The Economics of Industrial Relations

Handouts for B44 (Spring 2002)

Jaap H. Abbring*

Department of Economics
University College London

April 7, 2002

Preface and advice on reading

These notes are incomplete. They are meant to complement rather than to substitute your own notes and the readings. Also, this is a first draft written in a relatively short amount of time. It may contain typos or other mistakes. Comments are most welcome.

These are partial transcriptions of my lecture notes. I have worked out some parts, like the analysis of the union models, in quite some detail. Some other parts are not included or worked out. I have usually described these parts between square brackets "[" and "]". Finally, I skipped over some of the material in the lectures, referring you to the reading.

The syllabus and the course web site list all required and other reading. I have added references throughout these notes, but consult the syllabus, web site and your own notes (or me) if in doubt. The most important texts for the exam are Booth (1995), Ch. 1–7, Milgrom and Roberts (1992), Ch. 10-13, and some text on efficiency wages (e.g. Milgrom and Roberts, part of Ch. 8; see references). The first part of McDonald and Solow (1981) is a good source for some of the union models. The complementary reading suggested in the notes, the syllabus and on the course web site will help you getter a better and broader understanding of the material.

Contents

1 Introduction

[Course overview along the course outline in the syllabus.]

2 Standard neoclassical labour market theory, unions and human-resource systems

Reading:

- Your favourite textbook (e.g. Borjas, 2000, or Bosworth, Dawkins and Stromback, 1996): Review of standard neoclassical model
- Booth (1995), Ch. 1: Introduction to unions
- First part of Milgrom and Roberts (1992), Ch. 10: Some background on labour contracts and human resources systems

[Review of neo-classical spot labour markets; Incentives for general and specific human-capital investments in such markets; Failure to describe observed labour relations.]

3 Trade unions and labour market power

3.1 Overview and history

Reading:

- Booth, Ch. 2: Historical overview of UK and US unionism
- Cully et al (1999, from p. 234) and Borjas (2000), Ch. 11: Recent data on resp. UK and US
- Layard, Nickell and Jackman (1991): Broader international comparison
- Polachek and Siebert (1993), Ch. 10, and Bosworth, Dawkins and Stromback (1996), Ch. 25: Additional sources of data and history

[Quick overview of some facts using tables and graphs from the texts above.]

3.2 Unionisation of a small sector

Reading:

- Booth (1995), Ch. 4 and 5;
- McDonald and Solow (1981), in particular the first part:
- Relevant sections from Borjas, 2000, Ch. 11, and/or Bosworth et al, 1996, Ch. 25: useful text book reviews of union models.

3.2.1 Competitive benchmark

Consider a small labour market sector with workers and firms that are pricetakers.

Workers

(i). inelastically supply 1 unit of labour (say, the equivalent of a 40 hour work week) to the economy;

(ii). can work at some wage w_0 elsewhere in the economy (outside the sector we are considering), where w_0 is "exogenously" given (that is, determined outside the model);

(iii). derive utility $u(w)$ from working at a wage w, with u (strictly) increasing ($u'(w) > 0$).

The fact that workers inelastically supply labour to the *economy* does not imply that they inelastically supply labour to the *sector*. Rather, in the competitive model they are free to move between the sector and the rest of the economy and try to sell their unit of labour services whereever they want. Thus, they will compare the wages offered for the type of labour services they are selling between the sector and the rest of the economy and will supply labour whereever the wage is highest. Because the workers can earn w_0 for their unit of labour elsewhere in the economy, they all want to work in the sector if the sectoral wage $w > w_0$. They prefer to work elsewhere in the economy if the sectoral wage $w < w_0$. If $w = w_0$ they will be indifferent. In conclusion, even though aggregate labour supply is inelastic, the supply of labour to the sector is perfectly elastic at the wage w_0: supply is 0 if $w < w_0$ and equal to all available labour if $w > w_0$.

Firms in the sector have a downward-sloping demand curve $n(w)$ for labour (that is, $n'(w) < 0$).[1]

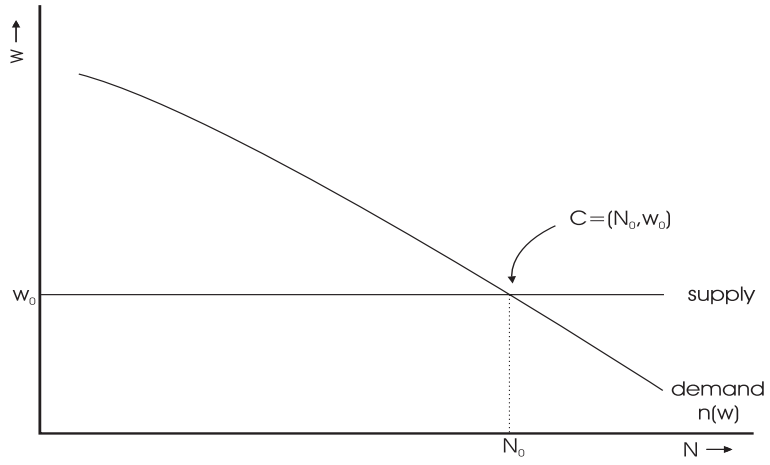
In competitive equilibrium, wages adjust so that supply equals demand. Figure 1 illustrates this. The figure has sectoral wages w on the vertical axis and quantity of sectoral labour N on the horizontal axis. The downward-sloping line is the labour-demand curve. For each wage w on the vertical axis, it gives the labour-demand level $n(w)$ on the horizontal axis. Competitive equilibrium is the pair $C = (N_0, w_0)$ at which supply and demand intersect. Of course, $N_0 = n(w_0)$ because C is on the labour-demand curve.

Remark 1. The labour-demand curve in Figure 1 is really the *inverse* labour-demand function n^{-1}: for each employment level N on the horizontal axis you find the wage $n^{-1}(N)$ on the vertical axis.[2] The inverse labour demand $n^{-1}(N)$ at employment level N

[1] I am using lower case n to explitly distinguish the labour-demand *function* from a given labour-demand or employment level N. In class, I used upper case N for both.

[2] The labour-demand curve is invertible by the assumption that $n'(w) < 0$.

Figure 1: Competitive equilibrium in a small labour market sector



can be interpreted as the firms' "willingness-to-pay" for one additional unit of labour if N units are already employed. We return to this later.

Remark 2. The sector is "small" in the sense that the wage w_0 elsewhere in the economy does not depend on employment in the sector. So, we can take w_0 to as exogenously given if we analyze equilibrium in the small sector. Obviously, because aggregate labour supply is fixed, labour supply to the rest of the economy is a fixed number minus employment in the small sector. If employment in the sector decreases, labour supply to the rest of the economy increases and wages w_0 there typically tend to fall. The "small" sector assumption basically says that the sector has more or less a zero weight in the bigger economy: employment in the sector is an ignorable part of aggregate labour supply, and any workers displaced from the sector to the rest of the economy will not noticeably add to labour supply there. This is of course an abstraction for simplification: it allows us to take w_0 as fixed. In Subsection 3.3, we will analyze a two-sector model in which changes in one sector "spill over" to the other sector.

Remark 3. The assumption that workers inelastically supply one unit of labour to the economy is made for simplicity. It is immaterial in the model of a small sector.

Now suppose that $M > N_0$ workers organise in a union that monopolises the supply of labour to this small sector. Unlike in the competitive model above, the workers are now not wage-takers anymore. Rather, they have market power in the sectoral market. We discuss three models, (*i*) the monopoly-union model, (*ii*) the efficient-bargaining model and (*iii*) the (general) right-to-manage model. One of the main distinctions between these models is whether they grant firms the right-to-manage or not. The firms have the

right-to-manage in this context if the amount of labour services hired is solely at their discretion, rather than being prescribed by the union or a binding contract agreed upon in bargaining.

Before discussing the three models in detail, we give a brief characterisation of each.

(i). *The monopoly-union model*

In this model,

(a) the union unilaterally sets the sectoral wage and

(b) the firm(s) in the sector choose employment.

The union only imposes a wage on the sector; it is not able to impose an employment level. Thus, the firms have the right-to-manage and pick an employment level that they prefer given the wage set by the union. Obviously, this will be the corresponding employment level on the labour-demand curve.

(ii). *The efficient-bargaining model*

In this model, the firms and the union bargain over both sectoral wages and employment. Bargaining requires the firms to somehow be organised (or to be just one firm). It is assumed that the firm and the worker can write a binding contract specifying both the wage and employment. Thus, this is not a right-to-manage model.

(iii). *The (general) right-to-manage model*

In this model,

(a) the firm(s) and the union bargain over the sectoral wage and

(b) the firm(s) in the sector choose employment.

Like the efficient-bargaining model, this is a bargaining model and therefore requires the firms to be organised (or to be one firm). Unlike the efficient-bargaining model, the union and the firm only bargain over the wage and leave the right-to-manage to the firm. The firm picks employment on the labour-demand curve, so that bargaining is effectively over points on the labour-demand curve. The monopoly-union model is a special case of the right-to-manage model in which the union has all the bargaining power (that is, can set the wage unilaterally) in the first stage.

Remark 4. At this point, you may wonder how the M workers have been able to monopolise the sectoral labour market and attain market power. We will return to this later. For now, we assume that they are somehow able to do so effectively.

3.2.2 Monopoly-union model

Point of departure is the competitive benchmark model of the small sector. $M > N_0$ workers organise in a union that monopolises the supply of labour to the sector. The union is a "monopoly-union" in that it can unilaterally set wages subject to the right-to-manage constraint that firms choose employment taking the wage set by the union as given. Thus, the model has two stages,

(i). the union unilaterally sets the sectoral wage and

(ii). the firm(s) in the sector choose employment.

First, we have to discuss the objective used by the union in setting wages. In general, unions prefer both higher wages and higher employment for their M members. We summarize union preferences by a union utility function $V(w, N)$ that is increasing in w and N. Here it is assumed that union membership M is exogenous, that is determined outside the model. Unions choose wages and employment that maximise union utility, taking membership M as given and subject to the right-to-manage constraint. More formally, the union chooses a wage w_m and union employment level N_m equal to

$$(w_m, N_m) = \arg\max_{(w,N)} V(w, N) \text{ such that } N = n(w). \tag{1}$$

Here, "$\arg\max_{(w,N)} V(w, N)$" simply means the value of (w, N) that maximises $V(w, N)$.

It may seem a bit strange to formulate the union's first-stage choice as a choice of both a wage and an employment level. However, note that the right-to-manage constraint $N = n(w)$ restricts this choice to be on the labour-demand curve. Thus, the union effectively chooses a wage w_m only. The corresponding employment level N_m then follows from $N_m = n(w_m)$.

We can make this explicit by substituting the right-to-manage constraint in the objective function, which gives the union's "indirect" utility $V(w, n(w))$ as a function of the wage w only. Then, we can reformulate (1) as

$$w_m = \arg\max_{w} V(w, n(w)) \quad \text{and} \quad N_m = n(w_m). \tag{2}$$

An increase in w has a direct positive effect on $V(w, n(w))$ because the union prefers higher wages over lower wages for its members employed in the union sector. However, an increase in w also has an indirect negative effect on $V(w, n(w))$ through the negative effect of higher wages on union employment. Thus, the union faces a trade-off between higher wages for its members who are employed in the union sector and the induced lower union employment.

To investigate this further, we assume that the union has *utilitarian* preferences, that is

$$V(w,N) = \begin{cases} Nu(w) + (M-N)u(w_0) & \text{if } N < M \\ Mu(w) & \text{if } N \geq M, \end{cases} \tag{3}$$

with $u(w)$ the utility derived by a single union member from income w. In words, utilitarian union utility $V(w,N)$ at a union wage w and union employment N is simply the sum of the individual utilities of all its M members. If some of the union members cannot be employed in the union sector, that is $N < M$, then N members hold union jobs at a wage w and have utility $u(w)$ and $(M-N)$ members work elsewhere at a wage w_0, from which they derive utility $u(w_0)$. If all union members are employed in the union sector at a wage w, that is $N \geq M$, then all M members have utility $u(w)$. Here, we assume that the union does not care about workers that are not members of the union. In our analysis and graphs, we will mostly ignore the case in which $N > M$. Students should however keep in mind that we should be careful if we encounter high levels of union employment.

We have already assumed that $u'(w) > 0$ for all w, so that union members prefer more income over less. Often, we also assume that $u''(w) \leq 0$ for all w, so that union members are either risk-neutral or risk-averse (see Problem Set 1).

For $N \leq M$, we can rewrite (3) as

$$V(w,N) = N\left(u(w) - u(w_0)\right) + Mu(w_0).$$

This shows that $V(w,N)$ can be interpreted as the sum of the "baseline utility" $Mu(w_0)$ and the "union surplus" $N\left(u(w) - u(w_0)\right)$. The baseline utility $Mu(w_0)$ is the utility the M union members can ensure without exploiting any union market power by simply working at the competitive wage w_0. The union surplus $N\left(u(w) - u(w_0)\right)$ is the total utility surplus the union generates by raising the wage from w_0 to w for N of its members. The baseline utility $Mu(w_0)$ is not affected by union behaviour (the wage chosen) under the assumption that union membership M and the outside wage w_0 are determined outside the model. Thus, maximising $V(w,N)$ is equivalent to maximising the surplus $N\left(u(w) - u(w_0)\right)$.

Remark 5. We have interpreted equation (3) as writing $V(w,N)$ as the sum of the utilities derived from by the M union members from their employment status *after* union jobs have been distributed among them (and only N are employed in the union sector and $M-N$ are employed elsewhere). Rather than this *ex post* interpretation, we can also provide an *ex ante* interpretation from the perspective of the union members *before* union jobs are distributed (see also Problem Set 1). To this end, we can rewrite (3) into the

sum of M identical "expected utilities",

$$V(w,N) = M\left[\frac{N}{M}u(w) + \frac{(M-N)}{M}u(w_0)\right].$$

To see this, suppose that *ex ante* all M union members have the same chance of securing one of the $N < M$ union jobs (that is, that the allocation of union jobs among the union members is "fair"). Then, each individual union member faces a probability N/M of being allocated to a union job and a probability $(M-N)/N$ of being forced to find employment elsewhere in the economy. So, *ex ante* each union member attached expected utility

$$\frac{N}{M}u(w) + \frac{(M-N)}{M}u(w_0)$$

to being in the union. This *ex ante* expected utility is the same for all M members (under the assumption that they are treated fairly in the union job allocation process). $V(w,N)$ is simply the sum of the M identical *ex ante* utilities of the M union members. If M is given, determined outside the model, then maximizing $V(w,N)$ is equivalent to maximizing the *ex ante* expected utility $V(w,N)/M$ of a representative union member. The same outcome will prevail. The distinction between utilitarian utility and expected utility stressed in Booth (1995) is only relevant if union membership is taken to be "endogenous", that is determined inside the model. We may expect that M increases with the expected utility $V(w,N)/M$ of a representative member, in which case our assumption that M is exogenous is false. We ignore that here.

We first illustrate the union's wage setting problem graphically using the original formulation (1). For any given (constant) utility level $\bar{V}$,

$$V(w,N) = \bar{V} \tag{4}$$

implicitly defines a union *iso-utility (indifference) curve*, a set of points (N,w) that give the same utility $\bar{V}$ to the union. Different choices of $\bar{V}$ give different indifference curves. With a utilitarian union, equation (4) becomes

$$N\left(u(w) - u(w_0)\right) + Mu(w_0) = \bar{V} \tag{5}$$

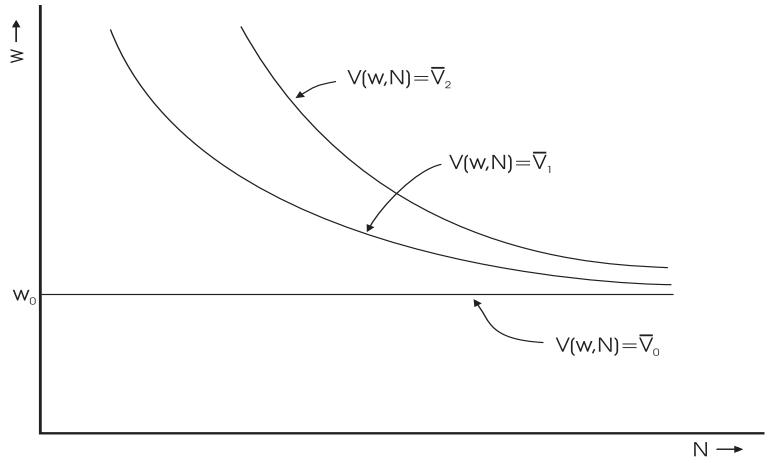
or, equivalently, in terms of the union surplus

$$N\left(u(w) - u(w_0)\right) = \bar{V} - Mu(w_0). \tag{6}$$

Recall that the minimum utility $\bar{V}_0$ the worker can guarantee its workers, by having all employed at the competitive wage w_0, equals $Mu(w_0)$, with a union surplus of 0. Thus, the corresponding indifference curve is given by the points (N,w) that satisfy

$$N\left(u(w) - u(w_0)\right) = \bar{V}_0 - Mu(w_0) = 0.$$

Figure 2: Union indifference curves for utility levels $\bar{V}_2 > \bar{V}_1 > \bar{V}_0$



This holds for all N if $w = w_0$. In this case, union members are indifferent between working in the union sector and working elsewhere because the wages are the same, and the union does not care about N. Thus, on each point on the horizontal line at w_0 in Figure 1, the union has utility $\bar{V}_0$. Thus, this line is (part of) the indifference curve corresponding to minimal utility $\bar{V}_0$. Note that the union utility is also $\bar{V}_0$ if $N = 0$, for all wages w. This corresponds to the vertical axis. This is trivial: if union employment is 0, all union members work elsewhere and the union does not care about the union wage w.

Above the horizontal line at w_0 (for $N > 0$), union utility is higher than $\bar{V}_0$, because the union prefers higher union wages at any union employment level $N > 0$. To to right of the vertical axis (for $w > w_0$), union utility is higher because the union prefer higher union employment at any union wage $w > w_0$.

To each utility level $\bar{V} > \bar{V}_0$ corresponds an indifference curve. Figure 2 plots indifference curves for utility levels $\bar{V}_2 > \bar{V}_1 > \bar{V}_0$. Indifference curves further out from the origin correspond to higher utility levels, because higher wages and employment give higher utility. The indifference curves are downward sloping and convex to the origin. Indifference curves corresponding to different utility levels do not intersect (why?). In particular, note that this implies that the indifference curves for utility levels $\bar{V}_1$ and $\bar{V}_2$ lie to the right of the vertical axis and above the horizontal line at w_0 (on both of which union utility is $\bar{V}_0$).

Technical details 1. We can formally check that the indifference curves are download-sloping and convex to the origin by taking the *total differential* of (5), which gives

$$(u(w) - u(w_0))\, dN + Nu'(w)dw = d\bar{V}. \tag{7}$$

Equation (7) describes how small changes dN, dw and $d\bar{V}$ in resp. union employment, the union wage and union utility are related according to equation (5). The right-hand side of equation (7) is the change in utility. The left-hand side consists of two terms. The first term is the change in utility induced by a change dN in union employment. It equals the union employment change dN times the surplus $u(w) - u(w_0)$ a single union member derives from employment in the union sector. The second term is the change in union utility caused by a dw change in wages for the N union employed. This equals marginal utility $u'(w)$ of a union employed, which is the change in utility of a union employed in response to a dw change in union wages, times the number N of union employed. Note that we have throughout imposed that M and w_0 do not change ($dM = dw_0 = 0$) by omitting terms related to changes in M and w_0.

Along any of the indifference curves, $d\bar{V} = 0$ (why?), and changes in union wages and employment are related according to

$$(u(w) - u(w_0))\, dN + Nu'(w)dw = d\bar{V} = 0.$$

Reorganizing this equation gives

$$\frac{dw}{dN}\rfloor_{d\bar{V}=0} = -\frac{u(w) - u(w_0)}{Nu'(w)} < 0. \tag{8}$$

I have added "$\rfloor_{d\bar{V}=0}$" to make sure that we do not forget that this equation holds only along an indifference curve (that is, with $d\bar{V} = 0$). $(dw/dN)\rfloor_{d\bar{V}=0}$ is the slope of the indifference curve. It is negative (for $w > w_0$), because $u'(w) > 0$ and $u(w) - u(w_0) > 0$. We have given the intuition for this result above.

Moving down on any indifference curve, w decreases and N increases. Then, $u(w) - u(w_0)$ decreases. If the workers are risk-averse, $u'(w)$ increases; if they are risk-neutral, $u'(w)$ is constant (why?). So, in either case the ratio in the right-hand side of (8) decreases (the argument can be extended to include risk-loving workers). Thus, moving down on any indifference curve $(dw/dN)\rfloor_{d\bar{V}=0}$ increases towards 0. This proves that the indifference curves are convex to the origin.

Finally, note that (6) can be rewritten as

$$u(w) - u(w_0) = \frac{\bar{V} - Mu(w_0)}{N}.$$

For a given $\bar{V} - Mu(w_0) > 0$, $w \downarrow w_0$ if $N \to \infty$ and $w \to \infty$ if $N \downarrow 0$. Here, we forget that the union is indifferent about union employment if $N > M$. This would translate into horizontal indifference curves for $N > M$ (rather than indifference curves that continue to fall towards $w = w_0$). As said before, we will ignore this in most of our analyses.

Figure 3: The monopoly-union utility-maximization problem

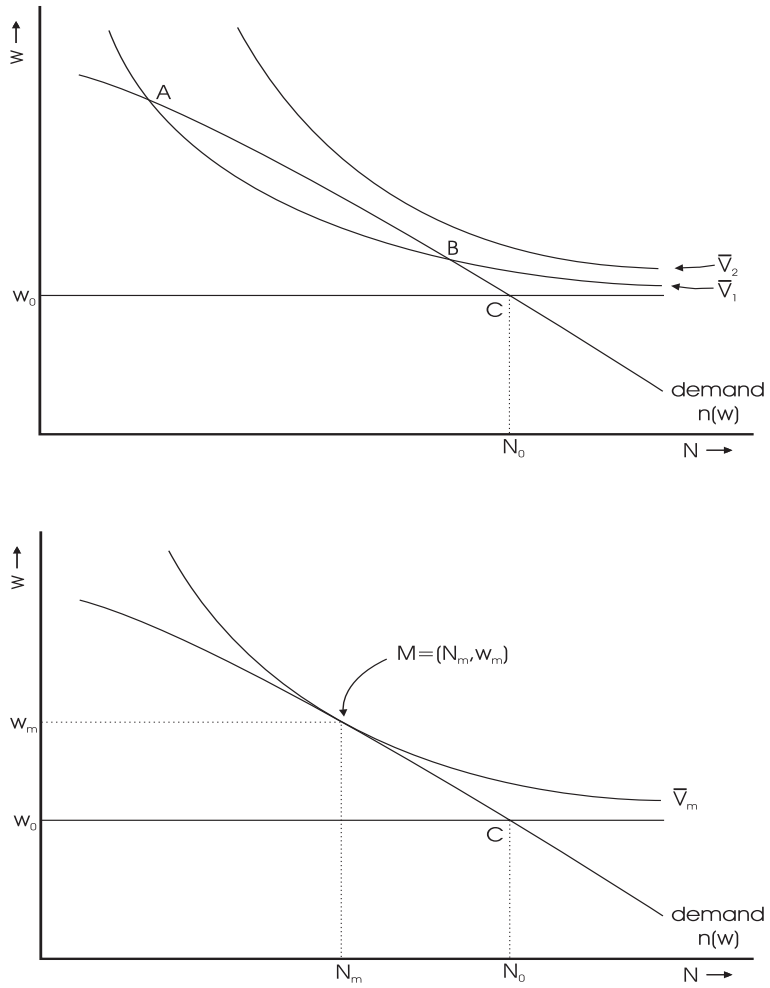


Figure 3 graphically illustrates the union utility maximization problem. The top panel is just Figure 2 with the labour-demand curve from Figure 1. The union will try to move from the utility level $\bar{V}_0$ in competitive equilibrium C to a higher utility level at a higher indifference curve. It could try to attain a utility level of $\bar{V}_2$ by moving to any point on the highest indifference curve shown. However, this curve has no point in common with the labour-demand curve. So, this utility level cannot be reached without violating the right-to-manage constraint. In contrast, the union could attain a utility level of $\bar{V}_1$, which is larger than $\bar{V}_0$ but smaller than $\bar{V}_2$, by moving to either point A or B. Both points are on both the labour-demand curve and the indifference curve corresponding to utility level $\bar{V}_1$.

However, neither of A and B would be the maximum utility the union can attain under the right-to-manage restriction. If we move to a slightly higher indifference curve, corresponding to a slightly higher utility level, this will still have two points of intersection with the labour-demand curve, both between A and B. If we move to higher and higher indifference curves, the intersection points come closer and closer. Either of the points on the labour-demand curve between A and B gives the union more utility than A and B and meets the right-to-manage constraint.

There is some indifference curve that is tangent to the labour-demand curve, that is has just point in common with the labour-demand curve. This indifference curve is drawn in the bottom panel. The tangency point is denoted by $M = (N_m, w_m)$ and the corresponding utility level by $\bar{V}_m$. If we move to an even higher utility level (like $\bar{V}_2$), the corresponding indifference curve can not have any points in common with the labour-demand curve. Thus, the point M maximizes union utility subject to the right-to-manage constraint.

Formally, the union maximization problem can be most easily solved using formulation (1). The wage w_m chosen by the union is the solution w to the first-order condition[3] (using the chain rule)

$$\frac{dV(w, n(w))}{dw} = V_w(w, n(w)) + V_N(w, n(w))n'(w) = 0. \tag{9}$$

Here, $V_w(w, n(w))$ and $V_N(w, n(w))$ are the partial derivatives of $V(w, N)$ with respect to respectively w and N, both evaluated at $N = n(w)$. Thus, $V_w(w, N)$ is roughly the change (increase) in union utility in response to a unit increase in the union wage w, keeping employment N fixed. $V_N(w, N)$ is approximately the change (increase) in union utility in response to a unit increase in union employment N, keeping the wage

[3]Utility maximisation also requires that an additional, "second-order" condition is satisfied: if a maximum (rather than a minimum) is attained at w_m, then marginal indirect union utility should be decreasing at w_m. We will not bother to check that here.

w constant. The left-hand side (9) is the total derivative of $V(w, n(w))$ with respect to w, that is, the change of $V(w, n(w))$ in response to a small increase dw in w. This total derivative includes both (i) the positive direct effect $V_w(w, n(w))$ of the increase in wages of those employed in union jobs and (ii) the negative indirect effect $V_N(w, n(w))n'(w)$ of the induced negative change $n'(w)$ in employment. At the wage w_m, these two effects should be balanced so that a further increase (or decrease) in wages does not improve union utility.

Using the definition (3) of V (for $N < M$), the first-order condition (9) for w_m can be written as

$$n(w_m)u'(w_m) + (u(w_m) - u(w_0))\, n'(w_m) = 0$$

or, after some rearranging,

$$\frac{w_m u'(w_m)}{u(w_m) - u(w_0)} = -\frac{w_m n'(w_m)}{n(w_m)}. \tag{10}$$

The solution w_m to this first-order condition is the wage chosen by the union. The union employment level N_m follows from the labour input choice of the firm, that is $N_m = n(w_m)$. The left-hand side of (10) is the elasticity of the surplus per union worker $u(w_m) - u(w_0)$ with respect to the union wage w_m. The right-hand side of (10) is minus the wage elasticity of labour demand at the wage w_m. This re-establishes the intuition given above. We have seen that the union effectively maximizes the (aggregate) union surplus (over the "baseline" union utility $Mu(w_0)$) $N(u(w) - u(w_0))$ with respect to the wage w subject to the right-to-manage contraint $N = n(w)$. This is equivalent to maximizing the "indirect" union surplus $n(w)(u(w) - u(w_0))$ with respect to w. If the wage w demanded by the union increases by 1%, the first, labour-demand factor in this surplus decreases by (approximately) $-wn'(w)/n(w)\%$. The second factor, the individual union-worker surplus, increases by (approximately) $w_m u'(w)/(u(w) - u(w_0))\%$. At the wage w_m that maximises the union surplus, these two effects should just cancel so that the union would not benefit from changing its wage demands. This is exactly what the first-order condition (10) requires.

Technical details 2. We have argued that, graphically, (N_m, w_m) is the point of tangency of the labour-demand curve and a union indifference curve. Rearranging (10), we find that the first-order condition alternatively reads as

$$-\frac{n(w_m)u'(w_m)}{u(w_m) - u(w_0)} = n'(w_m).$$

From (8), we know that the left-hand side is the (inverse) slope $dN/dw\rfloor_{d\bar{V}=0}$ of the indifference curve at the point $(N, w) = (n(w_m), w_m)$ on the labour-demand curve. The

right-hand side is the slope of the demand curve at the wage w_m. Thus, the first-order condition indeed requires tangency of an (inverse) indifference curve and the labour-demand curve. Equivalently, it requires tangency of an indifference curve and the inverse labour-demand curve as drawn in the figures.

Example 1. Suppose that labour demand n is such that its wage-elasticity does not vary with the wage level. Denote the (constant) absolute value of the wage-elasticity of labour demand by ε, that is

$$\varepsilon := -\frac{wn'(w)}{n(w)},$$

and assume that $\varepsilon > 1$. This implies that $n(w) = \delta w^{-\varepsilon}$ for some $\delta > 0$ (check!).

Furthermore, let utility be linear, so that, for example, $u(w) = w$ (nothing would change if we would take $u(w) = \alpha w + \beta$, for $\alpha > 0$, instead). The wage-elasticity of $u(w) - u(w_0)$ then equals

$$\frac{wu'(w)}{u(w) - u(w_0)} = \frac{w}{w - w_0}.$$

Note that this elasticity tends to ∞ for $w \downarrow w_0$ and decreases towards 1 for $w \to \infty$.

Therefore, for w close to w_0 the percentage increase in the surplus per union employed in response to an increase in the union wage far outweighs the percentage decrease in union employment. Thus, for w close to w_0 the union surplus $n(w)(u(w)-u(w_0)) = \delta w^{-\varepsilon}(w-w_0)$ increases if the wage w increases. Because the wage-elasticity of the surplus per union employed decreases with w and the labour-demand elasticity is constant, the relative (percentage) increase in the aggregate union surplus becomes smaller and smaller as the wage w increases. At the point $w = w_m$ where (10) holds, that is

$$\frac{w_m}{w_m - w_0} = \varepsilon,$$

the relative increase in surplus per union employed and the relative decrease in union employment are in balance. At wages above w_m, the labour-demand effect outweighs the individual surplus effect. Thus, if the union would increase the wage above w_m, the aggregate union surplus would fall.[4] In this example, we can solve explicitly for w_m, which is given by

$$w_m = \frac{\varepsilon}{\varepsilon - 1} w_0 > w_0.$$

It immediately follows that w_m is increasing in w_0. Thus, if the outside market wage w_0 is higher then the wage w_m set by the union is higher as well. Union employment N_m

[4] Note the close relation of this argument to the second-order condition discussed before.

is lower, as we move up the downward sloping labour-demand curve (which, by itself, is unchanged). This makes sense, as the union is willing to accept more union unemployment (that is, employment of its members outside the union sector) if market conditions are better.

Next consider a change in the wage-elasticity of labour demand. It is easy to check that the union wage w_m decreases in ε. So, if sectoral labour demand is more elastic, then the union wage w_m is lower (and closer to w_0). This makes sense: if labour demand is more elastic, then high union wage demands have stronger negative union employment effects. Therefore, the union is less eager to push for high union wages in this case. The corresponding effect on union employment is not immediately clear. A change in ε is a change in the labour demand curve. Therefore, we cannot simply argue in terms of moving along a given labour-demand curve. The change in ε corresponds to a rotation in the (inverse) labour-demand curve. The union employment effect depends on around which point we rotate the labour-demand curve, that is, on how we change δ together with ε. For an example, see Problem Set 1.

Remark 6. Note the close analogy between the example and product-market monopoly theory. The monopolist union maximises a surplus $n(w)(w-w_0)$ which is now very similar to a monopoly profits function.

Remark 7. The last result in the example may explain why crafts unions (monopolising the supply of very specific and hard-to-substitute skills) have traditionally been more effective than general unions (of course, crafts unions also had effective ways of monopolising supply; see Booth (1995), Sect. 3.4).

Remark 8. In the light of the previous remark, you may want to study the determinants of the elasticity of the sectoral demand for labour at this point (elasticity of substitution with other inputs, share of union labour in production, etcetera). See e.g. Bosworth et al. (1996), Section 7.5.

3.2.3 Efficient-bargaining model

Before we can discuss the efficient-bargaining model, we first have to zoom in on the firm's labour-demand problem. Rather than simply assuming a labour-demand function n, we derive it from profit-maximising behavior of the (single) union firm.

Let revenues when employing N workers be $R(N)$. We assume that $R(0) = 0$ and that marginal revenues $R'(N)$ are large for small w, decrease with N ($R''(N) < 0$) and become very small or negative for large enough N. Costs of employing N workers at a wage $w > 0$ are wN, so that profits, as a function of wages w and employment N, are

$$\pi(w, N) = R(N) - wN.$$

For given wages $w > 0$, marginal costs are simply w for all N, so that marginal revenues are larger than marginal costs for small N and smaller than marginal costs for large N. Thus, for a given wage, profits are first increasing in employment and then decreasing in employment. Our simplifying assumptions therefore imply that there is a unique profit-maximizing employment level $n(w)$ for each wage.

[Draw a graph of revenues and costs as functions of N for a given wage w that satisfies the assumptions above. You may also want to draw profits as a function of N for given w. Profits are maximised where the gap between revenues and costs is widest. The assumptions guarantee there is a unique such point.]

Remark 9. Revenues are output price times output sold. R' is usually called the "marginal revenue product of labour" (MRP). It is also the value of the marginal (physical) product of labour (that is, the output price times the marginal physical product of labour) if the firm is a price-taker in the product market in which it operates. In that case, the firm does not take into account any effects of its own labour-demand choice on the output price and revenue changes are simply the given price times output changes. If the firm has market power in the product market, changes in employment will typically both affect output price and physical output and the equivalence breaks down.

Formally, if the firm takes wages w as given, it will choose an employment level N that maximises profits, that is, satisfies the first-order condition[5]

$$\frac{\partial \pi(w,N)}{\partial N} = R'(N) - w = 0 \Longleftrightarrow R'(N) = w.$$

This simply says that marginal revenues should equal marginal cost. The employment level N that solves the first-order condition is labour demand at the wage w and is again denoted by $n(w)$. Thus, the labour-demand curve is now implicitly determined by the first-order condition $R'(n(w)) = w$. If w goes up, then $n(w)$ should go down because R' is decreasing in its argument. Thus, labour demand $n(w)$ is decreasing in the wage w as before.

Alternatively, substitute $w = n^{-1}(N)$ in the first-order condition $w = R'(n(w))$ to find that inverse labour-demand is $n^{-1}(N) = R'(n(n^{-1}(N))) = R'(N)$. This inverse labour demand $n^{-1}(N)$ can be interpreted as the wage the firm would be willing to pay at most for an extra worker if it is already employing N workers. Because $n^{-1}(N) = R'(N)$ the firm is willing to pay at most a wage equal to the MRP and the (inverse) labour-demand

[5]Note that $\partial^2 \pi(w,N)/\partial N^2 = R''(N) < 0$, so that the second-order condition is satisfied by assumption. This is closely related to the verbal argument in the previous paragraph: marginal profits are positive for low N, fall with employment and eventually become negative.

curve, as drawn in Figure 1, is also the MRP curve. For a given wage w and at low levels of employment, the MRP is larger than the marginal cost w_0 and the firm makes a profit equal to the difference between the MRP and w by hiring an additional worker. If N increases, MRP and this profit margin fall. The firm hires workers up to $n(w)$, at which point the MRP $R'(n(w))$ equals marginal cost w.

Technical details 3. The labour-demand curve itself is

$$n(w) = (R')^{-1}(w),$$

where $(R')^{-1}$ is the inverse of the marginal revenue function (which exists because $R'' < 0$). We can differentiate both sides of $R'(n(w)) = w$ with respect to w to get (using the chain rule)

$$R''(n(w))n'(w) = 1,$$

which boils down to

$$n'(w) = \frac{1}{R''(n(w))} < 0.$$

This shows that the labour-demand curve is downward sloping. Alternatively, we can show that the inverse labour-demand curve is downward sloping:

$$\frac{dn^{-1}(N)}{dN} = R''(N) < 0.$$

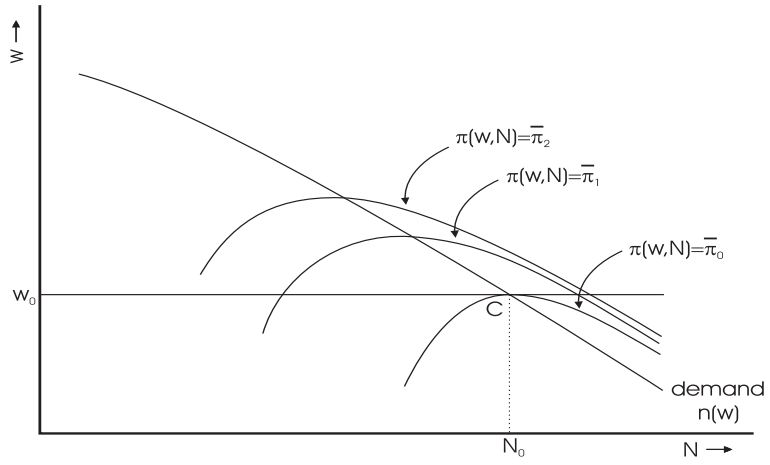
It is useful to have a second look at the firm's problem, based on "iso-profit curves". An iso-profit curve corresponding to a profit level $\bar{\pi}$ consists of all points (N, w) such that $\pi(w, N) = \bar{\pi}$ [compare to the definition of union indifference curves]. It is easy to verbally analyze the shape of iso-profit curves.

(i). First, consider increasing employment starting from a point (N, w) on an iso-profit curve. If N is small enough, an increase in employment for a given wage leads to an increase in profits. In order to keep profits constants the wage have to increase as well. Thus, for low N, employment and wages increase (or decrease) together along any iso-profit curve.

(ii). Similarly, for large N, profits decrease if employment increases. This has to be compensated by a decrease in the wage. Thus, employment and wages are inversely related along any iso-profit curve for large N.

(iii). Finally, if we start from a point $(N, w) = (n(w), w)$ on the labour-demand curve, a small change in employment does not change profits, by the first-order condition for

Figure 4: Profit maximisation and iso-profit curves



profit maximisation. So, we do not have to change the wage along with employment in order to keep profits constant and stay on the iso-profit curve. Thus, iso-profit curves are flat (with employment on the horizontal axis and wages on the vertical axis) where they intersect the labour-demand curve. They are increasing (as in (i)) to the left of the labour-demand curve and decreasing (as in (ii)) to the right of the labour-demand curve.

Technical details 4. The total differential of the iso-profit curve $R(N) - wN = \bar{\pi}$ is

$$(R'(N) - w)\, dN - N dw = d\bar{\pi} = 0. \tag{11}$$

Equation (11) formally relates small changes dN and dw in, respectively, N and w along the iso-profit curve (as in the verbal argument above). The first term in (11) is the effect of a small change dN in employment on profits (for given w): revenues increase by $R'(N)dN$ and costs by wdN. The second term is the effect of a small change dw in wages (for given N) on profits: the total wage bill increases by Ndw. The net effect of these two changes in profits should, by definition, be zero along the iso-profit curve: $d\bar{\pi} = 0$.

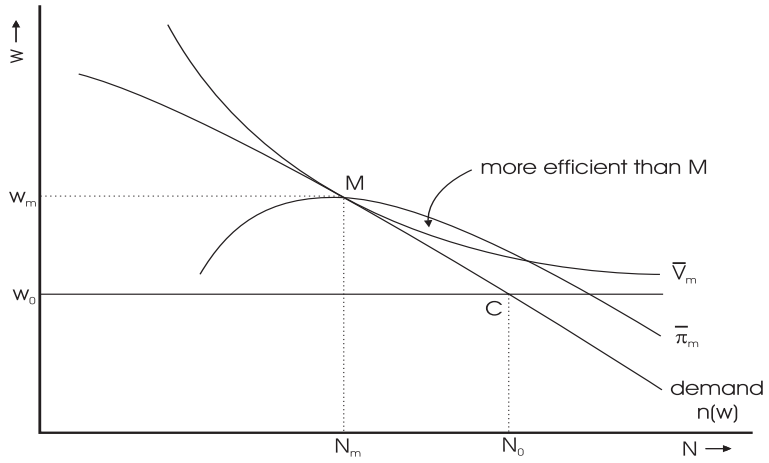
Rearranging (11) gives (in notation introduced for union indifference curves before)

$$\frac{dw}{dN}\Big\rfloor_{d\bar{\pi}=0} = \frac{R'(N) - w}{N} = \begin{cases} = 0 & \text{if } N = n(w) \\ > 0 & \text{if } N < n(w) \\ < 0 & \text{if } N > n(w), \end{cases}$$

which confirms our earlier, verbal conclusions.

Figure 4 plots some iso-profit curves. Like indifference curves, iso-profit curves cannot intersect (why?). Higher iso-profit curves correspond to lower profits (because higher

Figure 5: Inefficiency of the monopoly-union outcome



curves correspond to higher wages at any given employment level). Thus, the profit-maximisation problem of the firm can be cast in terms of iso-profit curves as follows. The firm want to end up on the lowest iso-profit curve (corresponding to the highest profit level) possible. Taken some wage w_0 as given, the firm will try to find the lowest iso-profit curve that contains a point (N_0, w_0) for some N_0. In other words, the firm will try to find the lowest iso-profit curve that has a point in common with the horizontal line at w_0. As before, it can easily be argued that this iso-profit curve should be tangent to this horizontal line. If we move to an even lower iso-profit curve, it has no point with a wage w_0. If it intersects the horizontal line at w_0, it is possible to find a lower iso-profit curve that has a point in common with the horizontal line. Thus, the maximum profit is obtained at the point (N_0, w_0) where an iso-profit curve is tangent to the horizontal line at w_0 and therefore is flat. Obviously, (N_0, w_0) is on the labour-demand curve n, so that $N_0 = n(w_0)$. Thus, we again find that the isoprofit curve is flat where it is intersected by the labour-demand curve.

We can use these results to argue that the monopoly union outcome is not Pareto-efficient. An outcome (N, w) is Pareto-efficient if there is no other outcome such that both the firm and the union are better off (that is, such that the firm has higher profits and the union has higher utility). Figure 5 again illustrates the monopoly-union outcome $M = (N_m, w_m)$ as the point of tangency between the labour-demand curve and the union indifference curve at utility level $\bar{V}_m = V(w_m, N_m)$. It also draws the iso-profit curve that cuts through M. Obviously, the corresponding profit level is the monopoly-union profit level $\bar{\pi}_m = \pi(w_m, N_m)$. Now consider the following argument.

(i). Because M is on the labour-demand curve, the iso-profit curve is flat at M.

(*ii*). Because the labour-demand curve is downward sloping and the union indifference curve through M is tangent to the labour-demand curve at M, this indifference curve intersects the iso-profit curve of (*i*) at M.

(*iii*). As a consequence, there are points to the southeast of M (the area indicated in the graph) that lie above the union indifference curve through M and below the firm's iso-profit curve through M. At each of these points, the union has higher utility and the firm has higher profits than in M. Hence, M is not efficient.

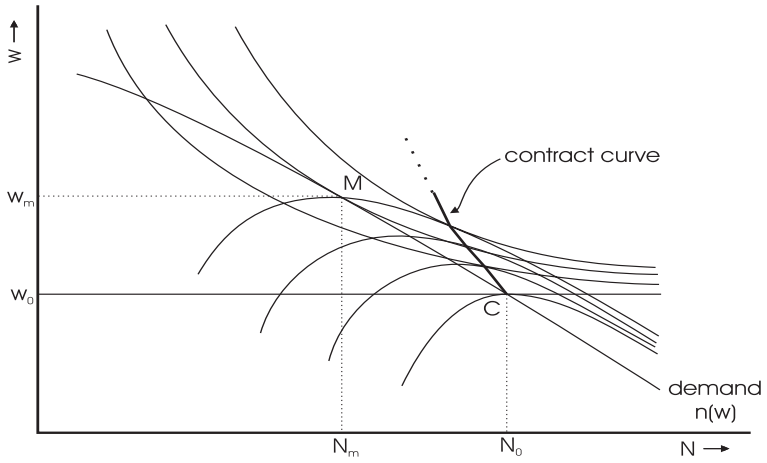
This argument can be repeated for any point (N, w) at which a union indifference curve and a firm iso-profit curve intersect. None of these points is efficient. On the other hand, any point (N, w) at which a union indifference curve and a firm iso-profit curve are tangent is Pareto-efficient. After all, from such points it is not possible to move to a higher indifference curve without decreasing the firm's profits, nor is it possible to move to a lower iso-profit curve without decreasing the union's utility. In other words, it is not possible to "Pareto improve" on such points and they are efficient.

A prime example of a Pareto-efficient outcome is the competitive equilibrium $C = (N_0, w_0)$. After all, we have seen that the horizontal line at w_0 is also the union indifference curve corresponding to utility level $\bar{V}_0 = V(w_0, N_0)$. By profit maximisation, the iso-profit curve through C, corresponding to profit level $\bar{\pi}_0 = \pi(w_0, N_0)$, is tangent to this indifference curve.

There are many— a continuum of— efficient outcomes. For any given utility level $\bar{V}$, an efficient outcome (N, w) maximises profits $\pi(w, N)$ subject to delivering at least utility $\bar{V}$ to the union (that is, $V(w, N) \geq \bar{V}$). After all, if profits are not maximised for the given utility level $\bar{V}$, it is possible to increase profits without decreasing union utility and the outcome cannot be Pareto-efficient. Different given utility levels $\bar{V}$ give different efficient outcomes. Alternatively, we can characterise the efficient outcomes as those points (N, w) that maximise union utility $V(w, N)$ subject to delivering at least a given profit level $\bar{\pi}$ to the firm ($\pi(w, N) \geq \bar{\pi}$). Different profit levels again correspond to different efficient outcomes. The two methods only lead to the same outcome if the iso-profit curve corresponding to $\bar{\pi}$ and the indifference curve corresponding to $\bar{V}$ are tangent. However, each efficient outcome can be characterised in either way. Graphically, either way leads to the conclusion that efficient outcomes are characterised by tangency of a union indifference curve with a firm's iso-profit curve (why?).

Technical details 5. Formally, either approach requires some knowledge of optimisation under constraints. If you have this knowledge, you know that the Lagrangian for e.g. the

Figure 6: Efficient outcomes and the contract curve



second problem is

$$\begin{aligned} L(w, N, \lambda) &= V(w, N) + \lambda \left(\pi(w, N) - \bar{\pi}\right) \\ &= N \left(u(w) - u(w_0)\right) + M u(w_0) + \lambda \left(R(N) - wN - \bar{\pi}\right). \end{aligned}$$

The first-order conditions follow by setting the partial derivatives of $L(w, N, \lambda)$ with respect to w, N and λ equal to 0, which gives

$$\begin{aligned} &u'(w) = \lambda, \\ &u(w) - u(w_0) = -\lambda \left(R'(N) - w\right) \quad \text{and} \\ &\pi(w, N) = \bar{\pi} \end{aligned}$$

respectively. We can substitute the first into the second condition, which gives

$$\frac{dw}{dN}\rfloor_{d\bar{V}=0} = -\frac{u(w) - u(w_0)}{N u'(w)} = \frac{R'(N) - w}{N} = \frac{dw}{dN}\rfloor_{d\bar{\pi}=0}.$$

Together with the third first-order condition $\pi(w, N) = \bar{\pi}$, this is the tangency condition we were after.

We are now ready to discuss the efficient-bargaining model. The idea underlying the efficient-bargaining model is that the firm and the union will not leave any possibilities of mutual gain unused. For example, if they foresee ending up in the inefficient monopoly-union outcome M, they will agree they can both do better by moving away from the labour-demand curve to any of the points in the area between the indifference and isoprofit curves indicated in Figure 5. The efficient-bargaining model takes this idea to the extreme by assuming that the firm and the union will always agree to settle for any of the Pareto-efficient outcomes. This requires that that can write a contract that specifies both wages

and employment. Thus, the efficient-bargaining model is not a right-to-manage model. In this sense, the efficient-bargaining model may have more intellectual appeal than practical importance. See for example the discussion in Booth (1995).

We have already seen that there are many Pareto-efficient outcomes. Not all of these make sense as possible outcomes of the efficient-bargaining model. Some efficient outcomes at very low wages may leave less utility to the union than it can get by not negotiating with the union firm at all and employing all its workers at the outside wage w_0.[6] Other efficient outcomes, at very high wages, may leave too little profit to the firm relative to what it can earn elsewhere in the economy. In sum, only those Pareto-efficient outcomes that meet so called "individual rationality" or "participation" constraints for the firm and the union are possible outcomes of the efficient-bargaining model. We call the set of all the Pareto-efficient outcomes that meet these participation constraints the "contract curve".[7]

Figure 6 plots the contract curve as the Pareto-efficient points (points of tangency between isoprofit curves and indifference curves) that satisfy the participation constraints. The union's participation constraint is that $V(w, N) \geq \bar{V}_0$, the utility level the union can attain by employing all its workers at w_0. The efficient outcome that just attains this utility level, and therefore lies at one end of the contract curve, is the competitive outcome C. The other end of the contract curve is determined by the firm's participation constraint. We will not explicitly discuss that here and it is kept implicit in the figure as well.

Remark 10. The figures are all freehand-drawings. They respect some of the main properties of the model, but are not based on actual numerical model computations. Thus, we should be careful in drawing conclusions from the graphs. The fact that the contract curve in Figure 6 is downward-sloping is not the only possible case, nor the most obvious one. If workers are risk-averse, as is often assumed, the contract-curve is upward-sloping. See Booth (1995), Section 5.4, and the following example.

Example 2. Let again $u(w) = w$. The tangency conditions for efficiency then imply that

$$\frac{dw}{dN}\rfloor_{d\bar{V}=0} = -\frac{w - w_0}{N} = \frac{R'(N) - w}{N} = \frac{dw}{dN}\rfloor_{d\bar{\pi}=0} \iff R'(N) = w_0,$$

so that each efficient outcome satisfies $N = n(w_0) = N_0$. Different wages w correspond to different efficient outcomes (N_0, w). The lowest wage consistent with participation of the union is w_0 and gives the lower end of the contract curve $C = (N_0, w_0)$. The curve is

[6] This involves contracts that force the union to employ a certain number of its members at a below-competitive wage. The union and its members would obviously not settle for such deals.

[7] Alternatively, the set of all Pareto-efficient contracts itself is often called the contract curve.

therefore a vertical line from C upwards to (possibly) some highest wage (that is, lowest profits) consistent with the firm's participation.

The special feature of this example is that union employment is always at its competitive level N_0. Thus, the presence of a union does not lead to an *allocative distortion* in this special case of the efficient-bargaining model. Borjas calls this *strong efficiency*. In general, Pareto-efficient outcomes can correspond to union employment levels different from N_0. We will later see that this implies allocative inefficiencies: labour is not optimally distributed between the union sector and the rest of the economy. This can occur even though the firm and the union in the unions sector restrict attention to efficient outcomes, because they (by assumption) are only interested in Pareto efficiency among themselves and do not take any effects on other agents in the economy into account. In this case of a very small union sector this is rather unimportant: the overall welfare effects of the allocative distortion arising from the behavior of a very small group of agents is (negligibly) small. However, in the two-sector model that follows, in which the union sector is not extremely small, such distortions would have noticeable adverse welfare implications.

So far, the efficient-bargaining model only restricts outcomes to be on the contract curve. Which point on the contract curve we end up with depends on the bargaining strengths of the union and the firm. If the firm is very powerful, it will be able to negotiate an efficient wage-employment contract close to the competitive outcome C. If the union has more bargaining power, it will be able to secure a deal far away from C and closer to the upper end of the contract curve. Two approaches to actually modelling the bargaining process can be distinguished:

(i). the Nash (1950, 1953) axiomatic bargaining model;

(ii). the strategic bargaining model.

The first approach simply requires that the bargaining outcome satisfies certain properties ("axioms") that seem a priori sensible. Axiomatic bargaining theory focuses on a particular set of such axioms that fully pin down a single point on the contract curve. If we all agree that these axioms should be satisfied by any bargaining outcome, we should also agree which outcome prevails: the unique outcome that satisfied the axioms.

In contrast, the second approach does not focus directly on the properties of the bargaining outcome, but rather on the actual bargaining process leading to the outcome. It models this process as a dynamic game in which both parties make offers and counter-offers and acceptance/rejection decisions over time. Using dynamic game-theoretic equilibrium concepts such models can in some cases be solved for a unique equilibrium outcome.

We will not study bargaining models in any detail. For this course, it is sufficient to understand that (i) there is a bargaining problem because efficiency and individual

rationality by themselves only restrict the possible outcomes to a whole range of outcomes, the contract curve, and (*ii*) that bargaining models solve this problem by modeling which single outcome in this range occurs.

Remark 11. It is interesting to note that Nash is indeed the same mathematician that Russell Crowe depicts in "A Beautiful Mind". Nash has not only pioneered axiomatic bargaining theory, but also the game-theoretic equilibrium concepts that underly the strategic approach. Professor Binmore of UCL and coauthors have analysed how the two approaches are related (that is, under which conditions they give the same outcome; Binmore, Rubinstein and Wolinsky, 1986) beyond the early work by Nash.

3.2.4 General right-to-manage model

Recall that in the general right-to-manage model

(*i*). the firm and the union bargain over the sectoral wage and

(*ii*). the firm chooses employment.

As in the efficient-bargaining model, the first, bargaining step can be modeled by either a strategic or, more commonly, an axiomatic bargaining model. Unlike the efficient-bargaining model, the union and the firm only bargain over the wage and leave the right-to-manage to the firm. The firm picks employment on the labour-demand curve, so that bargaining is effectively over points on the labour-demand curve (rather than points on the contract curve) under the relevant participation restrictions. The monopoly-union model is a special case of the right-to-manage model in which the union has all the bargaining power (that is, can set the wage unilaterally) in the first stage. In that case, we end up in the monopoly union outcome M. If the firm has all the bargaining power, it will choose the lowest wage consistent with union participation and we end up in the competitive outcome C. In non-extreme cases the outcome lies in between M and C on the labour-demand curve.

3.3 Two-sector model

Reading:

- Booth (1995), Ch. 3
- Relevant sections of textbooks.

3.3.1 Introduction

Consider the following initial observations on allocative efficiency.

(i). In the competitive model, the MRP in the (what will later be a) union sector equals the outside wage w_0, and therefore the MRP in the (competitive) rest of the economy. Thus, aggregate revenues cannot be increased by reallocating workers from the (later) union sector the rest of the economy or *vice versa*. The MRP lost in the union sector by removing one worker would just be gained elsewhere in the economy. Labour is allocated efficiently between the sectors.

(ii). With strong efficiency, as introduced in the previous section, employment and therefore the MRP in the union sector do not change after unionisation. So, the union sector's MRP is still w_0 and equal to the MRP elsewhere. There is allocative efficiency.

(iii). In general, employment and therefore the MRP changes in the unionised sector. In the monopoly-union model, for example, union employment decreases and the union MRP and wage are larger than w_0, the MRP in the rest of the economy. Aggregate revenues could be increased by removing a worker from outside the union sector (at a revenue cost of w_0) to the union sector (yielding revenues equal to the union sector's MRP). In this case, union employment is too low. In general, there is *allocative inefficiency* if union-sector employment is not equal to the employment level that would have prevailed in a competitive market.

(iv). In the case of a very small union sector, the effect of this possible allocative inefficiency on aggregate revenues is negligible. We will now discuss a two-sector model in which it is not.

3.3.2 The competitive benchmark

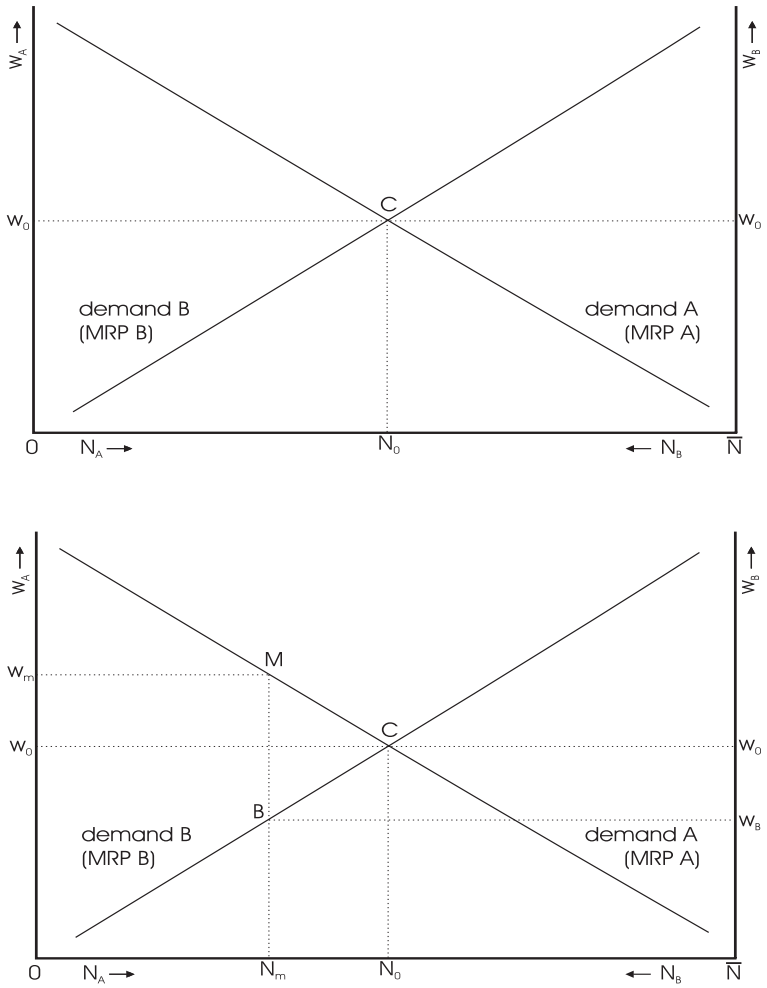
The primitives of the model are the following.

(i). The economy consists of two sectors, the (later) union sector A and an other sector B. Note that sector A is now not very small.

(ii). Labour demand is given by downward-sloping sectoral labour-demand curves n_A and n_B.

(iii). Aggregate labour supply is fixed at $\bar{N}$. Workers only choose in which sector to work, as before.

Figure 7: The two-sector model



In competitive equilibrium, in which sector A is not unionised,

(i). wages in sector A and B should be equal in equilibrium (otherwise, all workers would flock to the sector with the highest wage) and

(ii). aggregate demand $n_A(w_0) + n_B(w_0)$ at the (common) equilibrium wage w_0 should equal aggregate labour supply $\bar{N}$.

The top panel in Figure 7 illustrates this equilibrium.

(i). It plots both labour-demand curves. Labour demand in sector A is plotted in the usual way. Labour demand in sector B is plotted from right to left, with origin $(\bar{N}, 0)$.

(ii). At the wage w_0, both labour-demand curves intersect, at employment level N_0 in sector A. Employment in sector B is then $\bar{N} - N_0$, because labour demand in B is measured from the right origin $(\bar{N}, 0)$. Thus, at w_0 aggregate labour demand equals $\bar{N}$ and there is equilibrium.

(iii). At any other wage, the labour-demand curves do not intersect, aggregate labour demand is either smaller (if the wage is higher than w_0) or larger (if the wage is lower) than aggregate labour supply.

Because wages and MRPs are the same in sectors A and B, the allocation of labour between A and B is efficient.

(i). Revenues in sector A equal the surface below sector A's (inverse) labour-demand curve between 0 and N_0. Of that revenue, $w_0 N_0$ is paid as wages to sector A's workers. This is also the surface of the rectangle below the horizontal line at w_0 between 0 and N_0. The surface of the area between sector A's (inverse) labour demand curve and the horizontal line at w_0 (between 0 and N_0) are sector A's profits.

(ii). Revenues in sector B equal the surface below sector B's (inverse) labour-demand curve between N_0 and $\bar{N}$, etcetera.

[You may want to shade all relevant areas. My lack of Coreldraw skills held me back here.]

Remark 12. You may be puzzled by the claim that the surfaces below the (inverse) labour-demand curves equal revenues. We can gain some insight in this fact by returning to the competive small-sector model of Figure 1. In this model, the surface of the area

below the (inverse) labour-demand curve up to N_0 equals revenues $R(N_0)$ in competitive equilibrium. Intuitively, this follows from the fact that the inverse labour-demand curve is also the MRP curve. The surface below it is the roughly the sum of the MRPs for each incremental unit of labour between 0 and N_0. Formally, the surface of the area below the (inverse) labour-demand curve between 0 and N_0 equals

$$\int_0^{N_0} R'(N)dN = R(N_0).$$

Note that the surface of the quadrangle below the horizontal line at w_0 between 0 and N_0 is simply the total wage bill $w_0 N_0$ in competitive equilibrium. Thus, the surface of the area between the inverse labour demand curve and the horizontal line at w_0 equals the firm's profits $\pi(w_0, N_0) = R(N_0) - w_0 N_0$ in competitive equilibrium.

3.3.3 Unionisation in the two-sector model

Now suppose that a union monopolises labour supply to sector A.

(i). The union manages to get a wage $w_m > w_0$ and employ $N_m = n_A(w_m)$ of its members in the union sector. This corresponds to a move from C to M in the lower panel of Figure 7.

(ii). All $\bar{N} - N_m$ workers not employed in sector A supply labour to sector B.

(iii). In equilibrium, the wage in sector B has to drop from w_0 to w_B to equalize labour supply and demand in sector B.

Because $w_m > w_B$, the MRP in sector A is larger than the MRP in sector B and there is allocative inefficiency.

(i). Revenues in A are now the surface below sector A's (inverse) labour-demand curve between 0 and N_m.

(ii). Revenues in B are the surface below sector B's (inverse) labour-demand curve between N_m and $\bar{N}$.

(iii). The loss in aggregate revenues relative to the competitive case is equal to the surface of the "triangle" CMB (why?).

Remark 13. In this model, unionisation of sector A affects the outside opportunities of the union members elsewhere in the economy (sector B): the wage in sector B drops from w_0 to w_B. This implies that

(*i*). the union models of the previous lectures do not apply without change, because the union will take this effect on the outside options for its displaced members into account;

(*ii*). there are various ways of defining union wage gaps and gains in this model, like $w_m - w_0$ (Booth's "wage gain") and $w_m - w_B$ (Booth's "wage gap"). We return to this when we discuss the empirical analysis of union wage effects. See Booth (1995), Section 3.2.1.

3.4 Empirical analysis of wage and employment effects of unions

Reading:

- Booth (1995), Ch. 6
- Relevant sections of text books
- Some (not much) background reading on linear regression analysis etc.

3.4.1 Crash introduction to linear regression

[I have given a quite extensive introduction to linear regression in class. I have not included it here, but only give a list of issues discussed. See Problem Set 2 for the type of things you have to take from this (and be able to reproduce). For (too many) details you can also check my econometrics lecture notes posted on the course web site or an introductionary econometrics text book.]

Brief overview of issues covered in class:

(*i*). *Simple (linear) regression*

(*a*) Relating two variables; sample; scatter plot of data; linear model; regression as "best predictor"; estimation (by ordinary least squares; "best fit"); residual; interpretation of estimator/estimates (in particular also: using logs to model relative changes).

(*b*) Examples of applications: returns-to-schooling (earnings and schooling), union wage-effects (wage and union status or density)

Regression of log earnings on years of schooling gives estimate that, multiplied by 100%, can under some assumptions be interpreted as the percentage change in earnings for one extra year of schooling.

(*ii*). *Multiple (linear) regression*

(a) Relating more than 2 variables, for example 3.

(b) Same issues (sample, model, estimation, etc.) as with simple regression.

(c) Interpretation in terms of *partial* relations (*ceteris paribus*); using multiple regression to control for selection, etcetera.

(d) Examples: earnings, unions and schooling; earnings, schooling and experience

Regression of log wage on years of schooling and years of experience uncovers (under some assumptions) the relative effect of one extra year of schooling on earnings holding experience constant, etcetera.

(e) Including interactions of regressors: heterogeneous effects, etcetera.

(iii). *Simultaneous equations*

[Skip technical details in Booth; see also Problem Set 2.]

(a) Union status dependent on (heterogeneous) union-wage effects: endogeneity/simultaneity problem.

(b) Extend model into simultaneous equations model by adding discrete-choice ("logit") model for union status.

3.4.2 Empirical analysis of union wage-effects

[I have not gone into too much detail in class. Read Booth (1995) , Ch. 6.]

We are interested in measuring the effect of unions on wages. First, suppose we have individual data on wages W and union status D. We can compare wages of union members ($D = 1$) to wages of non-members ($D = 0$) and typically find that union wages are higher (on average) than non-union wages. Is that a satisfactory final answer to our problem? No, because of the following (potential) problems.

(i). We may not have used the appropriate measure of union status D (e.g., membership versus coverage).

(ii). We cannot observe both union and non-union wages for single individual, so cannot measure individual contrast between potential union and non-union wages directly.

Instead, we have to compare wages between (union and nonunion) individuals. This introduces some new problems.

(a) Individuals differ in other dimensions than union-status as well.

For example, suppose that union members are, on average, more skilled than non-union members. Then, the wage difference between union members and non-members confounds union effects and returns-to-skill (that is, skilled workers earn more than unskilled workers anyhow, irrespective of union status).

(b) Even more worrying, union wage-effects may differ between individuals (agents) and agents may act on that by self-selecting into unions: endogeneity problem. Suppose for example that there are two equally sized groups of agents, called A and B. Let wages depend on group and union status as follows.

group	union member	non-member
A	£10	£5
B	£11	£11

If there is a small fee for joining the union, workers in group A will join and earn £10 and workers in group B will not joint and earn £11. In the data, union members (all in group A; $W = £10$) will earn less than non-members (all in group B; $W = £11$). A simple comparison of union and non-union wages therefore does not reflect the fact that half of the workers earns twice as much in a union job than they would have earned in a non-union job and that the other half would have earned the same in a union job as they are earning in their non-union jobs. The problem is that those with relatively high earnings have relatively little incentive to join unions (in this example), so that a selected group of workers ends up in unions.

(iii). We cannot measure equilibrium effects.

Recall the discussion on wage effects on the non-union sector in the two-sector model (wage gap versus wage gain).

The selection problems can to some extent be dealt with by using multiple regression techniques. If we can regress wages on all determinants of wages that the agent uses to decide on union membership, we can isolate the (partial) effect of union membership on wages. Again, we can allow for observed heterogeneity in union wage-effects by including interactions of union status and other observed wage determinants. Broadly, the empirical findings of such regression analyses are (but read Booth, Ch. 6)

(i). a mildly positive and heterogeneous union wage-effect;

(ii). that most of the variation in wages between individuals is not explained by union status and other individual characteristics (see also Problem Set 2).

Individual panel data provide observations of wages and union status for a set of individuals over time. These data are richer and allow, for example, to exploit variation over time for a single individual (by comparing wages in periods where an individual was member with wages in periods in which the individual was not). However, similar problems as in the cross-section case arise in a slightly different form (e.g., what causes the union status of an individual to change over time?).

Aggregate data can to some limited extent be used to deal with equilibrium effects. Cross-sectional aggregate data would typically provide average wage and union status (union density, coverage rate, etcetera) data for sets of sectors, regions, countries. Time-series would provide such data over time for a single sector, etcetera. Aggregate panel data combine both dimensions. Aggregate data can be used to assess the "overall" effect of the level unionisation on average wages, but also on wage dispersion. Similar problems arise:

(i). What is the appropriate measure of unionisation (e.g., union density versus union coverage)?

(ii). There are other causes of average wage differences etc. between sectors, countries, over time than differences in unionisation.

This can to some extent again be solved by using multiple regression techniques.

Also, the link of the effects measured in aggregate data to the underlying theory is less clear (just see Booth, Ch, 6!).

Studies of equilibrium (macro) and micro effects ideally uses a combination of individual and aggregate data. Such data allow (again under some assumptions) to measure the effect of *aggregate* union (density, coverage) conditions on the wages of *individuals* in both union and non-union jobs.

Summary of main points:

(i). Data: cross-sectional, time-series and panel; individual and aggregate.

(ii). Compare individuals, firms, sectors, regions, times with different union status to assess the effects of unionisation. Use multiple regression to isolate union effects from other differences over individuals, sectors, regions, time. Include equation for union membership etc. to analyze endogeneity (but: relation weak because of free-riding, institutional constrains; see Problem Set 3).

(iii). Results:

(a) positive union wage gaps;

(*b*) most individual variation in wages not explained by union and other characteristics;

(*c*) within-sector wage dispersion lower in unions sectors;

(*d*) need aggregate variation and data on union and non-union wages to say something about effect of unions on non-union sector;

3.5 Union membership, source of monopoly power, insider-outsider theory, strikes

Reading:

- Booth (1995), Sec. 3.4; parts of Ch. 4 and 6

[Apart from what has been discussed so far, I have skipped over this and referred you to the relevant reading in Booth. You are supposed to study this yourself. See also Problem Set 3. You are not required to study insider-outsider theory beyond what you may encounter in Booth Ch. 1–7 (in particular, you do not have to study Booth, Ch. 8).]

4 Minimum wages

[We have not discussed minimum wages in class and you do not have to study minimum wages for the exam either.]

5 Human-resource systems

5.1 Overview

Overview of reading material for this part:

- Milgrom and Roberts, Ch. 10-13
- Booth, Sec. 3.2.3, 3.3 and Ch. 7 (unions, efficiency and organisation)
- Borjas, Ch. 12 (incentives) and Sec. 7.9 (training)
- Bosworth et al, Ch. 18-22 and Sec. 16.4 (training)
- Polachek and Siebert, Ch. 9 and Sec. 4.4 (training)
- Some text on efficiency wages, e.g. Milgrom and Roberts, part of Ch. 8
- Cully et al (1999)

(*i*). So far, focused on "monopoly-face" of unions:

(a) unions raise wages in union sector above competitive level;

(b) this introduces allocative inefficiency in the two-sector model.

Remark 14. This assumes, as we did throughout, that the union monopolizes an initially competitive labour market. In such a competitive labour market the allocation of labour is efficient and a union indeed creates a distortion. Alternatively, we could consider introducing a union in a monopsonistic labour market, that is, a labour market in which firms have the power to drive wages below competitive levels. If the union would organize labour supply in such a monopsonistic labour market sector, this could well lead to an allocative efficiency increase if this leads to a move closer to competitive wages.

(ii). Unions may also change efficiency:

(a) direct negative technical efficiency effects of restrictive practices/featherbedding;

(b) union wage increases may force managers to operate more efficiently (to stay in business); related to more general

(c) in a world of imperfect information and incomplete contracts (see introduction and next lecture), union acts as 'agent' on behalve of workers in

1. information provision,
2. contract negotiations,
3. monitoring and
4. contract enforcement.

This may be beneficial to both parties if they are interested in low turnover and stable relationships (*exit-voice hypothesis*).

This begs the question what the benefits of stable employment relationships (rather than spot market transactions) could be and, if there are such benefits, whether we need unions to safeguard those benefits.

(iii). This takes us to the second part of the course:

(a) Why do we see long-term employment relationships rather than spot markets?

(b) How are these relationships organized and why?

(c) How can these relationships be managed efficiently? More in particular, what kind of human resource management practices do we observe and practice and how can these be designed efficiently?

(d) Can unions indeed be helpful in organizing relationships more efficiently?

Clearly, the question whether union may increase efficiency through their agency services is a derived question. We first have to discuss the role of employment relationships and contracts before we can say more about the role of unions in all of this.

Remark 15. It should be said in advance that the empirical evidence is not in favour of positive efficiency effects of unions. See Booth (1995), Ch. 7.

5.2 History

Reading:

- Cully et al (1999)
- Occasional discussions in other texts.

[I have not spend lectures on an institutional overview, but this can easily be studied from the books.]

5.3 Incomplete contracts and labour relations

Reading:

- Milgrom and Roberts, Ch. 10

(i). Long-term relationships (rather than competitive spot labour markets) because of

(a) switching costs and

(b) specific investments.

An example of specific investments are investments in firm-specific human capital. These are skills that are only useful in the firm, and cannot be used at other firms. This could be social contacts within the firm, knowledge how to operate specific machinery, or even such a simple thing as knowledge where to find office supplies. Once these investments are made, the worker and the firm can only cash the future return if they stick together. Neither firm or worker can threaten to employ their specific capital elsewhere. If they move partners, they would have to make a new specific investment to attain a similar level of productivity, all else equal. So, they are to some extent locked into their relationship.

Switching costs can be seen as a special case of relationship-specific investments. A simple example of switching costs are the costs of actually moving your office

and filling in tax forms etcetera. A more involved example are search costs. Search models of the labour market assume that finding a job is a costly and time-consuming activity. Reversely, a firm spends a significant amount of time and money locating suitable employees. Because of these "search frictions", there are switching costs. We can view the resources spent in search by the firm and the worker as specific investments in establishing the firm-worker match. Once matched, the firm and the worker are to some extent locked into their relationship, as they would both have to incur new search costs before they can trade with an other partner.

Remark 16. One difference between search and investment in specific human capital is that in the case of search the future partners are not in contact *ex ante* (that is, before investments are made) and, therefore, cannot write a contract on the division of the costs of and returns on the search investment. In contrast, in the case of specific human capital the worker and the firm can possibly contract how the costs and benefits of training are divided between them.

(*ii*). In general, specific investments create (future) opportunities of trade. For example, if you train the specific skills necessary to operate a firm's specific machines now, you can actually operate the machines later. Thus, the specific investment allows you to produce more in the future than you could have without the training. Because of the specific nature of the skills developed, it also allows you to produce more at the firm to which the skills are specific than at "outside" firms, where only general skills (and new specific investments) count. One usually calls the gap that arises ex post (after the investment is made) between trading labour services with the "inside" firm on the one hand and with "outside" partners on the other hand, taking into account the fact that they would have to incur new specific investment costs, a *quasi-rent* (see Remark 17 below).

(*iii*). To reap the benefits from specific investments, the firm and the worker should somehow ensure that they will actually use the specific skill ex post if this is beneficial. Furthermore, they should somehow ensure that both are rewarded appropriately, so that they are willing to invest ex ante. More formally, efficiency requires:

(*a*) *Ex post efficient trade in labour services*

The firm and worker should stick together and produce whenever they are more productive inside their relationship than in the outside market (that is, as long as there is a positive quasi-rents).

(*b*) *Ex post division of quasi-rents that ensure efficient ex ante incentives*

Both the firm and the worker should be confident that they will get their share of the quasi-rent, if any, in order to entice them to invest in specific capital.

Remark 17. In this context, *rents* (as opposed to quasi-rents) would be the excess returns on entering a employment relationship. For example, a worker would earn a rent if (s)he is hired into a job at a higher wage than can be earned elsewhere in the economy. In a competitive market, such rents should disappear because of arbitrage by workers and firms (all workers would flock to the jobs that pay rents). Quasi-rents arise *ex post*, that is after specific investments are made, and are ordinary returns to these specific investments. Thus, quasi-rents may persist even if ex ante competition for jobs and workers ensures ensures there are no rents. See Milgrom and Roberts (1992), Ch. 8.

(*iv*). Problem: ex post firm and workers are in bilateral monopoly and there is no guarantee that the quasi-rents are distributed ex post to ensure efficient investment ex ante. This is called the *hold-up* problem. Two extreme cases are easy to study:

(*a*) *Spot labour markets*

We have already seen in one of the first lectures that in spot labour markets neither firms and workers are very keen on paying for firm-specific training, even if the potential return is very high. After all, without any long-term commitment, the workers can only fear that ex post the firm will capture the entire quasi-rent by paying only the competitive spot wage, which is determined by their general skills only. The firm then faces the fact that the worker is not committed to efficient ex post trade and will move to another firm whenever an even slightly better opportunity comes along. This is especially true if the firm cannot (or actually, does not) commit to reward the worker for specific training by granting the worker a share in the quasi-rent (through payment of a higher wage than the competitive spot wage).

(*b*) *Complete contracts*

Complete contracts are contracts that specify what should happen (who should do what and pay or receive what) in all situations that may possibly arise ("contingencies"). Such contracts are fairly trivial if there is no uncertainty about the future at the contracting date. In general, however, there is a lot of uncertainty about the future and such contracts are far from trivial.

In the ideal situation enforceable complete contracts can be written, the firm and the worker can (and arguably will) ensure efficient (that is, mutually beneficial) investments by stipulating in such a contract the required investment

levels. The contract would also have to specify who pays for training and a distribution of returns (in any imaginable contingency) such that the contract is agreeable to both parties.

With complete contracts, the specific investment and relationship problem can be solved by markets not unlike the simple spot markets. Rather than competing in terms of spot wages, agents would be competing in terms of complete contracts. This competition would (again, arguably) ensure that actual contracts guarantee the appropriate shares in the cost and returns of investment to firm and workers.

(v). So, if enforceable complete employment contracts would exist, we would not have to worry about relationships (let alone the role unions can play in establishing these). However, contracts are typically incomplete, because of *transaction costs*:

- (a) costs of identification of relevant contingencies (problem of unforeseeable contingencies);
- (b) costs of writing complex contracts with many contingencies (problem of abundance of contingencies; bargaining costs);
- (c) costs of monitoring contract performance (problem of observing all relevant information);
- (d) costs of enforcing the contract (problem of verifying all findings; legal costs, etctera).

(vi). Rather, employment contracts are typically very incomplete *relational contracts*. Such contracts frame the relationship in a world of uncertainty and (possibly) asymmetric information by only explicitly specifying a few contingencies and procedures to handle unforeseen contingencies. Such procedures are

- (a) ex post bargaining;
- (b) granting "authority" to one of the parties to make decisions (residual control) and distribute returns (residual claims) whenever the contract falls short;
- (c) arbitration (by a third party).

Ex post bargaining comes with very low contracting costs, because it is basically the default action (bargaining is what naturally happens if no alternative is specified). Because there is no guarantee that an outcome will prevail ex post that provides the proper ex ante incentives (hold-up problem), there are perverse incentive effects.

Authority with a single agent has the advantage of avoiding bargaining and contracting costs.

In employment contracts,

(a) authority is typically given to the firm; employment is to a large extent "at will" (firm can to great extent decide on employment);

(b) workers can always decide to quit (no slavery).

Remark 18. So far, we have only argued that there is an advantage to granting authority to one party, not that it should be the firm . You may wonder why authority typically rests with the firm. Milgrom and Roberts (1992), Ch. 10, review two approaches [not discussed in class]:

(a) Firm as nexus of contracts; physical capital owners have highest risk of appropriation of their capital.

(b) Firm as bearer of reputation (longer horizon).

(vii). To conclude, in the modern view of labour relations:

(a) Labour is much more complex than acknowledged in the classical model.

(b) Long term relations are important because of specific investments/switching costs.

(c) Relations are governed by incomplete relational contracts granting authority to firms.

(d) This results in a decoupling of productivity and pay and allows firms, within some bounds set by the market (but that are not as sharp as in the spot market case), to deploy a variety of human resource policies to

1. attract and retain good people (selection);
2. help them to develop their potential and utilize their skills (investment);
3. inform, motivate and reward workers (motivation and insurance).

(e) Such human resource policies involve various compensation policies and so called "internal labour markets" (MR 11) that are to some extent insulated from the outside market.

(f) We will investigate what kind of human resource policies exist and how they may attain these goals.

5.4 Compensation policies and internal labour markets

Reading:

- Milgrom and Roberts, Ch. 10-13
- Borjas, Ch. 12 (incentives) and Sec. 7.9 (training)
- Bosworth et al, Ch. 18-22 and Sec. 16.4 (training)
- Polachek and Siebert, Ch. 9 and Sec. 4.4 (training)
- Some text on efficiency wages, e.g. Milgrom and Roberts, part of Ch. 8

5.4.1 Introduction

Specific investments and long-term relationships allow for a wide(r) range of compensation policies and internal-labor-market arrangements to pursue a variety of goals in a world of uncertainty and (possibly) asymmetric information

First, what are compensation policies and internal labour markets?

(i). *Compensation policies* are ways of paying workers, in particular in long-term relationships. Because of the specific investments and the resulting insulation from the outside market, there is some flexibility in the way workers are compensated over the course of the relationship. For example, they may receive pay below their MRP first and above their MRP later. Firms and workers can tailor compensation schemes to provide insurance, proper incentives to invest and exert effort and stimulate (self-)selection of workers.

(ii). More generally, firms may employ *internal labour markets*, which are characterized by

(a) limited ports of entry (most employees start in some specific entry-level jobs) and limited interaction with external labour market in general;

(b) career paths within firms and promotions from within;

(c) and are typically found within large firms (because of increasing returns-to-scale in setting up internal-labour markets).

Internal labour markets provide a set of rules for selection, training and promotion through various levels of jobs.

Because of the specific investments and the resulting insulation from the outside market, there is some flexibility in the way internal labour markets are organized.

Obviously, compensation policies are an important aspect of internal labour markets

5.4.2 Overview of objectives (and instruments) of human resource policies

(*i*). *Insurance (against bad and good days, ability risk, etcetera)*

We observe that wages are more stable than competitive (spot) labour market theory predicts. One explanation is transaction costs: a fixed (or simple linear) wage schedule is cheap to contract and enforce. It could also be a symptom of insurance.

(*a*) If workers are risk-averse then they would like to minimize earnings fluctuations, i.e. they would like to insure against income fluctuations.

(*b*) Such fluctuations could come about because of "idiosyncratic" shocks (bad and good days), but also more aggregate shocks to productivity (demand may temporarily fall for certain products, etcetera).

(*c*) If firms are risk-neutral they may provide insurance by offering pay that is constant or fluctuates less than worker's MRP. Workers are willing to get paid a bit less on average in return for more stable pay. So, insurance both increases profits on average and increases worker utility.

(*d*) Firms will definitely be risk-neutral with respect to idiosyncratic shocks that do not affect the productivity of their total workforce, but may also be risk-neutral with respect to company-wide or even sectoral/national fluctuations if their owners have well-diversified portfolios.

(*e*) The existence of specific investments and long-term relationships make it possible to pay wages that deviate from (are higher than) outside spot wages. Thus, the firm can to some extent pay time-constant wages even though outside wages fluctuate over time.

(*f*) Spot labour markets without specific investments, even though they may provide some short-term insurance, cannot provide such long-term income insurance.

This story seems to work quite well, but there are a few caveats, leading to partial insurance.

(*a*) First, why does the firm provide insurance to its own workers? Why don't they get their insurance from their insurance company that also provides health insurance, homeowners insurance, etcetera?

Possible answer: Individual productivity may not only depend on shocks that are outside the control of the worker, but may also be directly affected by the worker's effort on the job. Assuming that the workers dislike effort, insurance

would allow the worker to shirk on the job without being punished by lower wages! This is the "moral hazard" problem.

It may be easier for the firm than for an outside insurer to monitor worker performance and separate between "exogenous" shocks and choices by workers. Then, the firm may be able to provide insurance that will not be provided in the insurance market because of moral hazard.

(b) This being said, even within a relationship it is typically still impossible to fully monitor and control worker effort. So, even risk-neutral firms will be reluctant to provide full insurance to their workers because of the disincentive effects.

This is a first example on how human resource (in this case, compensation) policies affect multiple human resources goals. They cannot be studied in isolation!

(c) Another problem with providing, in particular, long-term insurance with incomplete contracts is opportunism (draw graph). Workers are typically eager to stay in the relationship if they get paid long-term average wages in bad times, but may be tempted to leave if they get paid the same wage in good times! So, even though specific investments and the corresponding quasi-rents to some extent isolate the relationship from outside spot market forces, these definitely put limits (without binding slavery-type contracts) on the compensation packages the firm can offer. This is a commitment problem in a world with uncertainty and incomplete contracts.

(d) Firms may not be completely risk-neutral, because of uninsurable risks (aggregate risks that affect the entire world so-to-speak, for example large-scale rare natural disasters and wars).

(e) The firm may also be risk-averse because of limited access to financial capital. The latter liquidity constraints may lead to bankruptcy, and loss of valuable (specific) investments, of companies that are profitable in the long term. If so, firms would prefer to reduce the bankruptcy risk by reducing wages in bad times. This is in the interest of both the worker and the firm (if there is a quasi-rent left).

Note that this problem creates an incentive for firm to lie about state of demand; union could be helpful here if more efficient monitoring; can also use strike to check whether firm is willing to hold out, which signals low opportunity cost of no production)

(ii). *Incentives*

The second objective of human resource management was already mentioned, provided proper incentives to the workers to exert efficient levels of effort.

If worker output depends on endogenous costly effort, the firm (and the worker) want to provide incentives to exert effort. Full insurance would be detrimental for efficiency in this case. Somehow, worker compensation should be positively related to their output.

If effort is observable, the firm and the worker can simply contract on effort. This does not necessarily require a contract that explicitly describes the effort required from the worker. For example, if output depends only on effort and there is no uncertainty, then effort is indirectly observable if output is observable. Then, a contract specifying the output would do the trick.

However, typically at the very best the firm observes some measure of individual output, which depends on both worker effort (observed by the worker) and some exogenous random shocks. Thus, there is asymmetric information.

Various ways to provide incentives in this situation:

(a) piece-rates as opposed to time-rates

(b) group-incentive pay and profit sharing

(c) efficiency-wages (and risk of being fired)

(d) deferred compensation (,,)

(e) promotions and tournaments

Rest in various degrees on stability relationship: opportunism again restrict applicability (for example, with deferred compensation worker should trust that firm will indeed pay higher wage later on).

In general, providing incentives requires making pay dependent on imperfect measures of effort (for example, number of units produced in case of piece-rates). Thus, there is in general a trade-off between incentives and insurance.

Only if worker risk-neutral, optimal incentives by having worker (agent in possession of private info) bear all risk (i.e. make the worker reap all benefits from effort!).

We will go through the various incentive mechanisms in detail later.

(iii). *Selection*

Selection through combination of firm screening, performance reviewing and self-selection (possibly induced by compensation schemes etcetera).

Various elements of internal labour markets contribute to proper screening, for example tenure in academic departments.

Interaction with incentive schemes: for good workers it may be more attractive to e.g. participate in tournaments than for bad workers. So, tournaments provide both incentives and means to select good workers for top-level jobs.

Also, good workers may prefer piece-rates over time-rates. If you compare otherwise similar time- and piece-rate workers, you find that the latter earn more and are more productive. What is the explanation?

(*iv*). *Investment in human capital*

Compensation specified in incomplete contract should provide incentives to invest to both worker and firm. For example simple contract specifying wage above outside wage but below maximum wage acceptable to firm may entice both parties to invest ex ante. If ex post shocks are not too large, both parties want to trade ex post against contracted wage, and neither can credibly threaten to not trade (in order to renegotiate wage).

By offering a upward sloping tenure profile of wages, both insurance, incentives to exert effort and incentives to invest in specific skills are provided!

5.4.3 Compensation policies

We now discuss various compensation policies in slightly more detail.

(*i*). *Time rates, piece rates, and more general individual incentive-pay schemes*

[See also Problem Set 4]

(*a*) Time rates: fixed wage or salary per time period;

(*b*) Piece rates: worker is paid fixed amount per unit of production;

(*c*) More general individual incentive-pay schemes: fixed base salary plus an output-dependent performance component.

In US manufacturing 25% of workers were on individual incentive pay in the 1960s and 20% in the 1970s. In the UK, 40% of male manual workers and 30% of female manual workers were on individual incentive schemes in the 1980s. Prevalence among white-collar workers is more than twice as low. For most of these workers on incentive schemes, the incentive component of compensation amounts to less than 10–15% of total pay. So, workers are, if anything, usually paid a mix of (a lot

of) base pay and (some) incentive pay. See Polachek and Siebert (1993) for these numbers.

First, note that these compensation schemes are also relevant outside long-term employment relationships. Even if you are hired for a quick temporary job (e.g. a day's work on the land of a farm), you may be paid in either of the above ways and this may affect your utility and the firm's profits.

Suppose that individual output is observed by both the firm and the worker (and can be verified in court, so that contracts can be contingent on output). Also, assume that the worker is risk-averse and the firm is risk-neutral.

We consider the following cases:

(a) *Output only depends on chance events that are outside the control of the worker.*

We have seen that in this case it is most efficient to pay the worker a wage (time rate) that is fixed independently of actual output. This provides insurance to the worker. The worker is willing to give up some expected wage as an insurance premium, so that the firm makes higher profits (relative to the case in which it employs piece rates).

(b) *Output depends only on worker effort and not on chance events.*

In this case, the worker chooses an effort level and given that choice neither the worker nor the firm faces risk (independently of the compensation scheme used). So, there is no need for insurance and the contract can be tailored to provide the appropriate incentives to work. Both time-rates with a contractual requirement to exert the right amount of effort and an incentive-pay contract that induces the right amount of effort can be used equivalently (because there is no uncertainty).

Example 3. (Along the lines of Problem Set 4) Suppose the worker can choose of only two effort levels, "work" and "shirk". If the worker works output is £100 and if the worker shirks output is £50. The worker incurs a cost of £20 if he works (and £0 if he shirks).

In this case, it is efficient if the worker works, rather than shirks: an output gain of £50 is achieved at an effort costs of only £20. So, the firm and the worker prefer that the worker works. Because there are no information problems the worker can simply commit contractually to working and the firm and the worker only have to decide on how to divide the gains.

Suppose the worker can earn £60 net of effort costs elsewhere in the economy. Then the worker will only accept a contract that offers at least £60 net of

effort costs. This is called the worker's *participation constraint*. As output of a working worker with the current firm is £100 > £60 + £20, the worker and the firm would like to come to an agreement (this would be efficient, i.e. to their mutual benefit, as they can both be better off if they do).

If the firm can unilaterally propose a contract, it will propose a contract that induces the right amount of effort and that just entices the worker to agree. The firm will not offer more, because it will only loose profits and does not change the participation or effort decision of the worker. This is the kind of situation discussed in Problem Set 4.

In this case, the contract could either stipulate that the worker should work and is paid a fixed wage £80, or it could only specify that the worker is paid 80% of his output. The latter arrangement would pay £40 is the worker shirks and £80 if he works, so the worker would decide to work. In either case, the worker is equally happy to get £80 for sure. Also, this £80 minus £20 effort costs just matches the net outside wage £60, so that the worker's participation constraint is satisfied.

The main ideas of the example are:

1. Without uncertainty and observability of output, effort is effectively observed and there is no incentive problem.
2. In this particular case it is efficient to work rather than shirk. That is, both the firm and the worker could be better of if the worker works. They can achieve this by either explicitly stating in the contract that the worker should work, or implicitly induce work effort through an appropriate pay scheme (the one giving is a linear piece-rate scheme, but other incentive schemes will do).

(*c*) *Output depends on both chance events that are outside the control of the worker and on worker effort, but worker effort is observed (and can be contracted upon, i.e. verified in court) by both the firm and the worker (symmetric information).*

In this case, the worker and the firm can still contract on effort because of the (somewhat unrealistic) assumption that the firm can distinguish the effort and chance components of output. However, there is now also a role for insurance because there is risk. Thus, they can contract on an efficient level of output and provide full insurance to work by specifying a fixed wage (time rate).

Example 4. Consider the previous example, but now extended with risk as in Problem Set 4. The probabilities of high and low output under shirking and

working are summarized in

	shirk	work
output	probability	
£100	1/5	4/5
£50	4/5	1/5

Unlike in the previous example, due to chance output can be low if the worker works and high if the worker shirks. Working however increases the probability of high output.

Because there are no asymmetric information problems, and therefore no incentive problems, the firm can provide full insurance to the worker by specifying a fixed wage. If it can unilaterally propose a contract, it will propose a fixed wage that just entices the worker to accept, which is £60 if the worker is told to shirk and £80 if the worker has to work. Expected profits are then $(1/5) \cdot £100 + (4/5) \cdot £50 - £60 = £0$ if the worker is asked to shirk and $(4/5) \cdot £100 + (1/5) \cdot £50 - £80 = £10$ if the contract specifies that the worker works.

Thus, the optimal contract will specify that the worker works and compensate the worker with a fixed wage £80. This solution is (Pareto) efficient, as it provides both full insurance to the worker and induces the output-maximizing level of output.

(*d*) *Output depends on both chance events that are outside the control of the worker and on worker effort, but worker effort is not observed by the firm (asymmetric information).*

In this case, the firm cannot tell whether output is low because the worker shirks or because of bad luck. The contract cannot specify effort levels and providing both full insurance and maximizing output is not feasible. Instead, there is a trade-off between insurance and incentives. If the worker is fully insured, he can shirk without consequences and will do so to avoid effort costs (moral hazard). An incentive (e.g. piece-rate) contract that induces the worker to work by making pay dependent on effort also makes pay dependent on chance and therefore offers partial insurance at most.

An optimal contract will typically offer partial insurance and incentives (dependence of pay on output) that are just sufficient to induce the worker not to shirk, unless full insurance and shirking is preferred by both parties.

Example 5. In the example above, the firm could offer full insurance through

a contract that specifies a fixed wage £60. The worker will be just willing to agree, will shirk and firm's expected profits will be £0.

At another extreme, the firm could provide a lot of incentives by paying all output as wages, so that it pays £100 if output is high and £50 if output is low. The firm's expected profits are again £0, so the firm is indifferent between this contract and the full insurance contract.

The worker would have expected pay equal to £60 if shirking. So, the worker would prefer earning £60 in the outside market over shirking with this firm. If working with the firm, the worker would earn £90, and therefore £70 net of search costs, in expectation. Provided that the worker is not too risk-averse (for example, if utility derived from consumption x is $\sqrt{x}$ as in Problem Set 4), it would prefer working with the firm over working in the outside market and shirking (see also Problem Set 4). Tables 1 and 2 summarize for the utility function of Problem Set 4.

Table 1: Fixed wage equal to £60

worker shirks

	output	prob.	wage	effort cost	utility	profits
high output	£100	1/5	£60	£0	$\sqrt{60} \approx 7.75$	£40
low output	£50	4/5	£60	£0	$\sqrt{60} \approx 7.75$	−£10
expected	£60		£60	£0	$\sqrt{60} \approx 7.75$	£0

worker works

	output	prob.	wage	effort cost	utility	profits
high output	£100	4/5	£60	£20	$\sqrt{40} \approx 6.32$	£40
low output	£50	1/5	£60	£20	$\sqrt{40} \approx 6.32$	−£10
expected	£90		£60	£20	$\sqrt{40} \approx 6.32$	£30

Thus, in that case, this contract is "incentive compatible" in that it induces the worker to exert the warranted level of output. If the worker strictly prefers to work and stay with the firm, the contract is not optimal for the firm nor efficient from their joint perspective. First, insurance can be increased without hurting incentives by decreasing the variance of pay without changing expected pay for a working worker (efficiency). Second, expected pay can be decreased leaving some positive profit to the firm (optimality).

Problem Set 4 provides a more worked-out example. We will not formally derive efficient/optimal contracts. A C-level course at UCL exists that rigorously and

Table 2: Wage equal to actual output

worker shirks

	output	prob.	wage	effort cost	utility	profits
high output	£100	1/5	£100	£0	$\sqrt{100} \approx 10.0$	£0
low output	£50	4/5	£50	£0	$\sqrt{50} \approx 7.07$	£0
expected	£60		£60	£0	≈ 7.66	£0

worker works

	output	prob.	wage	effort cost	utility	profits
high output	£100	4/5	£100	£20	$\sqrt{80} \approx 8.94$	£0
low output	£50	1/5	£50	£20	$\sqrt{30} \approx 5.48$	£0
expected	£90		£90	£20	≈ 8.25	£0

much more extensively discusses this and other models of information.

In the most realistic case (d) of asymmetric information, there is a tradeoff between insurance and incentives. In general, under asymmetric information a solution as efficient as the symmetric information solution (with full insurance) cannot be obtained. Some risk has to be borne by the risk-averse worker to elicit proper work effort.

In Problem Set 4, an example is also given of the possible selection effects induced by the choice of contract in the case that workers are heterogeneous and that firm's cannot distinguish between workers of different types. Loosely, lazy workers are less attracted to piece-rate contracts and more to time-rate contracts. Normal, or very able, workers prefer incentive-pay. Thus, by offering incentive-pay schemes like piece-rates, the firm can try to attract (only) normal, able workers.

In the problem set an example is given of a so called "menu of contracts". The firm can offer the worker a choice of a variety of contracts, for example a time-rate and a piece-rate contract. If well-designed, different workers (lazy and normal) will choose different contracts and thus reveal their type! Thus, menus of contracts are useful tools for selection.

Empirically, this causes interpretational problems. It can be shown that workers on piece-rate contracts are on average 15%-35% more productive than otherwise similar workers on time-rate contracts. You would like to ascribe this effect to the incentives provided by piece-rate contracts. However, we have just seen that more able (and less lazy) people are more likely to select jobs that offer piece-rate pay.

Thus, the measured 15%-35% productivity gap confounds incentive and selection effects.

One problem with menus of contracts in a dynamic setting is that the firm designs its menu without knowing the type of working it is dealing with. Once the worker has revealed his type, the contract offered may not be the best option for the firm anymore. If the firm is not committed to sticking to the contract originally offered, it may propose another contract to the worker (who is now of a known type) that yields higher profits. Foreseeing this lack of commitment, the worker may not choose the contract he really wants in the first place. This is called the "ratchet" effect.

Individual incentive contracts (and piece-rate schemes) are simple and effective ways of providing incentives to workers:

(a) simple and easily understood;

(b) provide strong incentives to exert effort and increase productivity;

(c) self-selection of able and hard-working workers;

(d) stimulates human-capital investment (consideration in long-term relationship);

(e) no costly performance reviews;

(f) based on objective measures of performance (avoids favoritism, manipulation).

Example 6. *Effects on human-capital investment*

Two-period relationship; forget about insurance, incentives and discounting first.

(a) First-period productivity/outside wage: $w_0 = £10$.

(b) Second-period expected outside wage is also $w_0 = £10$. First-period investments of $£1$ by both firm and worker turn the expected productivity in period 2 into $w_0 + £5 = £15$.

(c) Fixed second-period wage w_0: worker does not invest
Second-period wage $w_0 + 0.5(\text{output} - w_0)$: worker's expected pay $£12.5$; firm's expected profits $£2.5$ compensate for investment (no credible threat to withdraw in second period; invest in first period)

With risk-averse workers this loads risk on the worker. Alternative is to write fixed wage contract that specifies a fixed $£12.5$ wage in the second period (but at will, etcetera). If they invest, same situation arises (limits to effects opportunism because of quasi-rent). If worker does not invest, firm can credibly theaten to fire worker and contract will be renegotiated down to outside wage. Thus, worker and firm will have incentive to invest.

However, disadvantages of piece-rate systems/individual incentives schemes:

(a) Constrained by risk-aversion of workers (flip-side of providing incentives);

(b) Individual output may be hard to define unambiguously (not so objective after all);

(c) Workers may have little control over their own output (for example if part of an assembly line);

(d) Workers may shift attention from quality to quantity if only quantity is rewarded (more general point about providing incentives with multi-dimensional effort). This introduces the need for (costly) quality inspections. Holding workers responsible for quality loads risk on worker. Also, workers may neglect maintainance of machines, helping their co-workers etcetera. See also the Lincoln Electric example in MR.

To some extent, these adverse effects can be mitigated in long-term relationships, for example by stable job assignments (need to maintain your own machine; co-dependence on co-workers).

You expect that piece-rates are more likely to be observed in jobs where

(a) individual output is easy to define,

(b) effort is hard to measure directly,

(c) workers have a lot of control over their output and

(d) quality is easy to monitor .

For example, sales commissions are less prevalent in large department stores than for door-to-door sales persons. Obviously, effort is easier to monitor for the first class of workers. Also, Borjas: footwear/shirts versus industrial chemicals.

One problem in implementing individual incentive schemes is that it may be hard to measure the relevant parameters that feed into the firm's decision problem (effect of individual effort on measure of output, effort costs, etcetera). This leads to various instances of the ratchet effect. Workers have incentive to pretend their task is hard when standards are set. They may have similar incentives if standards (piece rates) are often reviewed in practice (example: Soviet central-planning system). Commitment problem. Similarly, firm may try to lower pay by claiming conditions are bad.

(ii). *Group incentive-pay and profit sharing*

(*a*) *Profit-sharing* used by 30% of US companies in 1988, but perhaps only 1% of employee income from profit-sharing (MR).

Extreme form is employee-ownership of firm (e.g. partnerships in law, consultancy, etcetera).

(*b*) *Gain-sharing*: group-bonuses for reaching certain (output, quality, etcetera) targets.

Why group incentives rather than individual incentives? (Closely linked to disadvantages piece-rate scheme)

(*a*) Individual contributions output hard to measure (design team)

(*b*) Groups of workers may have better information than management. Group-incentives then stimulate monitoring within group.

(*c*) Maybe easier to implement changes if they benefit group as whole.

(*d*) Group-incentives stimulate co-operation, which may be productivity-enhancing.

(*e*) Group may be better able to change production methods etcetera in response incentives.

(*f*) Better risk-sharing (within groups).

What group size? See MR.

Fairness issues. See MR.

(*iii*). *Efficiency wages*

Shapiro-Stiglitz model (e.g. MR 8 p. 250–; Borjas Sec. 12.5).

Idea: pay $w > w_0$, monitor worker and fire worker if found shirking. Requires quasi-rents (to allow $w > w_0$), for example due to search/unemployment. Interacts with internal labour markets: Japanese system with low-level hiring only.

Some evidence:

(*a*) (Borjas) 10% higher wages reduces dismissals for displinary reasons by 5%.

(*b*) (Borjas) Huge wage differentials, in particular between industries. Could be compensating differentials for job characteristics or worker traits. Mixed evidence.

(*c*) (Polachek and Siebert) 8% higher wages in UK plants with > 1000 employees relative to plants with ≤ 100 employees.

Advantages of efficiency-wages over direct incentive pay:

(*a*) Incentive-pay schemes may not be feasible, for example if worker cannot survive a period of low pay (liquidity constraints on worker).

(*b*) Efficiency wage schemes can use subjective evaluations in a way that direct incentive pay schemes cannot. With direct performance pay, the firm has an incentive to underrate the worker's performance in a subjective evaluation so that she has to be paid less. With efficiency wages, the only option is to fire the worker, or keep the worker at a fixed wage (independent of the review). The employer has no incentive to underrate the worker, as he would be firing a perfectly fine worker that brings in profits.

For simplicity, suppose one period model. Worker only gets paid outside wage w_0 if found shirking.

(*a*) Outside wage w_0.

(*b*) Inside wage $w > w_0$.

(*c*) Probability of monitoring (and detection of shirking) p.

(*d*) Cost c of exerting effort (working).

(*e*) Worker works if $u(w-c) \geq (1-p)u(w) + pu(w_0)$.
The wage w_e such that $u(w_e - c) = (1-p)u(w_e) + pu(w_0)$ is the "efficiency wage" that just induces the worker to work.

(*f*) Note that $w_e - c \geq w_0$, so that the worker participation constraint is always satisfied. We have to check though that the firm finds it worthwhile to pay efficiency wages (monitoring cost?).

For example, if the worker is risk-neutral (linear utility), then

(*a*) worker works if $w \geq w_0 + c/p$ and $w_e = w_0 + c/p \geq w_0 + c$ is the efficiency wage.

(*b*) w_e increases if c increases or p decreases. Thus, there is a trade-off between monitoring cost (that is, establishing higher p is more costly) and the level of efficiency wages. More extensive model would explicitly model choice of p for given monitoring technology (cost). See MR.

In earlier numerical example, $w_0 = £60$ and expected profits would be zero if $w = w_0$. Suppose that $p = 4/5$. The efficiency wage $w_e = £60 + (5/4)£20 = £85$ would induce the worker to work and deliver expected profits $£5$ (but, monitoring cost?).

If worker risk-averse, the efficiency wage can be lower, as there is only risk if the worker shirks. For example, with the preferences of the numerical example, a wage of £85 would give utility $\sqrt{85-20} \approx 8.06$ if working and $(1/5)\sqrt{85}+(4/5)\sqrt{60} \approx 8.04$ if shirking.

This model is extremely stylized in not explicitly dealing with dynamics. Efficiency wages are more effective in long-term relationship in which firing because of shirking interrupts a stream of high wages.

Note that efficiency wages typically limits turnover, and may have insurance, selection and HC effects.

Efficiency wage theory is typically discussed as a theory of *involuntary unemployment*. The quasi-rent arises because separated workers have to live through a spell of unemployment. The unemployed at any point in time would strictly prefer to work at the efficiency wage.

Criticism of efficiency wage theory: unless quasi-rent due to specific investment (search, HC, switching cost), unemployed have incentive to underbid incumbent workers by offering entrance fee/posting bond (up to the point they are indifferent). This may seem like a bribery story, but is really a deferred compensation story.

(*iv*). *Deferred compensation*

As noted in efficiency-wage context, in long-term relationships shirking worker typically runs risks of getting fired and loosing *future* wages.

Increasing wages over tenure (deferred compensation) can further increase such incentives (and may reduce turnover, reward HC investment, lead to selection and provide insurance).

It is sometimes said that the worker "posts a bond" that gets repaid over the course of the relationship. This is related to the "golden handcuffs" discussion in MR.

Note that deferred compensation is, like all truly long-term arrangements, particularly vulnerable to opportunism/firm moral hazard. Firm has incentive to fire worker (or induce the worker to quit) before the bond has to be repaid. With sufficiently large quasi-rents, this may not be a huge problem. Also, firm may be long-lived and worry about its reputation with future employees.

5.4.4 Internal labour markets, promotions, tournaments

(*i*). Within large firms, we typically find *internal labour markets*, which are characterized by

(*a*) long-term relationships

(*b*) limited ports of entry and limited interaction with external labour market in general;

(*c*) career paths within firms and promotions from within;

(*ii*). Often, wages are attached to jobs/job classifications rather than individuals.

Incentives can then be provided through performance-based promotions through a sequence of jobs.

This requires commitment to an internal labour market by the firm.

(*iii*). This may interact/conflict with other objectives of internal labour markets, like the selection of the right worker for the right job and human capital accumulation.

It may be easier to collect information on workers within the firm than on outsiders, so that careers within internal labour markets may be the efficient way of selecting managers.

Higher level positions may require specific skills that can only be acquired on-the-job within the firm.

(*iv*). Promotions can be based on (often subjective) absolute performance evaluation, but also on relative performance evaluations, *tournaments*.

Note that pay should rise even if task in new job not more complex (because of selection of better workers).

For simple formal model of tournaments (Lazear and Rosen), see PS, Sec. 9.4.

(*v*). *Some disadvantages of promotions/tournaments as an incentive device*

(*a*) Very blunt relative to performance pay, etcetera.

(*b*) Cannot be used at top: if promotion is only incentive scheme, those at the top drink coffee all day. Other incentives have to be given to managers.

(*c*) Competition between workers may be counterproductive (sabotage).

(*d*) Misassignment of workers to jobs may result. System may be such that everyone is "promoted to his level of incompetence". Also, best salesman may not be best manager.

(*vi*). *Some advantages of tournaments as an incentive device*

(*a*) evaluation costs (only measure of *relative* performance required at a few points in time);

(b) gets rid of common error in performance;

(c) with fixed prize, no room for managers/owners to cheat.

(vii). MR 12 describe some human-resource systems, which you can and should study:

(a) Hierarchy of jobs plus entry position, with, for each job level, a pay level, minimum standards and (higher) promotion standards.

Incentives and selection come together as discussed before.

(b) If different skills needed on different levels, mix of incentive pay and promotions can be answer.

(c) Up-or-out and tenure.

We have discussed tenure before (e.g. providing proper incentives in selection of new workers).

Combination with up-or-out useful to take away possibility to delay tenure in order to capture more rents.

[This last part on features of internal labour markets was rather sketchy. I do expect you to study these issues in the reading, notably Milgrom and Roberts (1992), Ch. 11.]

5.5 The role of trade unions

Reading:

- Booth (1995), Sec. 3.2.3 and 3.3; Ch. 7.
- The other references on unions (notably, Freeman and Medoff, 1984) provide additional material.

[The organisation theories discussed in e.g. Booth link unions to the human resources theory discussed so far.]

5.6 Empirical analysis of human-resource systems and unions

Reading:

- Booth (1995), Ch. 7.
- Bits and pieces scattered throughout the other texts.

[We did not have time for a comprehensive discussed of the empirical analyses of human resources systems, but we did encounter bits and pieces as we went. I do expect you to pick up these bits and pieces from the reading, classes and lectures, but that will suffice.]

References

Binmore, K.G., A. Rubinstein, and A. Wolinsky (1986), "The Nash bargaining solution in economic modelling", *Rand Journal of Economics*, **17**, 176–188.

Booth, A. (1995), *The Economics of the Trade Union*, Cambridge University Press, Cambridge, UK.

Borjas, G.J. (2000), *Labor Economics*, second edition, McGraw-Hill, Boston.

Bosworth, D., P. Dawkins, and T. Stromback (1996), *The Economics of the Labour Market*, Prentice Hall, New York.

Cully, M., S. Woodland, A. O'Reilly, and G. Dix (1999), *Britain at Work: As Depicted by the 1998 Workplace Employee Relations Survey*, Routledge, London, UK.

Layard, R., S. Nickell, and R. Jackman (1991), *Unemployment; Macroeconomic Performance and the Labour Market*, Oxford University Press, Oxford, UK.

McDonald, I.M. and R.M. Solow (1981), "Wage bargaining and employment", *American Economic Review*, **71**, 896–908.

Milgrom, P. and J. Roberts (1992), *Economics, Organization and Management*, Prentice-Hall International, London, UK.

Nash, J.F. (1950), "The bargaining problem", *Econometrica*, **18**, 155–162.

Nash, J.F. (1953), "Two-person cooperative games", *Econometrica*, **21**, 128–140.

Polachek, S. and S. Siebert (1993), *The Economics of Earnings*, Cambridge University Press, Cambridge, UK.