overline{w})>1$. Involuntary separations is more likely the higher is r, $f_\beta(\overline{w})/f_\alpha(\overline{w})$ and $f_\beta(m)/f_\alpha(m)$. Optimal $x_{c\alpha}$, if above $\overline{w}$, is found where (IT) holds with equality

A high r reflects a long duration of period 1. Since firing is a way of making wage schedules flatter, this option becomes more attractive when productivities are revealed slowly. To fire workers of productivity, say, x_f, becomes less costly the lower is $f_\alpha(x_f)$ and more effective the higher is $f_\beta(x_f)$. Similarly, for a higher $f_\beta(m)/f_\alpha(m)$, the wage schedule has to be steeper to keep the β-workers from applying, and thus ex post inefficient firing becomes more attractive as a self-selection device.

It is also worth noting that Proposition 3 is derived under the assumption that reservation wages are equal for the two types of worker. If I had assumed different alternative wages, with $\overline{w}_\alpha > \overline{w}_\beta$, as e.g. in Weiss (1980) (IT) would become.

$$r\left[\frac{(x_{c\alpha} - \overline{w}_\beta)f_\beta(x_{c\alpha})}{(x_{c\alpha} - \overline{w}_\alpha)f_\alpha(x_{c\alpha})} - 1\right] > 1 - \frac{f_\beta(m)}{f_\alpha(\mathrm{m})} \qquad \text{(IT')}$$

and there would *always* be some involuntary separations (as $x_{c\alpha}$ goes to $\overline{w}_\alpha$ the denominator goes to zero).

Thus, if $\overline{w}_\alpha > \overline{w}_\beta$, or if $\overline{w}_\alpha = \overline{w}_\beta$ and the condition (IT) holds, we will have an equilibrium where a proportion of the firms will only employ β-types and the other ones only employ α-types and terminate some productive relationships. The workers can then choose between firms that give a high premium for being a top performer, and/or a higher initial wage, where requirements for staying are strong, and firms but where requirements for staying (i.e. making partners, getting tenure or passing an intermediate exam) are less harsh, but top performers less well paid and/or receive lower initial wages.

4. Rank-order contracts

The property that only the production level m should be honoured by a prize looks fragile. The very best production level will almost never be attained by any worker, and the premium would have to be enormous. If productivity is not perfectly verifiable, no amount of reputation would be enough to keep the firm from cheating if the case were to arise. In short, the contract is not convincing. Also, as mentioned in footnote 4, if the interval of possible productivities had been unbounded from above, the problem would have no solution.

A more robust contract, avoiding both these problems, would be a rank-order contract, the firm paying wages according to relative performance. This makes the firms payment in period 2 invariant to the actual productivity of the workers in period 1. Such rank-order tournaments have earlier been suggested by Carmichael (1983) and Malcomson (1984, 1986) to mitigate the problem of non-verifiability of productivity levels. The following notation will be used:

$P_\iota(i)$: the probability that an ι-type is of rank i, given an ι-contract

$P_{\beta\alpha}(i)$: the probability an β-type is of rank i, given that he opts for an α-contract

$E_\iota[x \mid i]$: the expected marginal productivity of an ι-type, given rank i and given an ι- contract.

c_ι: the highest rank that will stay in the firm given an ι-contract. (The higher is the rank the lower is the productivity).

Consider the case of observable types and the following rank-order contract. An ι-worker of rank i is paid his expected productivity given his rank i.e. $w_{2\iota}(i)=E_\iota[x \mid i]$. The firm commits to keep only workers of lower rank than, c_ι^*. Those of higher rank (the higher the rank the lower the productivity) than c_ι* gets fired, where $E_\iota[x \mid i < c_\iota^*] > \overline{w}$ and $E_\iota[x \mid i > c_\iota^*] < \overline{w}$. It is easy to see that such a rank-order contract mimics the first best contract when the number of employees is large.

To see how the rank order contracts work under asymmetric information, I have to modify the model slightly. To obtain an equilibrium with more than one firm under the assumption of asymmetric information and rank order contracts, the production technology can no longer yield constant returns to scale. Instead, I assume that there is a positive entry cost T and that production is concave in total labour input, Z, where $Z=\Sigma x$. (These features could have been introduced from the beginning at a cost of slightly more complex expressions). The reason why it is necessary with a concave production function, $g(Z)$, is that, in a separating α-contract, unit labour cost will be a decreasing function of number of people employed in the firm. The positive entry cost ensures that equilibrium is well defined.

Competition in the labour market will ensure that no wage group, will earn less than their expected marginal productivity, i.e. $w_{2\iota}(i') \geq g'(Z)E_\iota[x \mid w_{2\iota}(i')]$. The actual rank i will only be observed by insiders. But the wage schedule is observable by outsiders. Also, any single worker can communicate his wage to outsiders. To prove Observation 1 under rank-order contracts, with any finite number of employees is difficult (even though it seems very plausible). However, in the limit, when the number of employees goes to infinity, we know that the rank-order contract is equivalent to a productivity contract. Thus, at least when number of employees is large we know that β-types will mimic α-types unless the wage contracts is adjusted so as to achieve self-selection. Define U_ι^* to be the equilibrium utility an ι-type can get under a separating rank-order contract. The maximization problem for finding the best separating α-contract now becomes;[17]

$$\text{Max } \Pi_{j\alpha} = g(N_\alpha E_\alpha + N_\alpha \sum_{i=1}^{c_\alpha} P_\alpha(i) E_\alpha[x \mid i]) - N_\alpha w_{1\alpha} - N_\alpha \sum_{i=1}^{c_\alpha} P_\alpha(i) w_{2\alpha}(i) \quad \text{(MP.3)}$$

$$\text{st. (3.a) } U_\beta^* \geq w_{1\alpha} + \frac{\sum_{i=1}^{c_\alpha} P_{\beta\alpha}(i) w_{2\alpha}(i) + \sum_{i=c_\alpha+1}^{N_\alpha} P_{\beta\alpha}(i) \overline{w}}{1+r}$$

$$\text{(3.b) } U_\alpha^* \leq w_{1\alpha} + \frac{\sum_{i=1}^{c_\alpha} P_\alpha(i) w_{2\alpha}(i) + \sum_{i=c_\alpha+1}^{N_\alpha} P_\alpha(i) \overline{w}}{1+r}$$

$$\text{(3.c) } w_{2\alpha}(i') \geq g'(Y) E_\alpha[x \mid w_{2\alpha}(i')]$$

where N_α is the number of people getting employed each period in a firm offering an α-contract. Condition (3.a) is the self-selection constraint, and (3.b) is the constraint that an α-type has to be paid at least as much in expectation in the firm as in another firm in the industry. Condition (3.c) is the constraint on the period 2 wage that is implied by competition in the labour market.

(3.a) will be binding since workers prefer flatter to steeper wage schedules and low cut-off ranks to high if $c_\alpha < c_\alpha^*$. (3.b) will be binding since profit always can be increased by decreasing the wage. I have omitted the condition that an α-type should prefer an α-contract to an β-contract. As usual, it does not bind.

MLRP implies that $P_\alpha(i)/P_{\beta\alpha}(i)$ is decreasing in rank (see Appendix 3). An α-worker, for a given expected wage payment, is indifferent between a contract which give each worker different wages for different ranks, and a contract that give the same

[17]The reason why I now maximize profit instead of maximizing workers utility is that there is no longer constant returns to scale, i.e. $\Pi_{jt} = 0$ will not hold in equilibrium.

wage to more than one rank. On the other hand, β-workers strictly prefer the latter (given an α-contract). This follows from (3.c) and the fact that $P_{\alpha}(i)/P_{\beta\alpha}(i)$ is decreasing in rank. Thus, it is never optimal to give α-workers of different ranks the same wage.

Because $P_{\alpha}(i)/P_{\beta\alpha}(i)$ is decreasing in rank, condition (3.c) holds with equality for all $i<1$. Thus, we can state a parallel result to that of Proposition 2.

Proposition 4: Under rank order contracts;

$w_{2\alpha}(i') = E_{\alpha}[x \mid i=i']$

$w_{2\alpha}(1) \geq E_{\alpha}[x \mid i=1]$

The variables that are not determined are $w_{1\alpha}$, $w_{2\alpha}(1)$ and c_{α}, the cut-off rank. We can now differentiate $\Pi_{j\alpha}$ with respect to c_{α} to get (To get (8), (9) and (10) I use Proposition 4, the fact that (3.a) and (3.b) hold with equality, that $P_{\alpha}(i)=1/N_{\alpha}$, and that $g'(Y)E_{\alpha}[x \mid \mathrm{i} = c_{\alpha}] - w_{2\alpha}(c_{\alpha}) = 0$.)

$$d\Pi_{j\alpha} = \left[-N_{\alpha}\frac{dw_{1\alpha}}{dc_{\alpha}} - P_{\alpha}(1)\frac{dw_{2\alpha}}{dc_{\alpha}} \right] dc_{\alpha} \tag{8}$$

Since

$$\frac{dw_{1\alpha}}{dc_{\alpha}}\bigg|_{U_{\alpha}} = \frac{(w_{2\alpha}(c_{\alpha}) - \overline{w})(P_{\beta\alpha}(1) - P_{\beta\alpha}(c_{\alpha}))}{(1 - P_{\beta\alpha}(1)N_{\alpha})(1+r)} \tag{9}$$

and

$$\frac{dw_{2\alpha}(1)}{dc_{\alpha}}\bigg|_{U_{\alpha}} = [w_{2\alpha}(c_{\alpha}) - \overline{w}]\left[\frac{N_{\alpha}(P_{\beta\alpha}(c_{\alpha}) - P_{\beta\alpha}(1))}{1 - P_{\beta\alpha}(1)N_{\alpha}} - 1\right] \tag{10}$$

we have that

$$\frac{d\Pi_{j\alpha}}{dc_\alpha}\Big|_{U_\alpha} = [w_{2\alpha}(c_\alpha) - \overline{w}]\left[1 - \frac{rN_\alpha(P_{\beta\alpha}(c_\alpha) - P_{\beta\alpha}(1))}{(1+r)\,(1 - P_{\beta\alpha}(1)N_\alpha)}\right] \qquad (11)$$

From this, it is easy to verify, the following result:

Proposition 3*: Under a rank-order contract, inefficient terminations are optimal if and only if $\frac{d\Pi_{j\alpha}}{dc_\alpha}\Big|_{U_\alpha} < 0$ at $c_\alpha = c_\alpha^*$ <=>

$$r\left[\frac{P_{\beta\alpha}(c_\alpha^*)}{P_\alpha(c_\alpha^*)} - 1\right] > 1 - \frac{P_{\beta\alpha}(1)}{P_\alpha(1)} \quad .$$

Proof: Use (11) and the fact that $P_\alpha(i) = 1/N_\alpha$. ♦

This condition for termination of productive relationships is almost the same as (IT).

As in the previous section, where wages where function of productivity (and not of rank), we have the possibility that self-selection is achieved solely through ex post inefficient separation. But when it is achieved at least partially through steepening the wage schedule, the wage will be equal to average marginal productivity in period 2 except for the first rank who earns a prize. The obtained wage schedule is thus similar to the one obtained in Rosen´s (1986) study of elimination career ladders. In his model, the large first prize is nessesary to induce contestants to keep their effort at a high level, also late in the game. What I have shown here is that this wage schedule, which have been used as an explanation of top-executives´ high wages, does not necessarily respond to a moral hazard problem but could result from adverse selection as well.

Can we use this framework to say anything about the relationship between firm size and workers´abilities? This is the question to which I now turn. Suppose that the production technology differs across firms. Then, the number of employees, *N*, would differ across firms as well. Let's say that $g=g(Y,K)$ where K is capital stock, and that g is increasing and concave both in Y and K. Clearly, a larger firm can have more ranks. Also, $P_{\alpha}(1)/P_{\beta\alpha}(1)$ will be greater the larger is the firm. My conjecture is therefore that larger firms will have a comparative advantage in employing α-types. The argument is that more ranks allow the payment in period 2 gets to be more precise, discouraging β-workers, because of the MLRP. This is a new explanation for the positive correlation of wages and ability with firm size. For empirical evidence and a survey of other explanations see Brown and Medoff (1989).

The model can also be used to study how top wages will change with changes in the product demand during the contract period. Recall that all period 2 wages except the top wage are set to their minimum, and the period 1 wage to its maximum, given zero profit (on the margin) and the self-selection constraint. Thus, the period 1 wage and all wages below the top wage in period 2 will change when demand conditions change. If we assume that the state of demand is not verifiable, the top wage (which includes prize-money) can not be state dependent, since the firm would then always have an incentive to claim a bad demand state (or the state where the top wage is the smallest). Consequently, the top wage must be insensitive to the state of demand, and hence the difference between the the top-wage and "ordinary wages" will be larger in bad demand states.

5. Final remarks

The three most important results of this investigation is that 1) termination of productive relationships may be used as a self-selection device, 2) that only the very best worker should be overpaid, and 3) that large firms have an advantage in attracting able workers. The second result hinges on the fact that workers are risk neutral. Had they been risk averse, the level above which workers were overpaid may be dependent on the degree of risk aversion.

In conclusion, let me now relate the first result more closely to the existing literature. Following the work of Weiss (1980), and Stiglitz and Weiss (1983), there have been a number of attempts trying to explain why productive labour market relationships are sometimes terminated. The original ideas were twofold. In Weiss (1980) the idea was that replying to a recession with a wage reduction would result in the most productive workers leaving the firm. Thus it would be better to keep the wage high and lay off workers randomly (*the adverse selection argument*). The second mechanism, presented by Stiglitz and Weiss (1983), is that the threat of termination can induce higher effort by workers in a model where the employer can observe production but not effort (*the effort inducement argument).* Variations of this latter idea are recently presented by Kahn and Huberman (1988), Kahn and Mookherjee (1988) and Waldman (1990)[18]. In their models, it is crucial that productivity is not verifiable. It is exactly the provision that workers with a sub-standard productivity are not retained which ensures that the employer does not falsely claim that a worker has a low productivity. However, the robustness of this type of argument has been seriously questioned by Mookherjee (1986). The simple Stiglitz-Weiss model creates involuntary retentions rather than involuntary layoffs for a range of plausible parameter values, and the more sophisticated model with two-sided uncertainty hinges on the assumption that the firm cannot use rank order tournaments. If the firm could commit to promoting a certain proportion of the work force, according to relative productivity, this would efficiently solve the incentive problems on both sides.[19] [20]

---

[18] Other authors who have considered up-or-out contracts are O'Flanery and Siow (1989). They are looking at hiring/firing policies when only experienced workers above a certain quality can supervise a new worker.

[19] This critique can be circumvented by assuming that the firm must incur a small cost in order to observe the workers' productivity, and that the workers cannot know whether or not the firm promotes randomly. Personally, I am unimpressed by this counter argument: in most cases firms learn something about workers' abilities whether they want to or not. Also, a firm with a random promotion policy would quickly earn a bad reputation.

[20]If application fees are allowed in the Stigliz-Weiss analysis the involuntary terminations would disappear. The reason why firing of productive workers can be rational is that workers are earning rents,

My model is immune to the critiques levelled against the models founded on incentive effects. On the other hand, the model's applicability is restricted to environments where the time between hiring and re-evaluation is considerable. Thus, I have presented a new explanation of ex post inefficient tenure and promotion decisions in places where the true productivity is learnt slowly - whereas it is inconsistent with inefficient terminations occurring shortly after hiring and (random) firing of highly productive workers. Therefore, it is probably better seen as an attempt to explain up-or-out contracts in certain professions like art, science and law,[21] than as a microeconomic foundation for Keynesian unemployment.

Although Guasch and Weiss (1981 p.278) hinted at the idea that termination threats may supplement wage profiles in sorting the work force, these authors never substantiated the argument and, as it stands it is incorrect. In this article, Guasch and Weiss analyse self-selection contracts both under the assumption of risk neutral workers and under the assumption that workers are risk averse, and they come to the conclusion that self-selection will always be optimal when workers are risk neutral, but not necessarily when they are risk averse. Throughout their paper they do however *assume* that self-selection implies that only those who pass the test will be hired. I show in Appendix 4 that if the workers are risk neutral, in their model all workers should be hired ex post. Only above a certain level of risk aversion can it be optimal to engage in inefficient firing. Again, of course, the intuition is that the steep wage schedule is costly - now due to workers' risk aversion - and it may be profitable to introduce a second self-selection device. But even with risk-aversion, it is necessary to have different reservation wages in order to produce the termination results in Guasch and Weiss' set-up. That was not necessary here.

Another important difference between the present model and Guasch and Weiss

due to efficiency wages. Thus, the threat to terminate induces higher effort. However effort depend only on the difference between wage if successful and wage if failing. With risk neutral workers an application fee makes it possible for the firm to widen this difference without decreasing profit.

[21] Spurr (1987) reports that it takes an average of seven years before a trainee in a law firm is considered for being promoted to partner. Around 30% are then accepted for partnership.

(1981) is that I allow for competition in the labour and product markets. In Guasch and Weiss (1980) and (1982), there is also competition in the labour and product markets, but ex post inefficient terminations are not optimal (even with different reservation wages). The reason terminations may be optimal in the present model is that there is asymmetric information between the firm which fire a worker and a firm which hire the same worker. Since a fired worker who is productive cannot be distinguished from other less productive workers, there is a net loss associated with having to leave the firm.

It is a key feature of my model that productivity as well as market conditions are learnt by both parties before layoffs. Thus, there is no asymmetry of information when the separation decision is made. Moreover, the results would go through even with costless renegotiation. Much other work on ex post inefficient layoffs, like that of Chari (1983), Hall and Lazear (1984), and Moore (1985) rely either on asymmetric information or costly renegotiation at the termination stage.[22] Of these models only the latter allow for heterogeneous workers. However, as pointed out by Moore (1985) and Mookherjee (1986), they also tend to produce involuntary retentions rather than layoffs.

Finally, I would like to apologize for this article adding to the already much too low facts/theory rate in this particular field. In order better to assess the relative merits of various theories, micro level data on inefficient terminations would be very welcome. The only indisputable example I am aware of where terminations of productive relationships have been used as part of a conscious strategy to increase overall efficiency, is from the Mongolian army under Ghengis Khan. There, a number of Mongolian soldiers would be executed if one of their leaders was lost in fight. The more prominent the leader, the higher was the number of soldiers punished. Although this practice was obviously ex post inoptimal, it seems to have been ex ante efficient.

[22] Here, I will not go into a detailed description of these models. The set of environments for which they are appropriate is sufficiently different from that studied here, that these theories are complementary to rather than competing with mine.

Few armies have been more successful.

Appendix 1[23]

I will here show that MLRP implies that $E_{\alpha}[x \mid w_{2\alpha}(x)] \geq E_{\beta}[x \mid w_{2\alpha}(x)]$

Assumptions;

1) $\dfrac{f_{\alpha}(x)}{f_{\beta}(x)}$ increasing in x (MLRP). $\qquad x \in R^{+} \qquad$ (A1.1)

2) Let A be a subset of R^{+} such that $\int_{A} f_{\alpha}(x)dx > 0$.

Put,

$$E_{\iota}[x \mid x \in A] = \frac{\int_{A} x f_{\iota}(x)dx}{\int_{A} f_{\iota}(x)dx} \tag{A1.2}$$

Proposition A1.1: Given assumption 1) and 2);

$$E_{\alpha}[x \mid x \in A] > E_{\beta}[x \mid x \in A]. \tag{A1.3}$$

To prove proposition A1.1 we will use the following lemma, which is proven afterwards;

Lemma A1.1: Let $h(x)$, $-\infty < x < \infty$, be a probability density , i.e. $h(x) \geq 0$ and $\int_{-\infty}^{\infty} h(x)dx = 1$. Then it holds for $u(x)$ and $v(x)$, $-\infty < x < \infty$, which both are non-negative and increasing that;

$$\int_{-\infty}^{\infty} u(x)v(x)h(x)dx \geq \left(\int_{-\infty}^{\infty} u(x)h(x)dx\right)\left(\int_{-\infty}^{\infty} v(x)h(x)dx\right).$$

[23] I am grateful to Bengt Rosén for help with this appendix and with appendix 3.

Proof of Proposition A1.1: By (A1.2) we have that

$$\int_A f_\alpha(x)dx\, E_\alpha[x \mid x \in A] = \int_A x f_\alpha(x)dx = \tag{A1.4}$$

$$\int_A x \left(\int_A f_\beta(x)dx\right)\frac{f_\alpha(x)}{f_\beta(x)} \frac{f_\beta(x)}{\int_A f_\beta(x)dx} dx \;\geq \tag{A1.5}$$

Use the lemma A1.1 with

$$h(x) = \frac{f_\beta(x)}{\int_A f_\beta(x)dx} \quad \text{for } x \in A, =0 \text{ for } x \notin A$$

$$v(x) = \left(\int_A f_\beta(x)dx\right)\frac{f_\alpha(x)}{f_\beta(x)} \qquad \text{increasing by (A1.1)}$$

$$u(x) = x \qquad \text{increasing}$$

$$\geq \int_A x f_\beta(x)dx \int_A f_\alpha(x)dx \Big/ \int_A f_\beta(x)dx = \int_A f_\alpha(x)dx\, E_\beta[x \mid x \in A] \tag{A1.6}$$

The claim in proposition A1.1 now follows from (A1.4) and (A1.6)

Proof of Lemma A1.1:

Step 1: It is readily checked that lemma A1.1 holds for one-step functions (jumping from zero to one).

Step 2: It is readily checked that if $u(x)$ and $v(x)$ are linear combinations of one-step functions ($u(x) = \sum_{\nu=1}^{N} a_\nu u_\nu(x)$, $v(x) = \sum_{\varphi=1}^{N} b_\varphi v_\varphi(x)$) lemma A1.1 holds.

Step 3: The claim in lemma A1.1 now follows from that every non-negative, increasing function can be obtained as limes of a serie of increasing one-step functions.

Appendix 2

Here I prove, in oppositing order, that a separating equilibrium in pure strategies exist for a sufficient high proportion of α-workers, and (ii) that there is no pooling equilibrium.

Expected profit given a pooling contract is;

$$\Pi_{jp} = \mu E_\alpha + (1-\mu)E_\beta - w_{1p} + \int_{\underline{w}}^{m} (x - w_{2p}(x))\, (\mu f_\alpha(x) + (1-\mu) f_\beta(x))\, dx \quad \text{(A2.1)}$$

where μ is the proportion of α-types in the economy.

Proposition A2.1: No pooling equilibria in pure strategies exist.

Proof: Denote expected profit from an ι-type in a poling contract $\Pi_{p\iota}$. If $\Pi_{p\alpha}>0$ and $\Pi_{p\beta}<0$, β-types can be profitably deterred from applying while α-types are still applying. This is brought about lowering w_{1p} by κ and increasing $w_{2p}(m)$ by $\kappa(1+r)/f_\alpha(m)$, leaving α-types utility unchanged but lowering that of the β-types. If $\Pi_{p\alpha}<0$ and $\Pi_{p\beta}>0$, it must be true that a wage group is paid more than their expected average productivity in period 2 and that α-workers have a higher probability of belonging to this wage group than do β-workers. Denote these probabilities q_α and q_β respectively. α-workers can then be profitably deterred from applying, while β-types still apply, if the relevant wage is lowered by κ and w_{1p} is increased by $q_\beta \kappa/(1+r)$. If $\Pi_{p\alpha}=\Pi_{p\beta}=0$, at least β-types will prefer the separating wage schedule to the pooling, since this is the best wage schedule for β-types given $\Pi_{p\beta}=0$. If instead of assuming productivity dependent wages in period 2, we had assumed rank order payments the proof is analogous. ♦

If however the proportion, μ, of β-types in the economy is so small that $1-\mu$

$< rf_\alpha(m)/(1+r)[f_\alpha(m) + f_\beta(m)]$, it might be profitable for a firm to deviate from the separating equilibrium. We run into the standard difficulties with non-existence of equilibrium in pure strategies when the proportion of "bad" agents in the economy is small, see e.g. Rotschild and Stiglitz (1976). On the other hand, for $1-\mu > rf_\alpha(m)/(1+r)[f_\alpha(m) + f_\beta(m)]$, no deviations from the separating equilibrium are profitable and hence an separating equilibrium exist. This can be shown as follows. Consider the pooling contract

$$W_p = \{\ w_{1p} = w_{1\alpha} + \frac{(w_{2\alpha}(m) - m) f_\alpha(m) - \int_{\overline{w}}^{x_{c\alpha}} (x - \overline{w}) f_\alpha(x) dx}{1+r}$$

$$w_{2p}(x) = x,\ x_{cp} = \overline{w}\ \} \tag{A2.2}$$

where $w_{1\alpha}$, $w_{2\alpha}(m)$ and $x_{c\alpha}$ takes the values of the best separating α-contract. Define W_t to be the best separating wage schedule for an α-type. α-workers are indifferent between W_p and W_α , and β-workers strictly prefer W_p to W_β. Clearly, W_p is the best possible pooling contract if it is not possible to increase profit by transferring wage payments from period 1 to period 2 (reducing the period 2 wage is impossible because of competition in the labour market). We have that

$$\frac{d\Pi_{jp}}{dw_{2p}(x)}\bigg|_{U_\alpha} = -\mu f_\alpha(x) - (1-\mu) + \frac{f_\alpha(x)}{(1+r)} =$$

$$[f_\alpha(x) - f_\beta(x)]\left[\frac{\frac{f_\alpha(x)}{1+r} - f_\beta(x)}{f_\alpha(x) f_\beta(x)} - \mu\right] \tag{A2.3}$$

If $f_\alpha(x) - f_\beta(x) < 0$, then $d\Pi_{jp} / dw_{2p}(x) < 0$. If $f_\alpha(x) - f_\beta(x) > 0$, then $d\Pi_{jp} / dw_{2p}(x) < 0$ iff $[f_\alpha(x)/(1+r) - f_\beta(x)] \ / \ [f_\alpha(x) - f_\beta(x)] < \mu$. The expression $[f_\alpha(x)/(1+r) - f_\beta(x)]\ /\ [f_\alpha(x) - f_\beta(x)]$ is increasing in $f_\alpha(x)/f_\beta(x)$. Therefore, $d\Pi_{jp}/dw_{2p}(x) < 0$ for all x iff

$$\frac{\frac{f_\alpha(m)}{1+r} - f_\beta(m)}{f_\alpha(m) - f_\beta(m)} < \mu \tag{A2.4}$$

Thus if (A2.4) holds, W_p is the best pooling wage schedule. It is straightforward to show that there exist $\mu < 1$, such that W_p is positive.

If $[f_\alpha(m)/(1+r) - f_\beta(m)]\ /\ [f_\alpha(m) - f_\beta(m)] > \mu$, then it will be profitable to raise some second period wage and lower w_{1p} until β-types are indifferent between the pooling contract and the separating contract. When both α-types and β-types are indifferent between the pooling contract and their first best separating contract, no further increases in the second period wage will be profitable since workers prefer flatter to steeper wage schedules. To see that this new pooling contract will be dominated by a separating contract, recall that W_β is the best wage contract for β-types given non-negative profit on β-types and that W_α is the best wage contract for α-types given that β-types are indifferent between W_β and W_α. Since W_α and W_β are not the same, this pooling contract will give negative profit.

Appendix 3

Here I will show that $\frac{f_\alpha(x)}{f_\beta(x)}$ increasing in x implies that $\frac{P_\alpha(i)}{P_{\beta\alpha}(i)}$ is increasing in rank i.

Assumptions:

1) $\frac{f_\alpha(x)}{f_\beta(x)}$ increasing in x (MLRP).

2) N α-workers gets a (random) productivity result; $\mathbf{X}_1, \mathbf{X}_2, ..., \mathbf{X}_N$ by the distribution $f_\alpha(x)$ and one β-worker gets a productivity result $\mathbf{Y}$ by the $f_\beta(x)$ distribution.

3) Let i_β = the rank for a β-worker when one ranks the values $\mathbf{X}_1, \mathbf{X}_2, ..., \mathbf{X}_N$, $\mathbf{Y}$ by size, staring with the largest.

The possible values of i_β are $1, 2, ..., N+1$.

Proposition A3.1: $P_{\beta\alpha}(1) < P_{\beta\alpha}(2) < ... < P_{\beta\alpha}(N+1)$.

Corollary A3.1: $\frac{P_\alpha(i)}{P_{\beta\alpha}(i)}$ is decreasing in rank, i.

$i=1, 2, ..., N+1$.

Proof of Corollary A1 (given Proposition A1):

$P_{\beta\alpha}(i) + NP_\alpha(i) = 1$. Thus, if $P_{\beta\alpha}(i)$ is increasing in rank i, $P_\alpha(i)$ is decreasing in rank i, implying that $\frac{P_\alpha(i)}{P_{\beta\alpha}(i)}$ is decreasing in rank i.♦

Proof of Proposition A3.1:

a) given $\mathbf{Y} = y$; $\mathbf{X}_1, \mathbf{X}_2, ..., \mathbf{X}_N$ are independent stochastic variables with the distribution $f_\alpha(x)$.

b) Given $\mathbf{Y} = y$, the β-type is of rank i if $N-i+1$ of $\mathbf{X}_1, \mathbf{X}_2, ..., \mathbf{X}_N$ are less than y and $i-1$ are greater than y.

$$P_{\beta\alpha}(i \mid \mathbf{Y}=y) = \binom{N}{N-i+1}\left(\int_{-\infty}^{y} f_\alpha(x)dx\right)^{N-i+1}\left(\int_{y}^{-\infty} f_\alpha(x)dx\right)^{i-1} \tag{A3.1}$$

c) $P_{\beta\alpha}(i) = \int_{-\infty}^{\infty} P_{\beta\alpha}(i \mid \mathbf{Y}=y) f_\beta(y)dy =$ (by (A3.1)) $=$

$$\binom{N}{N-i+1}\int_{-\infty}^{\infty}\left(\int_{-\infty}^{y} f_\alpha(x)dx\right)^{N-i+1}\left(\int_{y}^{\infty} f_\alpha(x)dx\right)^{i-1} f_\beta(y)dy \tag{A3.2}$$

We can rewrite (A3.2) on the following form;

$$P_{\beta\alpha}(i) = \binom{N}{N-i+1}\int_{-\infty}^{\infty} h(y)G(y)dy \tag{A3.3}$$

where

$$h(y) = f_\alpha(y)\left(\int_{-\infty}^{y} f_\alpha(x)dx\right)^{N-i+1} = \frac{d}{dy}H(y) \tag{A3.4}$$

where

$$H(y) = \frac{1}{N-i+2}\left(\int_{-\infty}^{y} f_\alpha(x)dx\right)^{N-i+2} \tag{A3.5}$$

and

$$G(y) = \frac{f_\beta(y)}{f_\alpha(y)}\left(\int_{y}^{\infty} f_\alpha(x)dx\right)^{i-1} \tag{A3.6}$$

Partially integrating (A3.3) we get;

$$P_{\beta\alpha}(i) =$$

$$\binom{N}{N-i+1}\left\{[H(y)G(y)]_{-\infty}^{\infty} - \int_{-\infty}^{\infty} H(y)\frac{d}{dy}G(y)dy\right\} = -\binom{N}{N-i+1}\int_{-\infty}^{\infty} H(y)\frac{d}{dy}G(y)dy \tag{A3.7}$$

From (A3.6) we get

$$\frac{d}{dy}(G(y)) = -(i-1)f_{\beta}(y)\left(\int_{y}^{\infty} f_{\alpha}(x)dx\right)^{i-2} + q(y)\left(\int_{y}^{\infty} f_{\alpha}(x)dx\right)^{i-1} \tag{A3.8}$$

where

$$q(y) = \frac{d}{dy}\left(\frac{f_{\beta}(y)}{f_{\alpha}(y)}\right) \tag{A3.9}$$

$q(y) < 0$ by assumption 1).

Inserting (A3.5) and (A3.8) into (A3.7) we get

$$P_{\beta\alpha}(i) = \binom{N}{N-i+1}\frac{i-1}{N-i+2}\int_{-\infty}^{\infty}\left(\int_{-\infty}^{y} f_{\alpha}(x)dx\right)^{N-i+2}\left(\int_{y}^{\infty} f_{\alpha}(x)dx\right)^{i-2} f_{\beta}(y)dy$$

$$-\binom{N}{N-i+1}\frac{1}{N-i+2}\int_{-\infty}^{\infty}\left(\int_{-\infty}^{y} f_{\alpha}(x)dx\right)^{N-i+2}\left(\int_{y}^{\infty} f_{\alpha}(x)dx\right)^{i-1} q(y)dy \tag{A3.10}$$

The first term in (A3.10) is equal to $P_{\beta\alpha}(i-1)$, thus

$$P_{\beta\alpha}(i) = P_{\beta\alpha}(i-1) - D \qquad \text{where } D < 0. \tag{A3.11}$$

(A3.11) proves proposition A1.♦

Appendix 4.

I now show that the claim made by Guasch and Weiss (1981,p.278) that involuntary separations occur in their model is incorrect. I stick to their terminology.

There are two types of workers, type 1 and type 2. Type 1 is more productive than type 2. Workers know their own type but prospective employers don't. The firm can charge an application fee and then test the applicants. The test is such that type 1 have a higher probability of passing than type 2. To discourage type 2 workers from applying, they condition the wage/hiring decision on the test result. The test is costly to the firm. There is only one firm, which faces a concave production function. Price is given and independent of production. Guasch and Weiss consider both the case of risk neutral and the case of risk averse workers and come to the conclusion that self-selection will always be optimal with risk neutral workers but not necessarily with risk averse.

Throughout the paper they *assume* that in self-selection contracts only those who pass the test will be hired. However, if the workers are risk neutral it is never optimal to do so. The optimal contract is to give those who fail the test their alternative wage and those who pass their alternative wage plus the application fee divided by the probability of passing. The application fee can still be chosen so that the low productivity types don't want to apply, and this scheme avoids turning able workers away because of a bad test result.

For a sufficiently high degree of risk aversion it might be optimal only to employ those who passed the test, but the authors do not consider the conditions under which this is so, they just - wrongly - assume that it always holds. To hire "failures" when employees are risk averse is more profitable the less risk averse they are, the higher is the firm´s testing cost and the smaller is the difference in the reservation wages. Formally:

Notation

A= The cost of the test per applicant tested

L= the number of workers tested by the firm.

C= the fee charged by the firm to take the test.

P_i= the probability of a type i worker passing the test

Q_i= the expected labour input of type i workers

w_i= the present value of the acceptance wage of a type i worker

w^*= the present value of the wage offered by the firm to all workers passing the test.

Assumptions: $Q_1>Q_2$, $w_1>w_2$ and $P_1>P_2$.

Minimum cost per efficiency unit of labour with a separating self-selection contract and when only those who pass the test are hired is

$$E^* = \frac{A + w_1 P_1}{P_1 Q_1} \tag{A4.1}$$

If, however, those who fail the test are hired to a wage equal to w_1, the minimum cost per efficiency unit will be

$$E^{**} = \frac{A + w_1}{Q_1} \tag{A4.2}$$

It is easily seen that $E^{**}<E^*$. Thus, it is profitable also to hire those who fail the test. Now it is straightforward to show that E^{**} can be achieved without violating either the participation constraint

$$P_1 w^* + (1 - P_1)w_1 - C = w_1 \tag{A4.3}$$

or the self-selection constraint

$$P_2(w^*) + (1 - P_2)w_1 - C < w_2 \qquad \text{(A4.4)}$$

The solution to this problem is any pair (C,w^*) satisfying

$$\text{(I) } w^* = w_1 + \frac{C}{P_1}$$

$$\text{(II) } C > \frac{(w_1 - w_2)(1 - P_2)P_1}{P_1 - P_2}$$

Obviously, since $E^{**}>E^*$, there exist some positive degree of risk aversion for which hiring of failures is profitable. Most importantly, if reservation wages are the same for both types of worker, inefficient terminations is never optimal, regardless of risk aversion. The reason is that workers who fail the test are indifferent between being retained or fired (since they are retained at their reservation wage), and so terminations can add nothing to what is already achieved by the upward sloping wage schedule.

References

Brown, Charles and James Medoff, (1989): The Employer Size-Wage Effect, *Journal of Political Economy*, 97, October, 1027-59.

Carmichael, Lorne, (1983), Firm Specific Human Capital and Promotion Ladders, *Rand Journal of Economics*, 14, Spring, 251-258.

Chari,V.V., (1983): Involuntary Unemployment and Implicit Contracts, *Quarterly Journal of Economics*, 98, Supplement, 107-122.

Greenwald, Bruce C., (1986): Adverse Selection in the Labour Market, *Review of Economic Studies*, 53, July, 325-347.

Guasch, J. Luis and Andrew Weiss, (1980): Wages as a Sorting Mechanism in Competitive Markets with Asymmetric Information: A Theory of Testing, *Review of Economic Studies*, 47, July, 653-664.

Guasch, J. Luis and Andrew Weiss, (1981): Self-Selection in the Labor Market, *American Economic Review*, 71, June, 275-284.

Guasch, J. Luis and Andrew Weiss, (1982): An Equilibrium Analysis of Wage-Productivity Gaps, *Review of Economic Studies*, 49, October, 485-497.

Hall, Robert. E., and Edward P. Lazear, (1984): The Excess Sensitivity of Layoffs and Quits to Demand, *Journal of Labour Economics*, 2, April, 233-257.

Jappelli, Tullio, (1990): Who Is Credit Constrained in the U.S.Economy?, *Quarterly Journal of Economics*, 105, February , 219-234.

Kahn, Charles and Gur Huberman, (1988): Two-sided Uncertainty and "Up-or-out" Contracts, *Journal of Labour Economics*, 6, October, 423-444.

Malcomson, James M., (1984): Work Incentives, Hierarchy, and Internal Labor Markets, *Journal of Political Economy*, 92, June, 486-507.

Malcomson, James M., (1986): Rank-Order Contracts for a Principal with Many Agents, *Review of Economic studies*, 53, October, 807-817.

Mayfield, Scott E., (1990): Borrowing Constraints Under Secured Loan Contracts, Boston College, Department of Economics, Working Paper No. 189.

Milgrom, Paul and Sharon Oster, (1987): Job Discrimination, Market Forces and the Invisibility Hypothesis, *Quarterly Journal of Economics*, 102, August, 453-476.

Mookherjee, Dilip, (1986): Involuntary Unemployment and Worker Moral Hazard, *Review of Economic Studies*, 53, October, 739-754.

Moore, John, (1985): Optimal Labour Contracts When Workers Have a Variety of Privately Observed Reservation Wages, *Review of Economic Studies*, 52, January, 37-67.

O'Flaherty, Brendan and Aloysius Siow, (1989): On the Job Screening, Up or Out Rules, and Firm Growth, University of Toronto, Department of Economics and Institute for Policy Analysis, Working Paper No. 8913,

Rosen, Shervin, (1986): Prizes and Incentives in Elimination Tournaments, *American Economic Review*, 76, September, 701-715.

Rotschild, Michael and Joseph, E. Stiglitz, (1976): Equilibrium in Competitive Insurance Markets: An Essay in the Economics of Imperfect Information, *Quarterly Journal of Economics*, 90, November , 629-649.

Stiglitz, Joseph, E. and Andrew Weiss, (1983): Incentive Effects of Terminations: Application to the Credit and Labor Markets, *American Economic Review*, 73, December, 912-927

Spurr, Stephen, J., (1987): How the Market Solves an Assignment Problem: The Matching of Lawyers with Legal Claims, *Journal of Labour Economics*, 5, October, 502-532

Waldman, Michael, (1984): Job Assignments, Signalling and Efficiency, *Rand Journal of Economics*,15, Summer, 255-267.

Waldman, Michael, (1990): Up-or-Out Contracts: A Signaling Perspective, *Journal of Labour Economics*, 8, April, 230-250.

Weiss, Andrew, (1980): Job Queues and Layoffs in Labor Markets with Flexible Wages, *Journal of Political Economy*, June, 88, 526-538.

Chapter 3: A MATCHING MODEL OF DISCRIMINATION[24]

1. Introduction

Numerous studies have shown that women earn less than men and that black men earn less than white men. Moreover, in an overwhelming majority of these studies, the differential is not fully accounted for by differences in pre labour market conditions, such as education. According to Blau and Ferber (1987) there is typically an unexplained wage differential of 20% or more.

From a theoretical point of view, discrimination has proved to be an elusive phenomenon. The most cited work is Becker's (1971) taste models. There, the lower wages for women and blacks result from the employer, co-workers or consumers requiring a premium to interacting with these groups rather than a white man. However, as pointed out by Arrow (1973) and Cain (1986), Becker's model does not explain discrimination in the long run. It should disappear either through segregation - people belonging to a particular group interacting with each other - or, in the case of employer discrimination, through entry by employers without prejudices. The latter firms would be more profitable and hence drive the discriminating firms out of the market.

Another strand of literature, often referred to under the label of "statistical discrimination," has been developed by Lundberg and Startz (1983), following early work by Phelps (1972). These authors argue that workers whose expected productivity is the same may be treated differently if one group's productivity is harder to measure. The impact on expected productivity, as measured by a potential employer, of a given investment in (invisible) human capital is smaller for the group whose ability is harder to measure, and therefore they will invest less. Also, as pointed out by

[24]Valuable comments and suggestions have been received from Daron Acemoglu, Tore Ellingsen, Alan Manning, Margret Meyer, John Moore and Asbjørn Rødseth.

Rothschild and Stiglitz (1982), if employers' payoff depends on how well they can match workers to tasks, the group whose productivity is easier to measure will be preferred even absent concerns for human capital investment. Other theories tracing discrimination to true differences in worker characteristics, are those of Bulow and Summers (1985) and Milgrom and Oster (1987). In Bulow and Summers' model workers with statistically lower turnover rates are better paid,[25] whereas the valuable characteristic emphasized by Milgrom and Oster is a worker's visibility. Although visibility is not valued per se, market forces imply that a visible worker will be assigned in better accordance with his productivity and hence have a higher wage. The basic message in all these theories is that unequal treatment can be traced to some underlying characteristic, different from productivity itself. Thus, there is no discrimination in the sense that groups with exactly the same (statistical) properties are treated differently.

A stricter definition of discrimination requires that workers who are identical in all relevant respects, including turnover, measurability and visibility, obtain different utilities. I will refer to this as strong discrimination. The first formal model capturing this possibility is Spence (1976). There, workers signal their productivity through education choice. The model has multiple equilibria, and it is possible therefore that the return to education will differ across ex ante identical groups. However, it has been argued that not all the equilibria in Spence's model are equally reasonable. Arguments that one particular equilibrium (viz; the Pareto-dominating separating equililibrium) is more reasonable than all of the others, have been put forward with increasing degree of sophistication by Riley (1979), Cho and Kreps (1987) and Nöldeke and van Damme (1990) respectively. If this argument is accepted, it is easily verified that the model loses its force in explaining discrimination. E.g., the "intuitive criterion" of Cho and Kreps (1987) eliminates the possibility that one group, women in Spence's paper, does not earn a return to education (are "pooled") whereas another identical group, men,

[25] However, the premise that women have higher turnover than men, has been disputed. Even though women's quit rates are higher than men's, women's quit rates in any given type of occupation are not. Studies by Viscusi (1980), Blau and Kahn (1981) and by Harber, Lamas and Green (1983) demonstrate that, when wage differentials are controlled for, there is no difference in turnover due to gender.

does earn such a return (are "separated").

An interesting early analysis of discrimination in a model with job search and incomplete information is McCall (1972). His model has the weakness, however, that firms' beliefs about the of workers' abilities are not correct, so discrimination is not an equilibrium outcome. Nevertheless, McCall deserves much credit for directing attention away from friction-free Walrasian models to markets with a more realistic description of interaction between firms and workers.

Perhaps the most convincing current theory of strong discrimination is Akerlof (1985). Here, a significant subset of the consumers are discriminating (in Becker's sense), and in a search framework, sellers employing black workers are at a disadvantage. This drives down black workers' wages. A similar point is made by Borjas and Bronars (1989) in a model of self-employment.

The present paper neglects the consumer side altogether, focussing exclusively on labour market interactions. In addition, no employers have prejudices against any group of labour. Groups differ only by their index, i.e. their sex, colour, looks or other irrelevant label. Still, it is shown that, for any commonly known labelling of workers, there exist equilibria entailing discrimination. Men can have a higher lifetime utility than women, white can be better off than black, blue-eyed better off than brown-eyed, or vice versa. Even more surprising, only equilibria involving discrimination are stable.

Key assumptions of the model are:

(i) workers search for jobs, firms search for workers,

(ii) workers are not equally able at every job (but have the same average abilities)

(iii) firms can not observe a workers productivity perfectly until after he/she has been employed for some time,

(iv) workers have some private information about their productivity in any specific job before they start working, and

(v) both profit and wages increase as the quality of the match increases (the productivity of the worker in that particular job).

In short, the idea of the paper can be described as follows. When a group of workers finds it difficult to get a job, workers belonging to this group will respond by applying for more jobs. As they start applying for jobs that they are not particularly well suited for, the average quality of applicants from this group is reduced, which in turn makes firms more reluctant to hire them. Thus, if some firm discriminates against this group (black, women or long term unemployed), it is rational for every other firm to do so too. Note, that no firm need to have a taste for discrimination in Becker's sense. Discrimination arises in a non-cooperative (Nash) equilibrium even if everybody knows that the groups are inherently equal.

The model is shown to be consistent with differential wages as well as unemployment rates. Furthermore, it is consistent both with discriminated groups being less well matched and being allocated to less attractive jobs. Another prediction is that observed skills of discriminated workers have to be significantly higher than those of other workers in order for the former to be preferred by an employer.

The theory offers an argument for affirmative action at the hiring stage in the public sector. Favouring the discriminated group in this way will move the whole economy to a less discriminating equilibrium.

We know that exit rates from unemployment falls with duration of unemployment. The model also offers a new explanation of this phenomenon.

The paper is organized as follows. In section 2, I go through the basic model under the simplifying assumption that no information about workers, except their colour, is available to the firm. Section 3 studies, in this prototype model, the effects of discrimination on unemployment and duration dependency in exit rates from unemployment. A more sophisticated version of the model, where firms can test applicants before making a hiring decision, is introduced in section 4. Finally, section

5 offers some concluding comments.

2. The Model

Firms with unfilled vacancies search for workers, and unemployed workers search for a job. Open vacancies are announced, and with them a starting wage. Each vacancy has a pre-determined closing time. After the closing time, the firm chooses among the applicants the worker on which they expect to earn the highest profit.[26] All workers have the same distribution of abilities over all jobs in the economy, but they are not equally able in every job. In other words, they have different comparative advantages. Moreover, an important assumption is that workers have private information about their ability.

Formally, let q be the per period value of a match between a worker and a firm. For any job match, q is a random drawing from a *p.d.f.*, $f(q)$. This distribution function is assumed to be the same for all firms and workers. Variations in productivities exist because matches are of different quality, and not because of intrinsic variations in skills or technologies. (I relax this assumption later). The firms must pay a training cost of K to make use of a new worker. When a worker applies, he has some private information about the quality of the match. To simplify, I will assume that they know their productivity q, with certainty. Only after the worker has been in the firm for one period, does the firm learn the productivity. This initial period is referred to as stage 1, and the periods thereafter constitute stage 2. The relative duration of the two stages does not play a role in the analysis. When a firm announces a vacancy, it also announces an initial wage, w_1, and for simplicity this is equal to

[26] For another search model with multiple applications and ranking among the applicants, see Blanchard and Diamond (1990). Empirical evidence on the number of applications per vacancy and details on firms' hiring procedures is given by Barron and Bishop (1985), Barron, Bishop and Duncelberg (1985), Meager and Metcalf (1987), Holzer, Katz and Krueger (1988) and Krueger (1988).

zero. At stage 2, a new per period wage, w_2, is negotiated. Contrary to w_1, this wage is not independent of productivity. Each period at stage 2 the firm has a net profit of $q - w_2$.

Workers and firms are infinitely lived and have a common discount rate r. In each period there is an exogenous probability of separation, s. At stage 2 when the productivity, q, is known to both parties, I assume that the wage is determined by the Nash bargaining solution. Both the firm and the worker has bargaining power since the labour market is not frictionless, also the firm must invest K to replace the incumbent worker.[27] The separation rate, s, is assumed to be the same regardless of whether the two parties are bargaining or have reached an agreement.

The analysis of Binmore, Rubinstein and Wolinsky (1987) is applied in choosing the appropriate disagreement points. These are what each party would get while bargaining and not, as is often used, the outside option. The latter (which for the worker is the value of being unemployed and for the firm the value of a vacancy) only constitute a minimum pay-off to the respective parties.[28] While bargaining, I assume that the worker gets his first stage wage, w_1 , (which I have assumed to be equal to zero) and that he produces q', where $q'<q$. For simplicity q' is set equal to zero.

The sequence of events are illustrated in Figure 1.

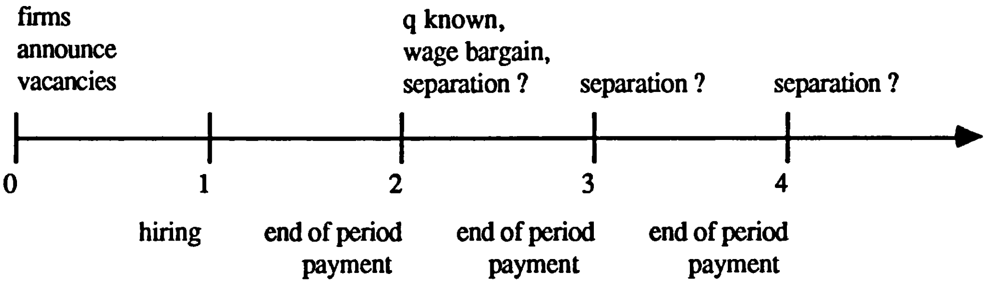

Figure 1

[27] The model has many similar features to Pissarides (1984, 1985), with stochastic job matching, wage determination through bargaining and endogenous determination of workers' reservation wages. The main difference is the assumption of incomplete information at the hiring stage.

[28] Binmore, Shaked and Sutton (1989) found in an experiment that this specification did significantly better than the conventional, and logically incorrect, approach to disagreement points.

We can now formalize the bargaining problem. The second stage wage is determined by[29,30]

$$\max_{w_2} \Omega = (q - w_2)^{\beta} w_2^{1-\beta} \qquad \text{(MP. 1)}$$

$$\text{st: } w_2 \geq rU$$
$$q - w_2 \geq rV$$

where U is the expected lifetime utility when unemployed, and V the expected value of vacancy. By multiplying U and V with r we get the per period value of the outside options. Solving (MP.1) we get

$$w_2 = (1-\beta)\, q \qquad (1)$$

$$\Pi_2 = \beta q \qquad (2)$$

We see that both wages and profit are increasing functions of q. This is nessesary for my results to go through. Both the firm and the worker will get more than their respective outside option, so the outside options are non-binding constraints. A worker's outside option is the value of being unemployed, which by the assumption that w_1=0, is less than what he will get in any job that he has accepted.[31] The outside option for the firm is the value of a vacancy, which in equilibrium will be zero.

[29]I have implicitly assumed that the worker is indifferent between producing q and q'. This simplification does not affect any of the results.

[30]The probability of separation, s, will not effect the outcomes since it is assumed to be exogenous and the same during agreement and bargain.

[31] A sufficient condition for a worker not to accept a job that yield a stage 2 utility less than the unemployment payoff is that $w_1 < w_2$. Recall that a worker can predict w_2 perfectly at the time of hiring.

The key to the model is the insight that workers will accept worse matches when their probability of being selected for a job decreases. This is what I will now show. On the job search is not permitted. While simplifying the analysis, the latter can be endogenized by, e.g., introducing a cost of changing jobs. Total value to a worker in a match of quality q at the beginning of stage 1, $J(q)$, is then

$$J(q) = \frac{R(q)}{1+r} \tag{3}$$

where $R(q)$ is equal to the total value to a worker who is in a match of quality q, at stage 2. Using (1) we get:

$$R(q) = \frac{(1-s)(1-\beta)q}{1+r} + \frac{(1-s)R(q)}{1+r} + \frac{sU}{1+r} \tag{4}$$

which gives that

$$J(q) = \frac{(1-s)(1-\beta)q + sU}{(1+r)(r+s)} \tag{5}$$

Using (2) we find the expression for expected total profit, $E[\Pi]$, in the same way :

$$E[\Pi] = \frac{E[q] - K}{1+r} + \frac{(1-s)\beta E[q] + sV}{(1+r)(r+s)} \tag{6}$$

The expected value of opening a vacancy, V is

$$V = \frac{\gamma E[\Pi]}{1+r} - \frac{I}{1+r} \tag{7}$$

where γ is the probability the vacancy is filled and I the per period cost of an open vacancy. In equilibrium $V=0$, thus

$$\gamma E[\Pi] = I \tag{8}$$

Job announcements arrives to the worker at a rate ϕ. The worker then makes the decision whether to apply or not. I assume that workers do not apply for jobs they will not accept if offered to them. (A small application cost would imply this.) A worker is more inclined to apply the higher he expects the second stage wage to be. Consequently, he selects a cut off rate q_c, above which he applies and below which he does not. Thus, the arrival rate of jobs that a worker applies for is, $\phi[1-F(q_c)]=\lambda(q_c)$. The value of being unemployed is then

$$U = \frac{p\lambda(q_c)E[J \mid q_c]}{1+r} + \frac{[1 - p\lambda(q_c)]U}{1+r} \tag{9}$$

where p is the probability that a worker gets a job given he applies and $E[J|q_c]$ is the expected value to a worker given that he gets a job and that q_c is his cut off rate. Rearranging, we get

$$U = U(p,q_c) = \frac{p\lambda(q_c)E[J \mid q_c]}{r + p\lambda(q_c)} \tag{10}$$

from which it follows that the value of being unemployed is increasing in the acceptance probability, p:

$$\frac{\partial U}{\partial p} = \frac{r\lambda(q_c)E[J \mid q_c]}{(r + p\lambda(q_c))^2} > 0 \tag{11}$$

It is also straightforward to check that a worker's productivity in the job that he is just willing to accept, q_c, is increasing in p. The optimal cut-off level q_c, is defined by

$$\frac{\partial U}{\partial q_c} = 0 \tag{12}$$

The second order condition is assumed to be satisfied.[32] A worker applies for all jobs that gives a higher or equal value to the value of being unemployed. Thus, q_c is also defined by

$$J(q_c) = U(q_c,p) \tag{13}$$

Using (5) and (13) we get

$$q_c = \frac{r(1+r+s)U}{(1\text{-}s)(1\text{-}\beta)} \tag{14}$$

from (11), (12) and (14) we then get

$$\frac{\partial q_c}{\partial p} = \frac{r(1+r+s)}{(1\text{-}s)(1\text{-}\beta)} \frac{\partial U}{\partial p} > 0 \tag{15}$$

Thus, if the acceptance probability, p, falls, the acceptance quality, q_c, will fall.

With this result in hand, the main insight is easily derived. Let there be two types of workers; black and white. The distribution function over different matches $f(q)$ are the same for all black and all white workers. The firms' hiring strategy is a rule for selecting between the applicants. This could be random or weighted in favour of one of the two groups. In the extreme case, the ranking of applicants is strictly in terms of colour. Let the resulting hiring probabilities be denoted p_B and p_W respectively. Let any combination of hiring rules which implies that $p_B=p_W$ on average be denoted a neutral policy. (An example of neutral policy is when all firms

[32] $\frac{\partial^2 U}{\partial q_c^2} < 0$ at $\frac{\partial U}{\partial q_c} = 0 <=> -\phi r f'(q_c)E[J \mid q_c] - 2\phi f(q_c)(r+p\lambda(q_c))\frac{\partial E[J \mid q_c]}{\partial q_c} + \lambda(q_c)(r+p\lambda(q_c))\frac{\partial^2 E[J \mid q_c]}{\partial q_c^2} < 0$

hire at random.) Any other policy is discriminating. The extreme case where all firms only employ blacks (whites) if there is no white (black) applicants are referred to as completely discriminating. I assume that it is always profitable for the firm to hire one of the applicants if there are any.

Proposition 1: In the model where only colour is observable, there are two types of hiring equilibria; neutral and completely discriminating.

Proof: (i) Existence: see appendix 1. (ii) Incomplete discrimination is not an equilibrium. Suppose $p_W > p_B$ on average. Then by (15) we get $q_{cW} > q_{cB}$. This means that expected productivity of whites is higher than expected productivity for blacks therefore all firms prefer to hire whites to blacks. ♦

Let me first present the intuition for why a neutral hiring strategy is an equilibrium. This strategy implies $p_B = p_W$ and hence that $q_{cB} = q_{cW}$. Consequently, both types of applicant have the same expected productivity, and any hiring strategy is optimal, including the neutral. Similarly, the intuition behind the completely discriminating equililibrium is the following. Suppose all firms hire black workers if and only if there is no white applicant. Then $p_B < p_W$, and by (15) $q_{cB} < q_{cW}$. This means that the expected productivity of a black applicant is lower than that of a white, confirming that complete discrimination is rational.

Obviously, in a discriminating equilibrium, black workers earn a lower average wage than do white workers. This is because on average they are not matched as well, so that w_2 is lower.

Another interesting observation is that only the completely discriminating equilibrium is stable. In other words, discrimination is the strong prediction of the model, not only a possible case.

Proposition 2: Only completely discriminating equilibria are stable.

Proof: Any deviation from the neutral equilibrium implies that $p_B \neq p_W$ and therefore all firms will discriminate by the logic of part (ii) of the proof of Proposition 1. But by the same reasoning, a small deviation from the completely discriminating equilibrium does not influence the hiring incentives of firms.♦

In order to check the robustness of the results, it is desirable to relax some of the assumptions of this simple model. In appendix 2, I show that introduction of (small) application costs does not affect the results. A more interesting question is whether discrimination disappears if the firms can observe some noisy signal of workers' productivity. Section 4 shows that this is not the case. On the contrary, allowing firms to test workers expands the set of interesting predictions generated by the model. Thus, I feel justified in performing some further analysis within the simple framework. First, however, let me discuss in some more detail the assumption that profit increases with quality of the match.

One may ask why firms do not adjust the first period wage in such a way that they become indifferent between hiring black and white applicants. Here, this would mean that the firm required a large fee in order to hire black workers. This issue is related to the bonding critique of efficiency wage models, see Carmichael (1985).

Two reasons why the first period wage cannot be manipulated so as to make the firm indifferent between applicants is that it may be infeasible to make a sufficiently detailed announcement (a hiring decision typically have many dimensions - the quality of the applicant being determined also by factors which are hard for the employer to formulate in advance), and that explicit discrimination may be prohibited by law.

At any rate, empirical evidence strongly suggests that firms condition their hiring decision on observable variables such as unemployment duration, work experience and other qualifications. The probability of a job offer is decreasing in the duration of unemployment and increasing in qualifications (see Meager and Metcalf (1987) and

Krueger (1988)). Also, the work by Holzer (1987) and Lynch (1989) confirms that blacks have a significantly lower job offer probability than do whites, i.e. $p_B<p_W$.

I now turn to an issue which is often brought up in discussions about discrimination, namely that of unemployment.

3. Discrimination and unemployment

It is well known that unemployment rates are substantially higher for black workers than for white with similar characteristics (see Rees (1986) and Pissarides and Wadsworth (1990)). It is also established that this is mainly due to the duration rather than the number of unemployment spells (see eg Clark and Summers (1982)). It is interesting to see whether the present model is consistent with this fact. Thus, let us investigate how unemployment rates changes with the probability of getting a job. If unemployment is negatively related to p, the model is consistent with higher unemployment rates for discriminated workers.

The exit rate from unemployment, z, is

$$z = p\lambda(q_c) \tag{16}$$

Clearly,

$$\operatorname{sgn}\frac{dz}{dp} = \operatorname{sgn}\frac{d\phi p[1 - F(q_c)]}{dp} = \operatorname{sgn}\frac{dp[1 - F(q_c)]}{dp}$$

Now,

$$\frac{d(p[1 - F(q_c)])}{dp} = 1 - F(q_c) - p\frac{dF(q_c)}{dq_c}\frac{dq_c}{dp} \tag{17}$$

Thus, we have two effects of a change in p:

1) $1 - F(q_c)$: For a given cut-off productivity, exit rates increase as acceptance probability increases. (+)

2) $p\frac{dF(q_c)}{dq_c}\frac{dq_c}{dp}$: When acceptance probability increases, optimal cut-off productivity increases. i.e. the workers make fewer applications. (-)

Therefore, $p_B < p_W$ implies that black workers have higher unemployment rates than whites if and only if

$$1 - F(q_c) - p\frac{dF(q_c)}{dq_c}\frac{dq_c}{dp} > 0 \tag{C1}$$

In other words, (C1) is both necessary and sufficient for lower acceptance probability to induce lower exit rates, and hence higher unemployment rates. It is not possible to sign the left hand side of (C1) for general functional forms. The ambiguity of the sign of unemployment as a response to changes in vacancies, was noted by Barron (1975). The number of vacancies is here described by the parameter ϕ. Since the sign of $d\phi[1 - F(q_c)] / d\phi$ is the same as the sign of $d(p[1 - F(q_c)])/dp$, Barrons analysis is directly applicable to my model. Burdett (1981) has investigated conditions on $f(q)$ (under the assumption of zero separation rate, i.e. s=0) for exit rates from unemployment to rise as the number of vacancies falls. However, the class of functions which is found to have the desired property is not large enough to be convincing on a priori grounds that unemployment will decrease with p.

But, although it is not possible theoretically to exclude the possibility that $d\phi[1 - F(q_c)] / d\phi$ is negative, the empirical studies by Barron (1975), Pissarides (1986), Jackman and Layard (1991) strongly suggests that it is positive. Thus, the empirical fact that exit rates fall as the number of vacancies falls implies that discriminated workers have higher unemployment rates in the present model.

A well established empirical regularity is that the exit rate falls with the duration of employment (see e.g. Johnson and Layard (1986)). Either this must stem from heterogeneity among workers (some have low exit rates others high), or from true duration dependence. True duration dependence is typically associated with loss of skills or with disillusion and less intensive job search (see e.g. Jackman et. al. (1991), chapter 5). Lockwood (1991) shows that it can also occur if workers differ in abilities, and abilities are not perfectly observable to an employer. The employer will then take unemployment duration as a signal of low ability, since a long term unemployed has failed to be hired by many other employers.

Empirical studies have tried to distinguish between heterogeneity and true duration dependency. Although the results are somewhat mixed, true duration dependency seems to exists. [33]

My model offers a new explanation for true duration dependency. Evidence of firms' hiring procedures shows that firms are reluctant to hire long term unemployed. The probability of being interviewed and obtaining employment decreases with the duration of unemployment spells.[34] If workers find it more difficult to get a job (given that they apply) as the duration of their unemployment spell increases, they will lower their reservation wage (and hence their productivity). This implies that the firms rationally believe the average quality of an applicant to decrease as unemployment durations increases, even without loss of skills, or sorting. This can be stated briefly as follows. Define the value of being unemployed for a worker with a unemployment spell of t as:

$$U(t) = \frac{p_t\phi[1 - F(q_c)]E[J \mid q_c]}{1+r} + \frac{[1 - p_t\phi[1 - F(q_c)]]\, U_{t+1}}{1+r} \qquad (18)$$

[33] See eg. Nickell (1979), Heckman and Borjas (1980), Lynch (1985), Blanchard and Diamond (1989) and Jackman and Layard (1991).

[34] See Meager and Metcalf (1987). For another model that assumes that firms condition hiring decisions on duration of unemployment see Blanchard and Diamond (1990).

with probability $p_t\phi[1 - F(q_c)]$ the unemployed finds a job and with probability $1 - p_t\phi[1 - F(q_c)]$ he stays unemployed. If the hiring probability falls with duration, ie if $p_t > p_{t+i}$ (i>0) then the value of being unemployed will fall with duration, i.e. $U(t) > U(t+i)$, which in turn leads the cut-off productivity q_c to fall with duration. Given condition (C1), it follows that exit rates falls with duration. Thus, the model suggests that we can find duration dependency even without loss of skills, discouragement or sorting.

4. Extension: Firms can test workers

So far, I have made the unrealistic assumption that firms are unable to make any judgement about specific applicants. I now allow firms to test (interview) workers. The test is assumed to be unbiased but not perfect. Empirical studies have shown that firms rarely interview all applicants (see eg. Barron and Bishop (1985), Barron, Bishop and Dunkelberg and Meager and Metcalf (1987)). They first make a short-list of applicants they will interview (test). Since profit is assumed to be linear in q, and firms risk neutral, the firms′ hiring decision will only depend on $E[q]$. The expected value of q, given a test result m' is

$$E[q \mid m'] = \frac{\int_{q_c}^{\infty} q g(m' \mid q) f(q) dq}{\int_{q_c}^{\infty} g(m' \mid q) f(q) dq} \tag{19}$$

where $g(\cdot \mid q)$ is the density function of m given q. Taking the first derivative with respect to q_c we get

$$\frac{\partial E[q \mid m']}{\partial q_c} = \frac{g(m' \mid q_c)f(q_c)\int_{q_c}^{\infty} (q - q_c)g(m' \mid q)f(q)dq}{\left(\int_{q_c}^{\infty} g(m' \mid q)f(q)dq\right)^2} > 0 \tag{20}$$

In words, the expected value of q for a given m, increases with q_c.

I will assume that the Monotone Likelihood Ratio Property (MLRP) holds, i.e. $\frac{g(m' \mid q)}{g(m'' \mid q)}$ is increasing in q if $m'>m''$. MLRP implies that $E[q|m]$ is increasing in m.[35]

Completely discriminating short-listing will refer to the situation where no firm puts a black (white) on the short-list if there are some white (black) applicants that are not on the short-list. Clearly, there are now two ways in which a class of workers can be discriminated against; in the short-listing decision and in the choice between short-listed candidates.

Proposition 3: (i) Any stable equilibrium is completely discriminating. (ii) In a stable equilibrium, a black worker who is short-listed together with a white, will have to score significantly better in order to get hired.

Proof: (i) I first show that incomplete discrimination is not an equilibrium. By definition, a discriminating equilibrium entails $q_{cW} \neq q_{cB}$. If $q_{cW} > q_{cB}$, it implies that the expected productivity of a white worker is higher than that of a black. Consequently, black applicants should be shortlisted only if all white applicants are short-listed. The stability argument is the same as in the proof of Proposition 2. (ii) follows from (20) and that $\partial E[q \mid m]/\partial m > 0$. ♦

[35]MLRP hold for example when the density function $g(\cdot \mid q)$ is normal, poisson or uniform. For a fuller discussion, see Milgrom (1981).

The novel feature of this proposition is the prediction that discriminated groups have to score higher (do better on the interview) in order to be chosen.[36] There is some empirical evidence that this is true. E.g., Cross et al (1990) found in an experiment that compared to equally qualified anglo looking Americans, hispanic looking Americans had a lower probability of getting a job given that they got to an interview. They were also less likely to get to the interview.

The fact that U goes down means that the expected value of a job, and/or the unemployment rate, goes down. Both are not necessarily worsened. Thus, it is possible to construct examples where the average wages of blacks will be higher than the average wages of whites, even though the white workers obtains a higher lifetime utility. In part this is due to the assumed homogeneity of the labour market. If we assume instead that some jobs are more attractive than others, it is very likely that discrimination will hurt along both dimensions.[37] Attractive jobs with high average productivity will have more applicants. Therefore blacks will constitute a higher proportion in the jobs with low average productivity. Thus, even if it is possible that black workers earn the same or even more than whites given a specific job, there will be a tendency that their average wage is lower due to them being over represented in "bad" jobs. Extended in this way, the model opens up the possibility that within a narrow job category wages will be the same for black and white and for men and women, but average wages will differ due to differences in proportions across jobs. So even though gender and racial wage differentials are relatively small when occupations are narrowly defined (see eg. Cain (1986) or Blau and Ferber (1987)), this is not reliable evidence that discrimination is negligible.

I have not compared the efficiency of different equilibria. However, there are

[36] Note that the model does not necessarily imply that discriminated groups will have better visible qualifications in any given job. It only implies that when a woman competes with a man (or a black with a white) she will have to have better visible qualifications in order to be employed.

[37] For evidence of the existence of good and bad jobs, see the literature on inter industry wage differentials (eg. Kruger and Summers (1988)).

other reasons, such as fairness, for disliking discriminatory equilibria. If so, the present model offers a strong case for affirmative action at the hiring stage. Note in particular that equal treatment in one sector will improve the situation for the on average discriminated group also in other sectors. If for example the public sector favours blacks and women at the hiring stage, these groups would find it easier to get a job also in the private sector. Although it would still be hard to get on the short-list, the test-requirements would be weaker. (Because q_c is increasing in p, they can do worse on the test an still be employed.)

5. Concluding remarks

To sum up, it has been shown that discrimination is the unique (type of) stable equilibrium in a model where no one has prejudices. There is no tendency towards segregation (as in Becker (1971), Lang (1986) and Milgrom and Oster (1987)). Nor is there any scope for profitable takeovers as in Becker's model. The model is consistent with many empirical phenomena, including lower wage, higher unemployment, and a higher proportion of blacks and women in bad jobs. An interesting implication of the model is that affirmative action in one sector also serves to reduce discrimination in other sectors. Finally, the paper makes two methodological points. First, equal treatment within narrowly defined occupations does not rule out discrimination. Second, it is important to look at hiring procedures and not only at wages when studying discrimination.

Appendix 1: Existence of equilibria

Here I will show the existence of equilibria. The notation is as in the main text and,

L_i: Number of workers of colour i (exogenous)

D_i: Number of employed workers of colour i

X_i: Number of unemployed workers of colour i.

H_i: Number of hiring, of workers of colour i.

M: Number of vacancies

t: time index

$\alpha_i = \dfrac{\lambda(q_{ci})X_i}{M}$

$i \in \{W, B\}$

In steady state the following must hold;

$$X_i(t+1)=X_i(t) \qquad \text{(A1.1)}$$

$$H_i(t+1)=H_i(t) \qquad \text{(A1.2)}$$

$$D_i(t+1)=D_i(t) \qquad \text{(A1.3)}$$

i.e. the numbers of unemployed, X_i, hirings, H_i, and employed workers, D_i, of colour i are the same each period. Furthermore, in steady state the number of workers of colour i who are hired has to be equal to the number of separation of workers of colour i. This yields the steady state hiring condition;

$$H_i = sD_i \qquad \text{(A1.4)}$$

The following identities are also true;

$$L_i=D_i+X_i \qquad \text{(A1.5)}$$

$$X_i(t+1)=[1-p_i\lambda(q_{ci})]X_i(t) + sD_i(t) \tag{A1.6}$$

(A1.5) says that number of workers of colour i is equal to number of employed plus unemployed workers of colour i. (A1.6) says that number of unemployed workers at the beginning of period t+1 is equal to number of unemployed workers at t multiplied with 1 minus the probability of finding a job, plus the number of separations at time t;

Using the four steady state conditions; (A1.1), (A1.2), (A1.3) and (A1.4) and the unemployment identity; (A1.6) we get the steady state unemployment equation;

$$X_i = \frac{sD_i}{p_i\lambda(q_{ci})} = \frac{s(L_i - X_i)}{p_i\lambda(q_{ci})} = \frac{sL_i}{s + p_i\lambda(q_{ci})} \tag{A1.7}$$

The number of vacancies in the economy, M, is determined by the zero profit condition (equation (8) in the main text).

$$\gamma E[\Pi] = I \tag{A1.8}$$

From equation (15) we know that q_c is a continuous increasing function of p. Thus, the optimal cut-off equation can be written

$$q_{ci}=l(p_i) \tag{A1.9}$$

where l is continuous and increasing, and $l(1) = q_c{}^* < \infty$. Since $J(q_{ci}) = U_i$ in equilibrium, and U_i=0 if p_i=0, we know that $l(0) = 0$. To ensure existence of equilibria I will make two further assumptions. The first one is that (C1) in the main text is satisfied i.e.,

<u>Assumption A1:</u> $\frac{dp_i\lambda(q_{ci})}{dp_i} > 0$

Let $\mathbb{P}(\eta_i{=}c)$ denote the probability that a vacancy has c applicants of colour i. When $\lambda(q_{ci})X_i$ and M are large c is approximately poisson distributed, i.e.,

$$\mathbb{P}(\eta_i{=}c) = \left(\frac{1}{\alpha_i}\right)^c \frac{e^{-\alpha_i}}{c!} \tag{A1.10}$$

Consider now the case where whites are favoured. The number of white hirings, H_w is equal to number of vacancies that have at least one white applicant, i.e.,

$$H_w = M(1-\ e^{-\alpha_i}) \tag{A1.11}$$

On the other hand, H_B is equal to the number of vacancies that have no white applicants and at least one black applicant;

$$H_B = Me^{-\alpha_w}\,(1 - e^{-\alpha_B}) \tag{A1.12}$$

(A1.11) and (A1.12) allow us to get expressions for p_w and p_B.

$$p_w = \frac{H_w}{\lambda(q_{cw})X_w} = \frac{1}{\alpha_w}\left(1 - e^{-\alpha_w}\right) \tag{A1.13}$$

$$p_B = \frac{H_B}{\lambda(q_{cB})X_B} = \frac{e^{-\alpha_w}}{\alpha_B}\left(1 - e^{-\alpha_B}\right) \tag{A1.14}$$

Equilibrium is defined by (A1.5); the steady state hiring identity, (A1.7); the steady state unemployment equation, (A1.8); the zero profit condition, (A1.9); the optimal cut-off equation and (A1.13) and (A1.14); the two acceptance probability equations. The unknown variables are q_{ci}, p_i, X_i , D_i and M. For simplicity I assume that ϕ is a given constant.

The existence proof contains two steps:

1) I show that (A1.7), (A1.9), (A1.13) and (A1.14) imply that q_{ci}, p_i and X_i can be described as continuous functions of M and exogenous variables.

2) I show that (A1.8) yields a solution for M and (A1.5) yields a solution for D_i.

<u>Step 1:</u>

Inserting (A1.9) into (A1.13) we get

$$p_w=k_w(M,p_w,X_w) = \frac{1 - e^{-\alpha_w}}{\alpha_w} \tag{A1.15}$$

where

$$\alpha_w = \frac{\lambda(q_{cw})X_w}{M} = \frac{\phi[1 - F(l(p_w))]X_w}{M}$$

k_w is a continuous function of M, X_w, p_w and exogenous variables. Taking the first derivatives of k_w, with respect to M, X_w and p_w we get

$$\frac{\partial k_w}{\partial M} = \frac{1}{\alpha_w M}\left(1 - e^{-\alpha_w} - \alpha_w e^{-\alpha_w}\right) > 0 \tag{A1.16}$$

$$\frac{\partial k_w}{\partial X_w} = \frac{-1}{\alpha_w X_w}\left(1 - e^{-\alpha_w} - \alpha_w e^{-\alpha_w}\right) < 0 \tag{A1.17}$$

$$\frac{\partial k_w}{\partial p_w} = \frac{-d\lambda(q_{cw})}{dp_w}\frac{1}{\alpha_w\lambda(q_{cw})}\left(1 - e^{-\alpha_w} - \alpha_w e^{-\alpha_w}\right) > 0 \tag{A1.18}$$

By Assumption A1,

$$\frac{dp_w\lambda(q_{cw})}{dp_w} > 0 \Rightarrow \lambda(q_{cw}) + p_w \frac{d\lambda(q_{cw})}{dp_w} > 0$$

$$\Rightarrow -\frac{d\lambda(q_{cw})}{dp_w}\frac{1}{\lambda(q_{cw})} < \frac{1}{p_w} \tag{A1.19}$$

Thus,

$$\frac{\partial k_w}{\partial p_w} < \frac{1}{\alpha_w p_w}(1 - e^{-\alpha_w} - \alpha_w e^{-\alpha_w}) = \frac{1}{(1 - e^{\alpha_w})}(1 - e^{-\alpha_w} - \alpha_w e^{-\alpha_w}) < 1$$

Since $k_w(M,0,X_w) > 0$, $k_w(M,1,X_w) < 1$, $\frac{\partial k_w}{dp_w} < 1$ and $k_w(M,p_w,X_w)$ is continuous in its arguments, we know that (A1.9) and (A1.13) gives p_w as a continuous functions of X_w and M.

Let $Y_w = sX_w + X_w p_w \lambda(q_{cw})$. From (A1.7) we then get; $Y_w(X_w) = sL_w$, and hence

$$\frac{\partial Y_w}{\partial X_w} = s + \lambda(q_{cw})\frac{\partial X_w p_w}{\partial X_w} + X_w p_w \frac{\partial \lambda(q_{cw})}{\partial q_{cw}}\frac{\partial q_{cw}}{\partial p_w}\frac{\partial p_w}{\partial X_w} \tag{A1.20}$$

Further,

$$\frac{\partial X_w p_w}{\partial X_w} = p_w + X_w \frac{\partial p_w}{\partial X_w} = e^{-\alpha_w} > 0 \tag{A1.21}$$

and

$$\frac{\partial \lambda(q_{cw})}{\partial q_{cw}}\frac{\partial q_{cw}}{\partial p_w}\frac{\partial p_w}{\partial X_w} > 0 \tag{A1.22}$$

Thus $\frac{\partial Y_w}{\partial X_w} > 0$. Using that $\frac{\partial Y_w}{\partial X_w} > 0$, $Y_w(0)=0$, and that $Y_w \to \infty$ as $X_w \to \infty$, we know that (A1.7) gives X_w as a continuous function of M. Thus, q_{cw}, p_w and X_w are continuous functions of M.

Analogously,

$$p_B = k_B(M, p_B, X_B, \alpha_w) = \frac{e^{-\alpha_w}}{\alpha_B}(1 - e^{-\alpha_B}) \qquad \text{(A1.23)}$$

Noting that $0 < e^{-\alpha_w} < 1$ and that $\frac{\partial \alpha_w}{\partial M} < 0$, we similarly establish $k_B(M,0,X_B,\alpha_w) > 0$, $k_B(M,1,X_B,\alpha_w) < 1$, $\frac{\partial k_B}{\partial M} > 0$, $\frac{\partial k_B}{\partial p_B} < 1$, $\frac{\partial k_B}{\partial X_B} < 0$. Consequently, (A1.7), (A1.9) and (A1.15) define q_{cB}, p_B and X_B as continuous functions of M.

Step 2:

I now show that that the zero profit condition , (A1.8) solves M.

$$I = \gamma E[\Pi] \qquad \text{(A1.8)}$$

Clearly,

$$\gamma E[\Pi] = (1-e^{-\alpha_w})E[\Pi \mid q_c = q_{cw}] + e^{-\alpha_w}(1 - e^{-\alpha_B})E[\Pi \mid q_c = q_{cB}] \qquad \text{(A1.24)}$$

Thus $\gamma E[\Pi]$ is a continuous functions of q_i, X_i and M. Since q_i and X_i are continuous functions of M, $\gamma E[\Pi]$ is a continuous function of M.

Assumption A2: $M=0$ is not an equilibrium, i.e. $E[\Pi \mid q_{ci}=0] > I$

Assumption A2 is a sufficient (but not necessary) assumption for there to exist a M

such that $\gamma E[\Pi] < I$. As $M \to \infty$, $q_{ci} \to q_{ci}^*$, and $\gamma E[\Pi] \to 0$. As $M \to 0$, $q_{ci} \to 0$ and $\gamma E[\Pi] \to E[\Pi | q_{ci} = 0]$. This, together with Assumption A2 imply that (A1.8) has at least one solution to M. Without making further assumptions we can not exclude the possibility of multiple equilibria.

Obviously, for a given X_i, (A1.5) solves D_i.

Since $p_w > p_B$ by the above hiring rule, also $q_{cw} > q_{cB}$ (by (A1.9)), confirming that the hiring rule is optimal.

The proof that there exists a neutral equilibrium is analogous.

Appendix 2: Application Cost

I will now consider the sensitivity of the model to the assumption of zero application cost (pecuniary or non-pecuniary). The worker has to be indifferent between the value of being unemployed and the expected value from applying for a job of value q_C. Thus, with a positive application cost, A, q_c is defined by

$$pJ(q_c) - A + (1-p)U(q_c,p,A) = U(q_c,p,A) \tag{A2.1}$$

This is the analog to (13) in the main text. (A2.1) simplifies to

$$J(q_c) = U(q_c,p,A) + \frac{A}{p} \tag{A2.1'}$$

Using (5) in the main text and (A2.1') we get

$$q_c = \frac{r(1+r+s)U + \frac{A(r+s)(1+r)}{p}}{(1-s)(1-\beta)} \tag{A2.2}$$

thus,

$$\frac{\partial q_c}{\partial p} > 0 \Leftrightarrow r(1+r+s)\frac{dU}{dp} - \frac{A(r+s)(1+r)}{p^2} > 0 \tag{A2.3}$$

The value of being unemployed is now (the analog to (10) in the main text),

$$U = \frac{p\lambda(q_c)E[J \mid q_c] - A\lambda(q_c)}{r + p\lambda(q_c)} \tag{A2.4}$$

For my results to hold, cut-off productivities, q_c, has to fall as p falls, i.e. $\frac{\partial q_c}{\partial p} > 0$. Using (A.3) and (A.4), we get the following condition for $\frac{\partial q_c}{\partial p} > 0$

$$\frac{E[J \mid q_c]}{A/p} > \frac{(r+s)(1+r)(r + p\lambda(q_c))^2 - r(1+r+s)p^2\lambda(q_c)^2}{r^2 p\lambda(q_c)(1+r+s)} \tag{A2.5}$$

where, $E[J \mid q_c]$ is the expected lifetime earnings, given that one has a job and A/p is the expected total application cost before getting a job. For small enough values of r, (A2.5) will not hold. As r goes to zero, the right hand side goes to infinity.

If (A2.5) doesn't hold, it implies that as the probability of getting a job goes down, the workers will increase their cut-off. Consequently, in bad times, when the number of applications per vacancy is high, workers should be more choosy. Conversely, if it can be established that workers apply less critically in bad times, application costs do not represent a problem to the present analysis.

Also, even if A was sufficiently high in some jobs, so that (A2.5) did not hold for these, this would not bring the economy all the way to a neutral equilibrium. In jobs where (A2.5) doesn't hold optimal hiring procedures is neutral, but where (A2.5) hold they are discriminating.

References

Akerlof, George, (1985): Discrimination, Status-Based Wages among Tradition Oriented Stochastically Trading Coconut Producers, *Journal of Political Economy*,93, April, 265-276.

Arrow, Kenneth, J., (1973): The Theory of Discrimination, in *Discrimination in Labor Markets* (ed. O.A. Ashenfelter and A. Rees), Princeton, N.J.:Princeton University Press, 3-33.

Barron, John M., (1975): Search in the Labor Market and the Duration of Unemployment: Some Empirical Evidence, *American Economic Review*, 65, December, 934-942.

Barron, John M. and John Bishop, (1985): Extensive Search, Intensive Search and Hiring Costs: New Evidence on Employer Hiring Activity, *Economic Inquiry*, 23, July,363-382.

Barron John M., John Bishop and William C. Dunkelberg, (1985): Employer Search: The Interviewing and Hiring of New Employees, *Review of Economics and Statistics*, 67, February, 43-52.

Becker, Gary S., (1971): *The Economics of Discrimination*, Chicago: University of Chicago Press. (Original edition, 1957)

Binmore, Ken, Ariel Rubinstein and Asher Wolinsky, (1986): The Nash Bargaining Solution in Economic Modelling, *Rand Journal of Economics*, 17, Summer, 176-188.

Binmore, Ken, Avner Shaked, and John Sutton, (1989): An Outside Option Experiment *Quarterly Journal of Economics*, 104, November, 753-770.

Blanchard, Oliver J. and Peter Diamond, (1989): The Beveridge Curve, *Brookings Papers on Economic Activity*, 1, 1-60.

Blanchard, Oliver J. and Peter Diamond, (1990): Ranking, Unemployment, and Wages, NBER working paper no. 3387.

Blau, Francine D. and Lawrence Kahn, (1981): Race and Sex Differences in Quits by

Young Workers, *Industrial and Labor Relations Review*, 34, July, 563-577

Blau, Francine, D. and Marianne A. Ferber, (1987): Discrimination: Empirical Evidence from the United States, *American Economic Review*, 77, May, 316-320.

Borjas, George J. and Stephen G. Bronars, (1989): Consumer Discrimination and Self-Employment, *Journal of Political Economy*, 97, June, 581-605.

Bulow, Jeremy I. and Lawrence H. Summers, (1986): A Theory of Dual Labor Markets with Application to Industrial Policy, Discrimination, and Keynesian Unemployment, *Journal of Labor Economics*, 4, July, 376-414.

Burdett, Kenneth, (1981): A Useful Restriction on the Offer Distribution in Job Search Models, In *Studies in Labor Market Behavior: Sweden and the United States* (ed. G. Eliasson, B. Holmlund and F.P. Stafford), Stockholm: The Industrial Institute for Economic and Social Research. 169-182.

Cain, Glen G., (1986): The Economic Analysis of Labor Market Discrimination: A Survey, in *Handbook of Labor Economics* (ed. O. Aschenfelter and R. Layard), 1, chapter 13, Amsterdam: North-Holland.

Carmichael, Lorne, (1985): Can Unemployment Be Involuntary? : Comment, *American Economic Review*, 75, December, 1213-1214.

Cho, In-Koo and David, M. Kreps, (1987): Signaling Games and Stable Equilibria, *Quarterly Journal of Economics*, 102, May, 179-222.

Clark, Kim B. and Lawrence H. Summers, (1982): The Dynamics of Youth Unemployment, in *The Youth Labor Market Problem* (ed. R. Freeman and D. Wise). Chicago: University of Chicago Press.

Cross, Harry, Genevieve Kenney, Jane Mell and Wendy Zimmermann, (1990): *Employer Hiring Practices: Differential Treatments of Hispanic and Anglo Job Seekers*. Urban Institute Report, no. 90-4, The Urban Institute Press, Washington D.C.

Harber, Sheldon E., Enrique J. Lamas and Gordon Green, (1983): A New Method of

Estimating Job Separations by Race and Sex, *Monthly Labor Review*, 106, June, 20-27.

Heckman, James, J. and George J. Borjas, (1980): Does Unemployment Cause Future Unemployment? Definitions, Questions and Answers from a Continuous Time Model of Heterogeneity and State Dependency, *Economica*, 47, August, 247-283.

Holzer, Harry J., (1987): Informal Job search and Black Youth Unemployment, *American Economic Review*, 77, June , 446-452.

Holzer, Harry, J. Lawrence F. Katz, and Alan B. Krueger, (1988): Job Queues and Wages: New Evidence on Minimum Wage and Inter-Industry wage structure" NBER working Paper no.2561.

Jackman, Richard and Richard Layard, (1991): Does Long-term Unemployment Reduce a Person's Chance of a Job? A Time-Series Test, *Economica*, 58, February, 93-106.

Jackman, Richard, Richard Layard and Stephen Nickell, (1991): *Unemployment: Macroeconomic Performance and the Labour Market*, Oxford University Press, (forthcoming).

Johnson, George E. and Richard Layard, (1986): The Natural Rate of Unemployment: Explanation and Policy, in *Handbook of Labor Economics* (ed. O. Ashenfelter and R. Layard), 2, chapter 16, Amsterdam: North Holland.

Krueger, Alan B., (1988): The Determinants of Queues for Federal Jobs, *Industrial and Labor Relations Review*, 41, July, 567-581.

Kruger, Alan B. and Lawrence H. Summers, (1988): Efficiency Wages and the Inter-Industry Wage Structures, *Econometrica*, 56, March, 259-293.

Lang, Kevin., (1986): A Language Theory of Discrimination, *Quarterly Journal of Economics,* 101, May, 363-382.

Lockwood, Ben, (1990): Information Externalities in the Labour Market and the Duration of Unemployment, *Review of Economic Studies*, 58, July, 733-753.

Lundberg, Shelley J. and Richard Startz, (1983): Private Discrimination and Social

Intervention in Competitive Labor Markets, *American Economic Review*, 73, June, 340-347.

Lynch, Lisa M., (1985): State Dependency in Youth Unemployment, *Journal of Econometrics*, 28, April, 71-84.

Lynch, Lisa M., (1989): The Youth Labor Market in the Eighties: Determinants of Re-Employment Probabilities for Young Men and Women, *Review of Economics and Statistics*, 71, February, 37-45.

McCall, John J., (1972): The Simple Mathematics of Information, Job Search, and Prejudice, in *Racial Discrimination in Economic Life* (ed. A. Pascal), Lexington, Mass,: Lexington Books, 205-224.

Meager, Nigel and Hilary Metcalf, (1987): Recruitment of the Long Term Unemployed, Institute of Manpower Studies Report , September, no. 138.

Milgrom, Paul, R. and Sharon Oster, (1987): Job Discrimination, Market Forces, and the Invisibility Hypothesis, *Quarterly Journal of Economics*,102, August, 453-476.

Milgrom, Paul, R., (1981): Good News and Bad News: Representation Theorems and Applications, *Bell Journal of Economics*, 12, Autumn, 380-391.

Nöldeke, Georg and Eric van Damme, (1990): Signalling in Dynamic Labour Market" *Review of Economic Studies*, 57, January, 1-23.

Nickell, Stephen, (1979): Estimating the Probability of Leaving Unemployment *Econometrica*, 47, September, 1249-1266.

Phelps, Edmund, S., (1972) : The Statistical Theory of Racism and Sexism, *American Economic Review*, 62, September, 659-661.

Pissarides, Chrisopher, A. and Jonathan Wadsworth, (1990): Who are the Unemployed?, London School of Economics, Centre for Economic Performance, Discussion Paper no. 12.

Pissarides, Christopher, A., (1986): Unemployment and Vacancies in Britain, *Economic Policy*, October, 499-559.

Pissarides, Christopher, A., (1985): Short-Run Equilibrium Dynamics of Unemployment, Vacancies, and Real Wages, *American Economic Review*, 75, September, 676-690.

Pissarides, Christopher, A., (1984): Efficient Job Rejection, *Economic Journal,* Conference Papers, 94, 97-108.

Rees, Albert, (1986): An Essay on Youth Joblessness, *Journal of Economic Literature*, 24, june, 613-628.

Riley, John G., (1979): Informational Equilibrium, *Econometrica*, 47, March, 331-359.

Rothschild, Michael and Joseph E. Stiglitz, (1982): A Model of Employment Outcomes Illustrating the Effect of the Structure of Information on the Level and Distribution of Income, *Economic Letters,* 10, 231-236.

Spence, Michael A., (1973): Job Market Signaling, *Quarterly Journal of Economics,* 87, August, 355-374.

Viscusi, Kip, W., (1980): Sex Differences in Quitting, *Review of Economics and Statistics*, 62, August, 388-398.

Addenda (In pocket at back)

Contents Page

Addenda

Addendum to Chapter 1

In chapter 1, I have characterized the solution to the Nash bargaining problem

$$\operatorname*{Max}_{w} X = (\Pi-\Pi_0)^{\beta}(Z-Z_0) \qquad \text{st. } e=e_2, N(w,e) \qquad (15)$$

by its first order condition, thus implicitly assuming that the second order condition is satisfied. (Interiority is ensured by mild technical restrictions.)

I now give the steps leading to the conclusion that wage and effort is positively related at this interior solution, and we consider the region above the contract curve.

Step 1) $\dfrac{\partial X}{\partial w} = 0 \, , \; \dfrac{\partial^2 X}{\partial w^2} < 0 \;\Rightarrow\; \dfrac{\partial Z}{\partial w} > 0.$

Step 1) follows straight forward from (17), (11) and the fact that $(\Pi-\Pi_0)$ and $(Z-Z_0)$ are greater than zero when X is maximized.

Step 2) From (11) it follows that; $\dfrac{\partial Z}{\partial w} > 0 \iff u'(w)w - \varepsilon r > 0.$

Step 3) Above the contract curve $u'(w)w - r - eg'(e) < 0$. Combining this with step 2) we have

$$r(\varepsilon - 1) - eg'(e) < 0 \; .$$

Using (12), we then know that $\dfrac{\partial Z}{\partial e} < 0.$

Step 4) If $\frac{\partial Z}{\partial w} > 0$ and $\frac{\partial Z}{\partial e} < 0$ it is now easy to check that $\frac{dw}{de}\big|_{\frac{\partial X}{\partial w}=0} > 0$. (Where $\frac{dw}{de}\big|_{\frac{\partial X}{\partial w}=0}$ is described in equation (18)).

Addendum to Chapter 2

Appendix 1

I will here show that MLRP implies that $E_{\alpha}[x \mid w_{2\alpha}(x)] \geq E_{\beta}[x \mid w_{2\alpha}(x)]$

Assumptions;

1) $\frac{f_{\alpha}(x)}{f_{\beta}(x)}$ is increasing in x (MLRP). $\quad x \in R^{+} \quad$ (A1.1)

2) Let A be a subset of R^{+} such that $\int_A f_{\alpha}(x)dx > 0$.

Put,

$$E_{\iota}[x \mid x \in A] = \frac{\int_A x f_{\iota}(x)dx}{\int_A f_{\iota}(x)dx}, \qquad \iota \in \alpha,\beta \tag{A1.2}$$

Proposition A1.1: Given assumptions 1) and 2);

$$E_{\alpha}[x \mid x \in A] > E_{\beta}[x \mid x \in A]. \tag{A1.3}$$

To prove Proposition A1.1 we will use the following lemma, which is proven afterwards. Although we shall use the lemma only for the case when the probability density $h(x)$ has its entire mass on $x \in R^{+}$, we formulate it more generally.

Lemma A1.1: Let $h(x)$, $-\infty < x < \infty$, be a probability density , i.e. $h(x) \geq 0$ and $\int_{-\infty}^{\infty} h(x)dx = 1$. Then it holds for $u(x)$ and $v(x)$, $-\infty < x < \infty$, which both are non-negative and increasing that;

$$\int_{-\infty}^{\infty} u(x)v(x)h(x)dx \geq \left(\int_{-\infty}^{\infty} u(x)h(x)dx\right)\left(\int_{-\infty}^{\infty} v(x)h(x)dx\right). \tag{A1.4}$$

Proof of Proposition A1.1: By (A1.2) we have that

$$\left(\int_A f_\alpha(x)dx\right) E_\alpha[x \mid x \in A] = \int_A x f_\alpha(x)dx =$$

$$\int_A x \left(\int_A f_\beta(x)dx\right) \frac{f_\alpha(x)}{f_\beta(x)} \frac{f_\beta(x)}{\int_A f_\beta(x)dx} dx \qquad \text{(A1.5)}$$

Using lemma A1.1 with

$$h(x) = \frac{f_\beta(x)}{\int_A f_\beta(x)dx} \quad \text{for } x \in A, =0 \text{ for } x \notin A$$

$$v(x) = \left(\int_A f_\beta(x)dx\right) \frac{f_\alpha(x)}{f_\beta(x)} \qquad \text{increasing by (A1.1)}$$

$$u(x) = x \qquad \text{increasing}$$

we get that;

$$\int_A x \left(\int_A f_\beta(x)dx\right) \frac{f_\alpha(x)}{f_\beta(x)} \frac{f_\beta(x)}{\int_A f_\beta(x)dx} dx$$

$$\geq \left(\int_A x f_\beta(x)dx\right)\left(\int_A f_\alpha(x)dx\right) / \int_A f_\beta(x)dx = \left(\int_A f_\alpha(x)dx\right) E_\beta[x \mid x \in A] \qquad \text{(A1.6)}$$

The claim in proposition A1.1 now follows from (A1.5) and (A1.6)

Proof of Lemma A1.1: Let h be as stated in Lemma A1.1.

<u>Step1:</u> Let $J(x;a)$ be the function that jumps from 0 to 1 in $x=a$. (J is a one-step function). (A1.4) holds for $u(x)=J(x;a)$ and $v(x)=J(x;b)$ (for arbitrary a and b). To prove this put $H(x) = \int_{-\infty}^{x} h(y)dy$. It is readily checked that

$$\int_{-\infty}^{\infty} J(x;a)h(x)dx = 1 - H(a)$$

$$\int_{-\infty}^{\infty} J(x;b)h(x)dx = 1 - H(b)$$

$$\int_{-\infty}^{\infty} J(x;a)J(x;b)h(x)dx = 1 - H(\max(a,b)).$$

The claim now follows from the above formulas and the following inequality.

$$1 - H(\max(a,b)) \geq (1-H(\max(a,b)))\,(1-H(\min(a,b))) =$$
$$(1 - H(a))(1 - H(b)).$$

Step 2: Let $u_1(x), u_2(x),..., u_N(x)$ and $v_1(x), v_2(x),..., v_M(x)$ be functions such that (A1.4) hold for every pair $(u_\gamma(x), v_\rho(x))$, $\gamma=1,2,...,N$, $\rho=1,2,...,M$, and let $c_1,c_2,...,c_N$ and $d_1,d_2,...,d_M$ be arbitrary non-negative numbers. Then (A1.4) holds for

$$u(x) = \sum_{\gamma=1}^{N} c_\gamma u_\gamma(x) \text{ and } v(x) = \sum_{\rho=1}^{N} d_\rho v_\rho(x) .$$

Proof:

$$\int_{-\infty}^{\infty} u(x)v(x)h(x)dx = \sum_{\gamma=1}^{N}\sum_{\rho=1}^{M} c_\gamma d_\rho \int_{-\infty}^{\infty} u_\gamma(x)v_\rho(x)h(x)dx \geq \text{(by assumption)}$$

$$\sum_{\gamma=1}^{N}\sum_{\rho=1}^{M} c_\gamma d_\rho \left(\int_{-\infty}^{\infty} u_\gamma(x)h(x)dx\right)\left(\int_{-\infty}^{\infty} v_\rho(x)h(x)dx\right) =$$

$$\left[\int_{-\infty}^{\infty}\left(\sum_{\gamma=1}^{N} c_\gamma u_\gamma(x)\right)h(x)dx\right]\left[\int_{-\infty}^{\infty}\left(\sum_{\rho=1}^{M} d_\rho v_\rho(x)\right)h(x)dx\right] =$$

$$\left(\int_{-\infty}^{\infty} u(x)h(x)dx\right)\left(\int_{-\infty}^{\infty} v(x)h(x)dx\right).$$

End of proof.

Steps 1 and 2 yield that (A1.4) holds for arbitrary non-negative, increasing step functions $u(x)$ and $v(x)$, i.e. functions of the type $\sum_{\gamma=1}^{N} c_\gamma J(x;a_\gamma)$, $c_\gamma \geq 0$.

Step 3: If (A1.4) holds for $(u_1,v_1), (u_2,v_2),..., (u_n,v_n)...$ and if $u(x) = \lim_{n\to\infty} u_n(x)$ and $v(x) = \lim_{n\to\infty} v_n(x)$ exist, then (A1.4) holds for $u(x)$ and $v(x)$. Proof is obtained by letting $n\to\infty$ in the inequality

$$\int_{-\infty}^{\infty} u_n(x)v_n(x)h(x)dx \geq \left(\int_{-\infty}^{\infty} u_n(x)h(x)dx\right)\left(\int_{-\infty}^{\infty} v_n(x)h(x)dx\right).$$

The claim in lemma A1.4 now follows from the fact that every non-negative, increasing function can be obtained as lim of a sequence of increasing step functions (approximation from below).

Addendum to chapter 3

Here I will show that $\frac{\partial U}{\partial q_c} = 0 \iff J(q_c) = U,$ (used at page 82)

$$U = \frac{p\lambda(q_c)E[J \mid q_c]}{r + p\lambda(q_c)} \qquad \text{(a1)}$$

where,

$$\lambda(q_c) = \phi[1 - F(q_c)]$$

$$E[J \mid q_c] = \frac{\int_{q_c}^{\infty} J(q)f(q)dq}{\int_{q_c}^{\infty} f(q)dq}$$

Differentiating $E[J \mid q_c]$ with respect to q_c we get,

$$\frac{\partial E[J \mid q_c]}{\partial q_c} = \frac{-J(q_c)f(q_c)\int_{q_c}^{\infty} f(q)dq + f(q_c)\int_{q_c}^{\infty} J(q)f(q)dq}{\left[\int_{q_c}^{\infty} f(q)dq\right]^2} =$$

$$\frac{-J(q_c)f(q_c)}{\int_{q_c}^{\infty} f(q)dq} + \frac{f(q_c)E[J \mid q_c]}{\int_{q_c}^{\infty} f(q)dq} = \frac{-J(q_c)f(q_c)}{1 - F(q_c)} + \frac{f(q_c)E[J \mid q_c]}{1 - F(q_c)} \qquad \text{(a2)}$$

Now,

$$\frac{\partial U}{\partial q_c} = 0 \quad <=>$$

$$\left(-\phi f(q_c)E[J|q_c] - \phi[1 - F(q_c)]\frac{J(q_c)f(q_c)}{[1 - F(q_c)]} + \phi[1 - F(q_c)]\frac{f(q_c)E[J|q_c]}{[1 - F(q_c)]}\right)\left(r + p\lambda(q_c)\right) +$$

$$\phi f(q_c)p\phi[1 - F(q_c)]E[J|\, q_c] = 0 \qquad \text{(a3)}$$

$<=>$

$$\{ -\phi f(q_c)E[J \mid q_c] - \phi J(q_c)f(q_c) + \phi f(q_c)E[J|q_c]\} \ \{r + p\lambda(q_c)\} +$$

$$\phi f(q_c)p\phi[1 - F(q_c)]E[J|q_c] = 0 \qquad \text{(a4)}$$

$<=>$

$$- \phi J(q_c)f(q_c) + \frac{\phi f(q_c)p\phi[1 - F(q_c)]E[J|q_c]}{r + p\lambda(q_c)} = 0 \qquad \text{(a5)}$$

$<=>$

$$J(q_c) = \frac{p\phi[1 - F(q_c)]E[J|q_c]}{r + p\lambda(q_c)} = \frac{p\lambda(q_c)E[J|q_c]}{r + p\lambda(q_c)} = U \qquad \text{(a6)}$$