



# CONTRACTS AND INCENTIVES IN LABOUR MARKETS

Six Essays

by

Kjell Erik Lommerud

Institute of Economics  
The Norwegian School of Economics  
and Business Administration  
Helleveien 30,  
5035 BERGEN-SANDVIKEN  
NORWAY

1987

## FOREWORD

These six essays constitute my thesis for the degree dr. oecon. They are all essays in labour economics, the first four lie within the field of labour contract theory. No particular reading order is required, apart from that the first essay, a survey article, might help putting especially the three next essays in perspective. Hopefully, this first essay also sheds some light on what I mean by 'contract theory', a central concept throughout this thesis. For further information on the contents of these essays, I refer to the abstracts preceding essays 2, 3, 4 and 6 and succeeding essay 5.

During the time I have worked on this thesis, I have accumulated debt to many persons. Foremostly, I would like to mention my supervisor Kåre Petter Hagen and the two remaining members of my supervising committee, Agnar Sandmo and Jon Strand.

The thesis was written mainly during my three years as a research fellow at the Norwegian School of Economics and Business Administration (NHH). I had the pleasure of spending one of these years at the London School of Economics (LSE), which I found enriching in many respects.

Many persons at these institutions have helped me at one point or another. I would like especially to mention three fellow doctoral students, and thank them for discussions and friendship: Jan Erik Askildsen, Bjørn Sandvik and Padmini Venkatnarayan.

Typing assistance has been offered by several secretaries: Peter Hansteen did the bulk of the work. Thank you!

Moreover, during my year in London I received financial support not only from NHH, but also from the Norwegian Research Council for Scientific Research (NAVF) and the Bank of Norway.

I doubt whether it is possible to complete a doctoral thesis without the support of a whole lot of people. I know it sounds like an empty cliché, but I feel real gratitude towards many of those people I have not found space to mention in these acknowledgements.

**CONTRACTS AND INCENTIVES IN LABOUR MARKETS**  
**Six Essays**

by

**Kjell Erik Lommerud**

**CONTENTS:**

Foreword .....	i
Contents .....	ii
CHAPTER 1: Explicit and Implicit Labour Contracts: A Survey .....	1
CHAPTER 2: Labour Contract Theory and the Insider- Outsider Dilemma .....	38
CHAPTER 3: Worksharing versus Layoffs in Labor Contracts .....	63
CHAPTER 4: Marital Division of Labor with Risk of Divorce: The Role of "Voice" Enforcement of Contracts .....	85
CHAPTER 5: Persistent Discrimination with Social Ability as a Productive Factor .....	127
CHAPTER 6: Educational Subsidies when Relative Income Matters .....	139

**Chapter 1:**

**EXPLICIT AND IMPLICIT LABOUR CONTRACTS: AN INTRODUCTORY SURVEY**

June 1987

Revised October 1987

## EXPLICIT AND IMPLICIT LABOUR CONTRACTS: AN INTRODUCTORY SURVEY

### 1. Introduction.

A natural first question when embarking on a survey like this, is why labour contract theory is an interesting topic. Of course, as a preliminary towards answering this question it is necessary to get some idea of what labour contract theory is. Unfortunately, this concept is not very precisely defined. A broad definition is the following: Labour contract theory is the study of the terms of exchange in the labour market. The very simplest form of a labour "contract" would then be the wage rate paid when labour and money are exchanged simultaneously in a spot market. However, economists have become increasingly aware that such spot market models are rather poor at capturing what really goes on in the labour market. One important issue is that for most people the value of their human capital is the lion's share of their total wealth. At the same time it is difficult to insure oneself against the uncertainty connected with the value of human capital. Accordingly, we would expect labour contracts to include implicit insurance arrangements. Moreover, the employer-employee relationship is often long-term, and both parties tend to invest in relationship-specific capital. Long-term contracts are called upon to lessen the vulnerability stemming from such relationship-specific investments. On the other hand there are large problems when it comes to foreseeing all eventualities, specifying the contract elements, verifying what has actually happened and enforcing the contract. Labour markets abound with informational problems: When a worker is hired, there is a lot the employee does not know about his future job, and much the employer does not know about his prospective worker. Also at later stages of the employment relationship will there be asymmetric information - as regards the true profits of the firm, the workers' effort levels, the workers' alternative opportunities. Also in many other ways real labour markets will differ from simple spot market exchange.

Of course, phenomena like these prevail in many other markets too. Labour contract theory is but an application of more general insights. Still, I believe the labour market offers the most important examples why the study of more complicated contracting processes is important.

If one accepts the above, broad definition of labour contract theory, it is rather easy to argue that this theory is interesting and important. One only has to argue that it is important to have as realistic as possible a picture of the labour market, the market which is perhaps the most important for people's daily lives.

The drawback with such a broad definition is that it makes contract theory almost synonymous with labour economics. In what follows I will restrict myself to trying to convey the main contents of three strands of literature which conventionally have been labelled "contract theory". First, the original "implicit labour contract theory" of the mid-seventies concentrated on the role of the firm as an insurer against uncertainty about the value of the workers' human capital. This theory is presented in section 2. A later stream of articles focused on the implications of asymmetric information between firm and workers concerning the true state of nature. In section 3 I present this theory, and argue that perhaps too much attention has been paid to this line of research. Finally, in section 4 we will have a look at what can be termed "transaction cost" contract theory. This body of theory concentrates on the costs of writing, verifying and enforcing contracts. One reaction to such transaction costs would be to rely on unwritten, tacit agreements, enforced by the parties' concern for their reputation. As I see it, this is implicit contract theory in the true sense. We will also have a look at the use of noncontingent contracts in combination with damage measures as a response to such a situation.

## 2. The Firm as an Insurer.

Azariadis (1975), Baily (1974) and Gordon (1974) count as the seminal papers in labour contract theory. Surveys of this early literature can be found in Sargent (1979) and Azariadis (1981).

The basic assumption in this literature is that workers cannot insure themselves against fluctuations in their labour income, in some unspecified manner due to moral hazard problems. However, the firms which employ the workers can overcome these informational problems: They know whether an employee's low income can be explained by an unfortunate state of nature or by low effort. This implies that the employment relationship does not only include a sale and purchase of units of labour, but an implicit insurance arrangement as well. In the case where the firm is risk neutral while its workers are risk averse, this will result in a constant wage across states of the world.

This is easily shown in a simple model. The model is a two period one. In period 0 the contract is agreed upon, in period 1 production takes place. In period 1, one of  $S$  states is realized. The probability of a state  $s$  is  $\phi_s$ . The firm is assumed to maximize expected profit, expressed as:

$$\Pi = \sum_s \phi_s \{ \theta_s f(L_s) - w_s L_s \} \quad (1)$$

$w_s$  denotes wages and  $L_s$  employment in state  $s$ .  $f$  is a production function, and we assume  $f' > 0$ ,  $f'' < 0$ ,  $\lim_{L \rightarrow 0} f'(L) = \infty$  and  $\lim_{L \rightarrow \infty} f'(L) = 0$ .  $\theta_s$  can be thought of as multiplicatively separable technological uncertainty, or alternatively, as an uncertain price of the output in state  $s$ .

To attract workers in period 0, the firm must offer a representative worker a level of expected utility at least equal to  $\bar{U}$ , presumably determined in the labour market. We write this as:

$$U = \sum_s \phi_s \{ r_s u(w_s) + (1-r_s) u(R) \} > \bar{U} \quad (2)$$

Here  $U$  denotes expected utility,  $u$  is ex post utility. We assume marginal utility to be positive, but diminishing, implying risk aversion.  $r_s$  is the percentage of workers under contract retained in a given state.  $(1-r_s)$  is thus the percentage of workers being laid off in this state. We assume that layoffs take place by random draw, implying that  $r_s$  and  $(1-r_s)$  are the individual probabilities of work and layoff.  $R$  is the "reservation wage", i.e., the monetary value of the benefits and costs accruing to a laid-off worker. This could include wage in new employment, value of leisure, public unemployment benefits, the psychological strains of unemployment, etc. It is restrictive to assume that leisure etc. has a monetary value independent of the wage level. With a more general specification our results would not survive exactly in the form presented here, but much of their flavour would remain.

Note further that condition (2) tacitly assumes that a worker either works a "full day" or is laid off. This is a restrictive assumption, to be commented upon below.

Also, note that we have presupposed that the firm deals individually with every worker. However, most of our results will rather easily carry over to a unionized setting - as long as we assume that the firm and the union strike an efficient bargain. In other models of union behaviour things would be more complicated. See Oswald (1985) for a survey of recent trade union theory.

The firm wants to maximize (1) given (2). Note that we can write  $L_s = r_s \bar{L}$ , where  $\bar{L}$  is employment in the best state.  $\lambda$  is the Lagrangian multiplier associated with the constraint. The first-order conditions for an interior solution are:

$$- r_s \phi_s \bar{L} + \lambda \phi_s r_s u'(w_s) = 0 \quad (3)$$

$$\phi_s (\theta_s f'(L_s) - w_s) \bar{L} + \lambda \phi_s (u(w_s) - u(R)) = 0 \quad (4)$$



Condition (3) can easily be restated as:

$$u'(w_s) = \frac{\bar{L}}{\lambda} \quad (3')$$

This condition implies that marginal utilities for employed workers are constant across states of nature. With  $u$  being concave, we then know:

$$w_s = w \quad \forall s \quad (5)$$

The wage is constant no matter whether the firm experiences good or bad times.

This result is perhaps not very surprising. We know from insurance theory (Arrow (1971), Borch (1962)) that efficient risk sharing between two parties implies that their marginal rates of substitution between consumption in any pair of states should be equal. If one party is risk neutral, this party will carry all the risk. Of course, this points to a rather straightforward generalization of our model. Suppose firms were risk averse. Their ex post utility function over profit is  $v$ , ex post profit in a state  $s$  is  $\pi_s$ . Efficient risk sharing would then require:

$$\frac{u'(w_s)}{u'(w_t)} = \frac{v'(\pi_s)}{v'(\pi_t)} \quad \forall s, t; \quad s \neq t \quad (6)$$

For a risk neutral employer the right hand side of (6) is 1, in turn implying  $w_s = w_t$ . Azariadis (1981) points out that "strict invariance of wages to the state of nature is not an essential element in the theory of labor contracts. The essence of these contracts is rather that wages differ from the marginal product of labor by an insurance indemnity in adverse states of nature and by a premium in favourable states."

But note that even if the workers are relieved of all uncertainty concerning the wage level in our model, they still face risk concerning the value of their human capital: They face the risk

of layoff. Let us now turn to the study of the firm's employment of labour. Using (3'), condition (4) can be restated as follows:

$$\theta_s f'(L_s) = w - \frac{1}{u'(w)} (u(w) - u(R)) \quad \forall s \quad (4')$$

This condition says that at optimum the value marginal product of labour should equal, not the wage, but something which can be thought of as the shadow cost of labour. This shadow cost consists of the wage minus a term Azariadis (1975) names the "marginal risk premium". This is the monetary value to a worker of a marginal reduction in the probability of layoff in a given state.

By concavity of  $u$ , we know:

$$u'(w) < \frac{u(w) - u(R)}{w - R} \quad (7)$$

Taken together with (4'), this implies (when  $w > R$ ):

$$\theta_s f'(L_s) < R \quad \forall s \quad (8)$$

This condition characterizes the optimal employment level as long as there is unemployment. It implies that in our contract model employment is higher than in a Walrasian spot market model of the labour market (where  $\theta_s f'(L_s) = R$ ). This result is rather intuitive. Our set-up bars the possibility that the firm can insure its workers against the risk of layoff by paying them redundancy pay. It is therefore optimal for the firm to employ more workers than would have been productively efficient, as a partial insurance against layoffs.

But even if employment is higher than with Walrasian spot markets, the unemployment in our model is ex post involuntary: The retained workers receive a wage exceeding their value marginal product and opportunity wage. Furthermore, at this wage, the laid-off workers would have preferred to work. But as the unem-

ployment in this model is the result of voluntarily agreed contracts, it is voluntary in an ex ante sense.

Note that we so far, rather arbitrarily, have presupposed that worksharing is not possible. Any reduction in the employment level therefore takes the form of layoffs. Intuitively one would think it preferable to share unemployment between all workers instead of placing the whole burden on a selected few. In contract theory, as in labour economics in general, it is necessary to assume some sort of nonconvexities to make layoffs the rational choice over worksharing. I.e., it is for some reason more profitable to lay off one worker instead of marginally reducing the workday for all workers. (See e.g. Mortensen (1978), Rosen (1985), FitzRoy and Hart (1985), Lommerud (1986) and Burdett and Wright (1986).) These reasons might be technological, or that the opportunity value of time is larger with layoffs than with worksharing (e.g. because totally laid-off workers more easily can find alternative employment), or that the tax system treats layoffs and worksharing differently.

To sum up so far: If somebody hoped that early labour contract theory should provide a microeconomic underpinning for Keynesian wage rigidity, they must have been disappointed. True, these models predict rigid wages, but the firms do not singlemindedly determine employment by equating the workers' value marginal product with this rigid wage. In fact, in the model investigated so far, employment is too high compared to a Walrasian situation, and unemployment is voluntary in the sense that it originates from voluntary contracts.

#### **Private Unemployment Insurance.**

As we have seen, early contract theory included an ad hoc ban on private unemployment insurance. This led to ex post involuntary unemployment, but still an employment level higher than with spot markets. How does these results change if we allow the firm to pay laid-off workers a redundancy payment,  $b_s$ ? We must then

deduct from profits  $b_s(\bar{L}-L_s)$ , and the utility of a laid-off worker will be  $u(b_s+R)$ . With these alterations, we will have the following first-order conditions for an interior optimum:

$$u'(w_s) = \frac{\bar{L}}{\lambda} \quad \forall s \quad (9)$$

$$u'(b_s+R) = \frac{\bar{L}}{\lambda} \quad \forall s \quad (10)$$

$$\theta_s f'(L_s) = w_s - b_s - \frac{\lambda}{\bar{L}} [u(w_s) - u(b_s+R)] \quad \forall s \quad (11)$$

Conditions (9) and (10) imply that wages are constant across states and that laid-off workers will be given the same utility as those retained, i.e.:

$$w_s = w = b_s + R \quad \forall s \quad (12)$$

Using (12), (11) can be rewritten as:

$$\theta_s f'(L_s) = R \quad \forall s \quad (13)$$

In other words, the employment level is Walrasian. The intuition is straightforward: Private unemployment insurance is used to make the workers indifferent between being laid off or retained. Therefore, the employment decision is left with the firm, which sets employment at the efficient level. Remember that the reason why employment was higher than the Walrasian level in the previous model was that as redundancy pay was not allowed, the firm wanted to give the workers at least some insurance against lay-offs by keeping employment high. When redundancy payment is included in the firm's policy kit, this motive obviously vanishes.

An argument against the present model is that we observe precisely little private unemployment insurance in real life. But why is this? Several authors have focused on informational prob-

lems: It is difficult to monitor the extent of laid-off workers' search for new jobs, or to get workers truthfully to reveal their true alternative opportunities. (See e.g. Geanakoplos and Ito (1985), Moore (1985), Kahn (1985).) Shapiro and Stiglitz (1984) point out the disciplining effect of unemployment: To be fired for shirking is a harsher penalty with high unemployment. If fired shirkers receive severance pay as other laid-off workers, such payments obviously reduce the disciplining power of unemployment. Another possible explanation for the lacking private unemployment insurance might be the following: The costs of supporting thousands of workers for months or years of unemployment in a deep slump might be very substantial. So even if the firm behaves approximately risk neutrally towards small variations in the wage level, it might be risk averse towards such substantial payments of redundancy pay. We know for instance that the possibility of bankruptcy makes a risk neutral firm behave as if it were averse against huge losses. One might then think that private unemployment insurance should be undertaken by private insurance companies with larger financial strength. However, such companies would encounter an adverse selection problem (not accounted for in the present model): Presumably the workers themselves (and their employers) know their own layoff risk better than an outside insurance company.

Still another explanation of lacking private redundancy pay is hinted at by Jon Strand (1983). Strand suggests that private unemployment insurance could drive out public unemployment insurance. I feel this could be a valid explanation, but it has not yet been explicitly modelled in a framework where the government acts rationally. I believe the way to go is to model the situation as a game between government and firms. The ability to commit oneself to a strategy will be very important for the solution. Further, I expect that a model along these lines might predict that there might be no private unemployment insurance at all. This is interesting, because the explanations above seem to explain why there is less than complete private unemployment insurance, not why there is none at all.

### Ex Post Mobility and New Entrants.

A crucial assumption in early labour contract theory is that workers are immobile ex post, or that they can commit themselves to stay on in the firm even if the wage should fall below the opportunity wage. This might be justified: We can picture that when a worker enters a firm he undertakes relationship-specific investments (not modelled) which severely limit his ex post mobility. However, I find it interesting to examine the opposite extreme assumption, that there are no mobility costs at all. I will here present a simple model based on Holmström (1983). It turns out that this situation is best modelled with production taking place in two periods. In period one the contract is entered, and production takes place with a known productivity parameter. Subscript  $s = 0$  refers to values of variables in this period. In period two, as before, one of  $s = 1, \dots, S$  states occur.

The firm's maximization problem is:

$$\text{Max } \Pi = \theta_0 f(L_0) - w_0 L_0 + \sum_{s=1}^S \phi_s \{ \theta_s f(L_0 r_s) - w_s L_0 r_s \} \quad (14)$$

s. t.

$$U = u(w_0) + \sum_{s=1}^S \phi_s \{ r_s u(w_s) + (1-r_s) u(R_s) \} > \bar{U} \quad (15)$$

$$u(w_s) > u(R_s) \quad \forall s \quad (16)$$

$$0 < r_s < 1 \quad \forall s \quad (17)$$

As the reader will have noted, I have (rather arbitrarily) chosen a model where private unemployment insurance is not allowed. Moreover, I have now (realistically) assumed that the opportunity wage  $R_s$  can vary with the state.

The first-order conditions are (with  $\lambda$  and  $\gamma$  being the Lagrangian multipliers associated with the constraints (15) and (16)):

$$- L_0 + \lambda u'(w_0) \quad (18)$$

$$\theta_0 f'(L_0) - w_0 + \sum_{s=1}^S \phi_s \{ \theta_s f'(L_s) - w_s \} r_s = 0 \quad (19)$$

$$\theta_s \{ f'(L_s) - w_s \} L_0 + \lambda \{ u(w_s) - u(R_s) \} > 0 \quad \forall s$$

$$\text{with strict equality if } r_s < 1 \quad (20)$$

$$- \phi_s L_0 r_s + \lambda \phi_s r_s u'(w_s) + \gamma u'(w_s) = 0 \quad \forall s \quad (21)$$

Now, if the constraint (16) does not bind,  $\gamma = 0$ . A comparison of (18) and (21) then yields  $w_s = w_0 \forall s$ . On the other hand, if (16) does bind, we know that  $w_s = R_s$ . Taken together this means that (21) can be replaced by:

$$w_s = \max [w_0, R_s] \quad \forall s \quad (21')$$

Let us assume that  $R_s$  and  $\theta_s/R_s$  increase in  $s$ . This means that a "better state" implies that both internal and external productivities (measured by  $\theta_s$  and  $R_s$ ) rise, while  $\theta_s$  rise the most. (21') then implies that the wage will be constant in bad states, but at a certain point it will start to rise with  $R_s$ , in order to prevent the firm from losing workers. This partial insurance arrangement is paid for by the workers accepting a lower wage in period 1 than they would otherwise have had. This is why it was essential to use a model framework with production in two periods: Insurance is prepaid.

Holmström describes this as second period wages being "upward mobile, downward sticky". I think this is an unfortunate use of terms. In a multi-period model, wages can be upward and downward mobile or upward and downward rigid over time, depending on which states occur in any pair of time periods.

Note further that if we assume there are two generations of workers, one arriving ex post, i.e., in period 2, and that there is a ban on seniority wage differentiation, then we will be in a situation almost as if all workers were ex post mobile, even though the senior workers are ex post immobile. The intuition is clear: New entrants are per definition mobile. If the firm wants to recruit new workers, it must pay them at least  $R_s$ . But as new and old workers in this setting must receive the same wage, this applies also for the immobile senior workers. A more thorough treatment of this question is found in Lommerud (1987).

More generally: The so-called insider-outsider literature (See e.g. Lindbeck and Snower (1986) and Carruth and Oswald (1986)) takes as a starting point that otherwise homogeneous workers cannot be given differentiated wages according to seniority. The traditional focus in this literature has been the involuntary unemployment among new entrants that might result. But it turns out that a ban on seniority wage differentiation also carries implications for the senior workers' opportunity to buy implicit insurance from their firm. If the wage in a state contains an implicit insurance indemnity, this benefit must also be given to new entrants. But the new entrants cannot be made to pay for this insurance: They are mobile, so they would not stay in the firm in bad states if wages fell below the opportunity wage. And they are not around in the first period to prepay for the insurance. The presence of a generation of younger workers will therefore severely limit the old workers' possibility to buy insurance from their employer. It might therefore be justified to ask if the role of the firm as an insurance contrivance has been exaggerated?

### 3. Asymmetric Information Contracts: A Blind Alley?

When early labour contract theory, properly understood, turned out to explain inefficiently high involuntary unemployment rather poorly, attention shifted to models of asymmetric information. More precisely, the firm was viewed as able to ascertain



ex post the true state of the world, while the workers were not.

In one particular setting the introduction of asymmetric information about the true state turns out not to matter. Let us start with this case.

Our starting point will be the symmetric information model with private redundancy pay allowed. This model has first-order conditions (9) - (11). Assume now that this model is altered so that only the firm observes the true state. But in this framework the firm has no incentive to misrepresent the true state of nature. The wage is not state-contingent, consequently it cannot be changed by lying about the state. But let us assume that the firm lies about the true state to alter the employment level. With symmetric information, employment is Walrasian. Reducing employment marginally saves the firm  $w - b$ . At the same time it loses production of value  $R$ . As  $w = b + R$ , this means that profits remain unchanged. Further reduction of unemployment, saves the firm  $w - b$ , but the reduction in produced value exceeds  $R$ . Hence, profit is reduced. By the same line of reasoning we could also show that it does not pay to lie to increase the employment above the Walrasian level.

This model includes three crucial assumptions:

- the firm is risk neutral
- the utility function is separable in consumption and leisure
- the firm uses redundancy payments to equalize the utility of retained and laid-off workers.

By changing any one (or several) of these assumptions, we can create models where asymmetric information matters, e.g. in the sense that the employment level will be non-Walrasian. Grossman and Hart (1981), Hart (1983) and Azariadis (1983) concentrate on a model where the firm is risk averse. Chari (1983) and Green and Kahn (1983) focus on the case where workers have a more general utility function. Only recently Oswald (1986) has inves-

tigated an asymmetric information model with no redundancy payments.

In this survey I will pay most attention to the Grossman-Hart framework, but also rather briefly outline the consequences of allowing more general utility functions.

### **The Grossman-Hart-Azariadis Version.**

All asymmetric information labour contract models more or less follow the same pattern. First, one has to construct a rationale for the firm to misrepresent the true state, either by overstating or by understating it. As only the firm observes the state, this means contracts cannot be conditioned directly on the state. The second best thing to do is to condition the contract on some observable variable that affects the firm's profits. In these labour contract models the wage is tied to the employment level, as wage and employment are the only observables. Whether under- or overemployment relative to the Walrasian employment level occurs, depends on the particular set-up. This theory is, in fact, nothing but an application of more general theories of hidden information and non-linear pricing: It is an example of how price is tied to quantity in order to convey a signal of which "type" you truly are.

I will here present a simplified model based on Grossman and Hart's work. Hopefully, this model captures some of the most important insights from this work in an easily accessible manner.

Grossman and Hart's basic premise is that the firm is risk averse. This might be because the shareholders have not been able to diversify away all the firm's risk (there is some undiversifiable risk or there are big shareholders for whom the shares in this company constitutes a significant part of their wealth), or because the firm's decisions are influenced by risk averse managers. We also know that the possibility of bankruptcy

can make the firm behave as if the firm were extremely averse of very bad states (Farmer (1984, 1985), Kahn and Scheinkman (1985)).

In the asymmetric information case, such risk aversion would imply that the firm would like its workers to take a wage cut in adverse states. But with unobservable states, this would give the firm an incentive always to claim that the state is terrible, to profit from correspondingly low wages.

For simplicity, I will assume here that there are only two states, G (Good) and B (Bad). I will also assume that the representative worker is risk neutral. Can this simplification be justified? In this model the firm wants the workers to carry some risk. One way of looking at things is the following: After providing the workers with complete insurance, the firm wants to buy back some insurance from its employees. By assuming  $u'' = 0$ , we isolate the firm's insurance purchase from the workers - and it is this purchase that is particularly interesting in this framework. The simplifying assumption of risk neutral workers highlights the huge difference between the early contract theory and the Grossman-Hart model. In the symmetric information model the key issue was the workers' desire to buy insurance from their employer. Now we focus on the firm's desire to buy insurance. And even if we assume here that this insurance is bought from the employees of the firm, it could just as well have been from anybody else unable to observe the true state. The key results would still remain (Hart and Holmström (1985)).

The firm is taken to maximize expected utility of profits,  $\Pi$ .

$$\text{Max } \Pi = v(\pi_B)\phi_B + v(\pi_G)\phi_G$$

$$\text{where } \pi_s = \theta_s f(L_s) - W_s \quad s = B, G \quad (22)$$

Here  $W_s$  denotes the total wage bill paid in state  $s$ . We assume this wage bill is divided equally among the workers under contract, either because there is worksharing or redundancy pay-

ment. One convenient way to model this is to assume that there is one worker who can sell more or less of his time to the firm. The reservation price of a unit time is  $R$ . This means that to attract units of labour from the (risk neutral) worker, the following condition must be fulfilled:

$$\phi_B(W_B - RL_B) + \phi_G(W_G - RL_G) > 0 \quad (23)$$

The firm also has to satisfy the following truth-telling (or incentive compatibility) constraints:

$$\phi_G(\theta_G f(L_G) - W_G) > \phi_G(\theta_G f(L_B) - W_B) \quad (24)$$

$$\phi_B(\theta_B f(L_B) - W_B) > \phi_B(\theta_B f(L_G) - W_G) \quad (25)$$

Condition (24) says that if the good state has occurred, the firm should not be able to increase its profits by reporting B instead of G. Condition (25) has a similar interpretation. From the discussion above, we should expect (24) to bind and not (25). This is indeed the case, which also can be shown formally (along the lines of Hart (1983, Appendix 3)). The binding condition (24) says that the firm must be induced not to under-represent the state.

It might be instructive to rewrite (24) as:

$$\theta_G(f(L_G) - f(L_B)) > W_G - W_B \quad (24')$$

This implies that if we want  $W_B$  to be lower than  $W_G$ , then incentive compatibility demands  $L_B < L_G$ . In this sense we can view the firm as facing an increasing function  $W(L)$ , or in other words, the wage to be conditioned on the employment level.

Note that condition (24') implies something about the gap between  $W_G$  and  $W_B$  for given  $L_G$  and  $L_B$ . It says nothing about the absolute level of  $W_G$  and  $W_B$ . This is determined by (23). We should therefore expect (23) to be binding, which can also be shown formally from the first-order conditions of the problem.

I will now take the firm's optimization problem to be the maximization of (22) subject to (23) and (24). The first-order conditions for an interior optimum are (with  $\lambda$  and  $\alpha$  being the Lagrangian multipliers associated with the constraints):

$$v'(\pi_G)\phi_G\theta_G f'(L_G) - \lambda\phi_G R + \alpha\phi_G\theta_G f'(L_G) = 0 \quad (26)$$

$$v'(\pi_G)\phi_B\theta_B f'(L_B) - \lambda\phi_B R - \alpha\phi_G\theta_G f'(L_B) = 0 \quad (27)$$

$$- v'(\pi_G)\phi_G + \lambda\phi_G - \alpha\phi_G = 0 \quad (28)$$

$$- v'(\pi_B)\phi_B + \lambda\phi_B + \alpha\phi_G = 0 \quad (29)$$

From (28) we have that  $\lambda = v'(\pi_G) + \alpha$ . Using this, (26) can be written as:

$$\theta_G f'(L_G) = R \quad (30)$$

This means that we will have Walrasian employment in the good state.

Correspondingly, from (29) we have  $v'(\pi_B) = \lambda + \alpha \frac{\theta_G}{\theta_B}$ . Inserting into (27) and rearranging terms, we get:

$$\theta_B f'(L_B) = R - \frac{\alpha}{\lambda} \frac{\phi_G}{\phi_B} (\theta_B f'(L_B) - \theta_G f'(L_B)) \quad (31)$$

This implies:

$$\theta_B f'(L_B) > R \quad (32)$$

In the bad state there will be underemployment relative to the Walrasian level.

One way of interpreting this result is the following: Only in the good state has the firm an incentive to misrepresent the state. But as it is less costly for the firm to cut employment

in the bad state than in the good one, an employment cut can therefore be used to prove that the adverse state really has occurred. We have modelled the firm as employing only one input, namely labour. A firm which uses many inputs can distort any or several of its input decisions to signal that it tells the truth, not only the employment of labour. Note how closely related this is to e.g. Rothschild and Stiglitz' (1976) account of how price- quantity ties can be used in the insurance market to segregate bad and good risks.

Moreover, note that the firm is worse off under asymmetric than under symmetric information. It must use resources to prove it is telling the truth. Formally, the asymmetric information problem equals the symmetric one with an added constraint. Hence, the opportunity set will generally be smaller. This means that the firm would like to give the workers insight into the real situation of the firm, were it able to. Public auditing, the representation of labourers on boards, etc., might be seen as attempts to lessen the costs of asymmetric information.

#### **Discussion.**

Of course, a crucially important question in this context is whether, in fact, there is, or is not, asymmetric information in the sense described. Note that this not only requires that there is asymmetric information at the time the contract is entered into, but also that the truth will not be disclosed at any later stage. E.g., the quality of the state should influence profits. And a large firm with many shareholders would presumably find it troublesome in the long run both to disguise its true profits and at the same time let the shareholders benefit from higher profits. And an optimal contract could include punishments for a lying firm, effective from the time the true state is deduced, which would prevent the firm from cheating without the employment level having to be distorted. Kovenock and Sparks (1985) have pointed out that employee stock ownership plans might play a role as an "automatic punishment" of the firm: Profit cannot

accrue to the shareholders without the workers also benefiting. Admittedly, things would be different in small family companies or in firms which do not expect to stay in business. Here asymmetric information might play a more important role.

Because of arguments like these, Oliver Hart has recently begun to model asymmetric information models where it is the managers who have the better information. Managers, supposedly, can consume invisibly any extra profits generated from cheating, e.g. by working less hard. But there are problems also with this kind of models. We know that shareholders (or anybody else who partially or wholly acts as insurers for the managers) can discipline the managers in two ways. First, corresponding to the model above, they can use direct incentive schemes, which might call for the managers to distort labour use (or use of any other input) to prove that a bad state has occurred. But managers are also disciplined indirectly by the managerial labour market (see e.g. Fama (1980)). If future employers of the managers judge managerial quality according to attained results, presumably this would be a very effective check against the managers understating the state too much and "eating up" the profits.

Furthermore, a firm normally has a whole hierarchy of managers, perhaps with conflicting interests and personalities. Managerial asymmetric information models tacitly assume that all these managers collude in misrepresenting the true state. But with many managers and extensive lying, there is always the chance that the truth will be revealed.

And finally, large scale managerial consumption of company profits would, I think, be rather conspicuous. The opposite assumption, that such consumption is invisible, can certainly be true only within limits.

Another objection against this theory could be the following: One might find that it is not credible for workers to demand that their employer should sack some of them in order to prove that a bad state has occurred. One interpretation of such a

statement might be this: In a model slightly different from that above, with no redundancy pay to laid-off workers, the laid-off workers will envy those retained. It might be impossible ex ante to write binding contracts which specify such ex post inefficiencies. We have so far, naively, assumed that any contract can be enforced. In Section 4 we will turn to what I find is one of the most interesting issues in contract theory, namely the forms contracts might take when there are limits to their enforceability.

Taking these reservations together I believe that perhaps too much research effort has gone into investigating models of this kind, but the reader will note that I cautiously have put a question mark after "blind alley" in this section's title.

#### **The Chari-Green-Kahn Version.**

So far, we have restricted our attention to the case where the workers' utility is separable in consumption and leisure. I.e., the money value of leisure is independent of the wage income level. If most of the utility accruing to a laid-off worker stems from unemployment benefits or wages earned in alternative employment, this might be a valid formulation. If not, this is a simplification. Had we used the more general utility function  $u(w_s l_s, -l_s)$  (where  $l_s$  is individual labour supply in state  $s$ ), all our results would have been more complicated. Among other things, the firm would then have an incentive to misrepresent the true state of the world even if it is risk neutral. Chari (1983) and Green and Kahn (1983) have used this to construct a model of asymmetric information labour contracts along a route alternative to that of Grossman and Hart.

Let us take as an example the utility function  $u_s = g(w_s l_s) - l_s$ , where  $g' > 0$  and  $g'' < 0$ . In a model with symmetric information, with a risk neutral firm, and in which the firm buys labour only from one worker (which, as mentioned above, in an economic



sense corresponds to worksharing being allowed), the maximization problem will be as follows:

$$\text{Max}_{W_s, L_s} \quad \Pi = \sum_s \phi_s \{ \theta_s f(L_s) - W_s \} \quad (33)$$

$$\text{s. t.} \quad U = \sum_s \phi_s \{ g(W_s) - L_s \} > \bar{U} \quad (34)$$

The first order conditions are:

$$g'(W_s) = \frac{1}{\lambda} \quad (35)$$

$$\theta_s f'(L_s) = \lambda \quad (36)$$

These conditions imply that total money income  $W_s$  will be held constant across states, whereas the labour supply will be larger the better the state. This is a quite natural result given the utility function: The worker has a preference against variations in  $W_s$  over states, but is indifferent as to uncertainty about  $L_s$ . Therefore, he is insured against variations in  $W_s$ , but the firm wants him to work harder the higher is his productivity.

Now, the important thing is that this gives the firm an incentive to overrepresent the true state: In a better state the worker is given the same income, but works harder. In an asymmetric information model, this result will lead to overemployment in good states. The firm must buy more labour in good states to prove that it is speaking the truth. Chari (1983) and Green and Kahn (1983) have shown that this kind of reasoning goes through as long as leisure enters the utility function as a normal good. If we combine the Grossman-Hart model and the Chari-Green-Kahn model (risk averse employers, more general utility functions) and allow for many states of the world, we will find that it is indeterminate whether we get over- or underemployment relative to the Walrasian level. The reader must determine for himself (herself) which of these two effects empirically is likely to play the most important role.

#### 4. Contracts with Transaction Costs.

Transactions are not costless. It is costly to foresee any eventuality relevant to the transaction, to write the contract, to gather information, to verify what actually has happened, and to enforce the contract. These costs are referred to as "transaction costs". In economics this concept is perhaps foremostly associated with the name of Oliver Williamson (e.g. 1975, 1983). A parallel focus within social anthropology is associated with Frederik Barth (e.g. 1966).

Many different attempts have been made at including such costs in a theory of transactions, or "contracts". But to model these costs in anything near a general way would be extremely complex. Instead, numerous more ad hoc models have been developed, focusing on more partial aspects of the problem. I will here limit myself to investigating two such sets of models. First, I will study models of "reputational enforcement", i.e., how the fear of loss of reputation partly can substitute for verifiable and enforceable contracts. Further, I will look at to which degree the use of noncontingent contracts in combination with damage measures can substitute for complex contingent contracts. This approach excludes the discussion of many related issues of interest, see e.g. Hart and Holmström (1985) for more complete references. Note also that we have defined the gathering of information as a transaction cost, in this sense the literature on asymmetric information contracts can be seen as a variant of transaction costs contract theory.

##### **Reputational enforcement**

We will take as our starting point that contracts cannot be enforced by a third party. This rules out the writing of contracts in the conventional sense. But we can still have self-enforcing agreements, i.e., agreements which neither party has an incentive to deviate from. Much attention has been paid to

the case where the fear of not being able to recruit business partners in the future disciplines the parties from breaching the contract. This we will call reputational enforcement. We will discuss this phenomenon with basis in a model which can be seen as a labour market adaption of Klein and Leffler's (1981) simple and seminal model of reputational enforcement in the market for high quality goods. Reputational enforcement in the labour market has also been discussed e.g. by Holmström (1981), Strand (1984), Carmichael (1984) and Bull (1987).

When a firm enters into business it undertakes a nonrecoverable investment  $I$ . At later stages it uses labour as its only input.  $L_t$  is the employment level at time  $t$ . The model is set in discrete time. Workers who at the outset of a period take on employment in a firm are completely immobile for the duration of the period. The discount factor  $\delta$  measures not only the time preferences of the agents, but also length of period. Ceteris paribus, a lower discount factor signifies longer periods, and thereby more serious immobility.

We assume that the production function  $f(\cdot)$  is concave with positive first-order derivatives. The price of the firm's single output is 1 at all dates.

We can now express the net present value of profits as:

$$\Pi = - I + \sum_t \delta^{t-1} \{ f(L_t; I) - w_t L_t \} \quad (37)$$

$w_t$  denotes wages paid at time  $t$ .

We will now assume that the firm cannot commit itself at the start of a period to pay a certain wage in that period. As workers are immobile, this leaves them in a vulnerable position. Of course, the assumption that wages are noncontractable is only credible if "wages" is interpreted in a broad sense, including e.g. job satisfaction, promotion possibilities, etc., or that the contract is very long term.

To highlight some main insights, I will model workers' expectations in a somewhat ad hoc manner. I will assume that a firm which in one period pays a low wage, in the next period suffers from a bad reputation, and consequently experiences problems in keeping and recruiting workers.

We assume that workers have adaptive expectations. With the simplest form of adaptive expectations, we can express the firm's "reputation" (i.e. what all workers expect the firm to pay in the future) as

$$R_t = w_{t-1} \quad \forall t \neq 1 \quad (38)$$

An alternative concept of reputation would be to envisage that every individual worker has his private view of how the firm has treated him and other workers in the past. It would then be interesting to study the process of how a reputation spreads (see e.g. Strand (1984)).

With adaptive expectations it is somewhat difficult to model the reputation of a newly started firm. I will here assume that there exists a firm which in all respects (except starting date) is equal to the newly started firm. If these firms have the same levels of initial investment, I assume that they have the same reputation. These assumptions mean that when it comes to reputation building it is "as if" a firm has existed forever. I agree that this treatment of a newly started firm is not wholly satisfactory, but no obvious alternative stands out.

We now postulate that the firm faces the following labour supply function:

$$L_t^S = \gamma(R_t - \bar{R}) \quad \forall t \quad (39)$$

$L_t^S$  denotes labour supply.  $\bar{R}$  denotes a "market level" of reputation.  $\gamma$  is a function which measures how strongly the firm will lose or attract workers if its reputation deviates from  $\bar{R}$ . We

can imagine that workers to a varying degree have undertaken relationship-specific investments in this or other firms. This justifies thinking of  $\gamma$  as being a smooth function. Alternatively, with free mobility at the end of a period a firm would have been able to recruit as many workers as wanted if  $R_t > \bar{R}$ , and none if  $R_t < \bar{R}$ .

The firm cannot buy more labour than  $L_t^S$ . Therefore we know:

$$L_t < \gamma(R_t - \bar{R}) \quad \forall t \quad (40)$$

The firm's problem now is to maximize (37) given the reputation building process and the labour supply constraint. Let  $\lambda_t$  be the Lagrangian multipliers associated with (40). For the time being we will assume that the first-order conditions describe an interior optimum. Two first-order conditions, with respect to employment and wage in period  $t$ , will be:

$$\delta^{t-1} \left( \frac{\partial f}{\partial L_t} - w_t \right) - \lambda_t = 0 \quad (41)$$

$$\delta^{t-1} (-L_t) + \lambda_{t+1} \gamma' = 0 \quad (42)$$

Note from (41) that if (40) is not binding ( $\lambda_t = 0$ ), we will have that  $\frac{\partial f}{\partial L_t} = w_t$ . We will argue below that this is not incentive compatible. Hence,  $\lambda > 0$ .

We now solve (41) for  $\lambda_{t+1}$  and substitute into (42). This yields:

$$L_t = \delta \left( \frac{\partial f}{\partial L_{t+1}} - w_{t+1} \right) \gamma' \quad (43)$$

As long as the technology, the discount factor and  $\gamma$  remain unchanged over time, the problem is symmetric over time periods (It is perhaps especially taxing to assume that  $\gamma$  is constant

across periods.). Therefore,  $L_t = L \forall t$  and  $w_t = w \forall t$ . We can now rewrite (43) as:

$$\frac{\partial f}{\partial L} - w = \frac{L}{\delta \gamma} \quad (44)$$

One immediate consequence is that in equilibrium we must have that

$$\frac{\partial f}{\partial L} > w \quad (45)$$

The firm must earn positive profit from the marginal worker. This is a rather intuitive proposition. Condition (43) says that in equilibrium the short run gain from lowering wages in one period must be offset exactly by the loss in the next period due to problems of recruiting workers. But had we been in a Walrasian equilibrium where  $\frac{\partial f}{\partial L} = w$ , the firm would earn zero profit from the marginal worker. The possibility of a marginal reduction of the workforce could therefore not discipline the firm from lowering wages. In a reputational equilibrium we must have that the firm earns positive profit from the marginal worker. This implies that (40) must be binding, and  $\lambda_t > 0$ .

Does this mean that the firms earn positive profits in equilibrium? If that is the case, we should experience infinite entry of new firms into this industry. In our setting profits cannot be competed away by raising wages relative to labour's value marginal product. This would simply not be incentive compatible. Therefore profits must be dissipated by driving the nonrecoverable investments,  $I$ , to a higher level than would have been chosen in a setting where enforceable contracts could have been entered. Therefore this model does not entail positive profits in a true sense, but "appropriable quasirents", to use a concept introduced by Klein, Crawford and Alchian (1978). Klein and Leffler (1981) picture these nonrecoverable investments as investments in advertising or other sales expenses. As an alternative, Shapiro (1983) has suggested that these investments could be investments in building up a reputation: In his model a high

quality producer initially must sell high quality goods at the low quality price in order to gain recognition.

It seems to be a tacit assumption in the literature that the size of these investments has a neutral effect on the firm's labour employment decision. That being the case, it can be seen directly from (45) that our reputational model implies lower employment than in a model with complete contracts.

So far we have assumed that the first-order conditions of the problem describe an interior optimum. This is not necessarily the case. E.g., there seems to be no natural assumption to impose on  $\gamma$  which ensures that the maximand is concave. More important economically: It might be that in some industries the firm's possibilities of short-run gains by exploiting workers' immobility were so large that no equilibrium with positive employment did exist.

Note that we here have used a framework with non-unionized workers. The ensuing reputational equilibrium does not entail contracts in a proper sense. The workers do not think of themselves as having entered an implicit contract with the firm, they just have adaptive expectations, and thereby a low-wage firm will be punished subsequently by an outflow of workers. To me, this is an attractive side of the model. One objection against much contract theory is that workers do not seem to be aware that they have entered an elaborate implicit contract with their firm. Such a criticism is not valid against a model of the present type. But admittedly, others might have preferred to analyze a situation where the employer and the workers (perhaps represented by a union) rationally played a game against each other, or struck a bargain, or "contract", in a more proper sense.

The main insight of this model is perhaps that a Walrasian equilibrium has no disciplining power. The threat of terminating a business relationship is an empty one when the parties earn zero profit/utility from a marginal relationship. This simple and

intuitive idea underlies many recent articles. In a labour market context it is interesting to point to Shapiro and Stiglitz' (1984) version of the efficiency wage model, namely the so-called "shirking model". In this model it is the reverse need for disciplining which is investigated, i.e., the employer's need to ensure that the workers perform. Because of a transaction cost (costly monitoring), the firm must resort to the threat of terminating the employment relationship with workers caught shirking. As this threat would be empty in a Walrasian equilibrium, it turns out that the need for discipline leads to an equilibrium with unemployment.

### **Reliance and Damage Measures**

The reputational model above was very extreme in that it assumed that transaction costs blocked any opportunity for writing contracts. The literature on damage measures takes a perhaps more sober approach: It is possible to write enforceable contracts, but it is not possible to make them contingent on which uncertain state of the world which eventually materializes.

Key articles in this literature are Shavell (1980, 1984) and Rogerson (1984). Neither of these articles is specially tailored to a labour market setting. I have here chosen to present a simple model due to Shavell (1984), and to discuss the relevance for the labour market below.

We study the interaction of two risk-neutral, wealth maximizing parties, a buyer and a seller. The buyer must undertake a sunk, relationship-specific investment before he finds out whether or not the seller will perform. In this literature, this investment is called "reliance expenditures", or simply "reliance". After the buyer has undertaken this investment, the seller observes the outcome of an uncertain contingency. E.g., if we study a production contract, he will learn his true production cost,  $c$ . Ex ante  $c$  is known only up to a probability distribution. If the seller then performs, the buyer enjoys a benefit which Shavell



calls the "expectancy", denoted  $v$ . Let us further denote  $B$  as the set of contingencies under which the seller will not perform "the breach set", and  $k$  as the contract price.

It is now easy to show, and rather intuitive, that a Pareto efficient contract will entail that the seller will not perform when production cost exceeds the buyer's expectancy, i.e.,  $B^* = \{c \mid c > v\}$ .

Shavell now assumes that due to transaction costs the seller and the buyer can only enter noncontingent contracts. However, they can also agree in advance that the seller shall pay a damage measure to the buyer in case the seller does not perform. Shavell then goes on to discuss the size of the breach set under some commonly used damage measures relative to that under Pareto efficient contracts.

First of all he discusses specific performance, which means that the seller is never allowed not to fulfill the contract. The breach set is empty, and therefore obviously too small relative to the Pareto efficient one. Against this, he discusses three different damage measures, the restitution measure, the reliance measure, and the expectation measure.

Under the restitution measure, a nonperforming seller must return the payment  $k$ . His breach set becomes  $B^* = \{c \mid c > k\}$ . Under the reliance measure, the seller must return payment  $k$  and compensate the reliance expenditure  $r$  (assumed to be measurable and verifiable). The breach set becomes  $B^* = \{c \mid c > r+k\}$ . Under the expectation measure, the seller must pay a court's estimate of the buyer's expectancy  $v$ . Let us denote this estimate  $u$ . The breach set becomes  $B^* = \{c \mid c > u\}$ .

Shavell now assumes that  $u$  always is larger than  $r + k$ , which seems a natural assumption. We then have:  $u > r + k > k > 0$ .

We then have that the breach set under the expectation measure is smaller than that under the reliance measure, which again is smaller than that under the restitution measure.

Our first observation is that if  $u$  is a close estimate of  $v$ , the expectancy measure will take us very close to a Pareto efficient situation. Second, note that if  $u$  underestimates  $v$  all three damage measures will yield too large breach sets, but the expectancy measure is closest to the optimal. Whether we would want to use this measure or specific performance, depends on the relative costs of having a too small or a too large breach set. On the other hand, if  $u$  overestimates  $v$ , the expectancy measure will lead to a too small breach set, but not so small as that under specific performance. Which is better of the expectancy measure and the reliance measure in this case, again depends on the relative costs of too small and too large breach sets.

We have seen that when only noncontingent contracts can be entered, this can to some extent be remedied by the use of damage measures. In some types of contracts the use of damage measures seems widespread. But is this a relevant insight concerning the workings of the labour market? I believe so, even though the situation in the labour market seems more complicated than Shavell's work indicates. A common arrangement in the labour market is to enter a noncontingent contract, but to leave both parties the option of terminating the relationship. In order to induce correct investments in relation-specific capital, the terminating party must pay a "termination penalty", or "damage measure". This can take the form of a direct payment of money, as when a firm pays a fired manager a "golden handshake", or more indirect forms, as when a quitting worker forfeits seniority wage rises or promotion possibilities. One complication is that in the labour market it is not always easy to ascertain which party really initiated the termination. I will not go into details on this question, only refer to the literature on investments in firm-specific human capital, which deals precisely with questions like these, notably Hashimoto (1981), Hashimoto and Yu (1980) and Carmichael (1984). As an aside: I find it

rather peculiar how few cross-references there are between the human capital literature and the labour contract literature, considering that much the same questions are studied.

#### 5. A Concluding Remark.

I have here tried to survey some of the very numerous contributions to labour contract theory. As I pointed out at the outset of this survey, "labour contract theory" seems to become more and more synonymous with "labour economics". All the same, I think we can group the reviewed literature in two categories: In some contributions it is the desire for risk-shifting which motivates long-term contracts, in others it is the desire to protect investments in relation-specific capital. Especially the risk-shifting version of the theory now seems to be reasonably well understood. However, I believe that many interesting questions in contract theory still remain unanswered.

Especially, I think it would be interesting to bring these two major strands of contract theory closer together. E.g., what is the scope for risk-shifting between employer and employee within the rather complex institutional set-up predicted by some of the literature on investments in relation-specific human capital?

References.

- Arrow, K. J. (1971), Essays in the Theory of Risk-bearing, Chicago: Markham.
- Azariadis, C. (1975), "Implicit Contract and Underemployment Equilibria", Journal of Political Economy 83, 1182-1202.
- Azariadis, C. (1981), "Implicit Contracts and Related Topics: A Survey", in Z. Hornstein et al. (eds.), The Economics of Labour Markets, HMSO.
- Azariadis, C. (1983), "Employment with Asymmetric Information", Quarterly Journal of Economics 18, Supplement, 157-172.
- Baily, M. N. (1974), "Wages and Employment Under Uncertain Demand", Review of Economic Studies 41, 37-50.
- Barth, F. (1966), Models of Social Organization, London: Royal Anthropological Institute of Great Britain and Ireland.
- Borch, K. (1962), "Equilibrium in a Reinsurance Market", Econometrica 30, 424-444.
- Bull, C. (1987), "The Existence of Self-Enforcing Implicit Contracts", Quarterly Journal of Economics 102, 147-159.
- Burdett, K., and R. Wright (1986), "The Effects of Unemployment Insurance on Layoffs, Hours per Worker, and Wages", mimeo, University of Cornell.
- Carmichael, H. L. (1983), "Firm Specific Human Capital and Promotion Ladders", Bell Journal of Economics 14, 251-258.
- Carmichael, H. L. (1984), "Reputation in the Labor Market", American Economic Review 74, 713-725.

- Carruth, A. A. and A. Oswald (1986), "On Union Preferences and Labour Market Models: Insiders and Outsiders", Working Paper No. 823, Centre for Labour Economics, London School of Economics.
- Chari, V. V. (1983), "Involuntary Unemployment and Implicit Contracts", Quarterly Journal of Economics 98, Supplement, 107-123.
- Fama, E. (1980), "Agency Problems and the Theory of the Firm", Journal of Political Economy 88, 288-307.
- Farmer, R. E. A. (1984), "Unemployment, Bankruptcy and Asymmetric Information", The Manchester School of Economic and Social Studies 52, 284-291.
- Farmer, R. E. A. (1985), "Implicit Contracts with Asymmetric Information and Bankruptcy: The Effect of Interest Rates on Layoffs", Review of Economic Studies 52, 427-442.
- FitzRoy, F. R. and R. A. Hart (1985), "Hours, Layoffs and Unemployment Insurance Funding: Theory and Practice in an International Perspective", Economic Journal 95, 700-713.
- Geanakoplos, J. and T. Ito (1985), "Implicit Contracts with Involuntary Unemployment in General Equilibrium", mimeo. Earlier version circulated as Discussion Paper, May 1981.
- Gordon, D. F. (1974), "A Neo-Classical Theory of Keynesian Unemployment", Economic Inquiry 12, 431-459.
- Green, J. and C. M. Kahn (1983), "Wage-Employment Contracts", Quarterly Journal of Economics 98, Supplement, 173-188.
- Grossman, S. and O. Hart (1981), "Implicit Contracts, Moral Hazard and Unemployment", American Economic Review 71, Papers and Proceedings, 301-307.

- Hart, O. (1983), "Optimal Labour Contracts under Asymmetric Information: An Introduction", Review of Economic Studies 50, 3-35.
- Hart, O. and B. Holmström (1985), "The Theory of Contracts", mimeo.
- Hashimoto, M. (1981), "Firm-Specific Human Capital as a Shared Investment", American Economic Review 71, 475-482.
- Hashimoto, M. and B. T. Yu (1980), "Specific Capital, Employment Contracts, and Wage Rigidity", Bell Journal of Economics 11, 536-549.
- Holmström, B. (1981), "Implicit Contracts, Moral Hazard and Unemployment", American Economic Review 71, Papers and Proceedings, 301-307.
- Holmström, B. (1983), "Equilibrium Long-Term Labor Contracts" Quarterly Journal of Economics 98, Supplement, 23-54.
- Kahn, C. M. (1985), "Optimal Severance Pay with Incomplete Information", Journal of Political Economy 93, 435-351.
- Kahn, C. M. and J. Scheinkman (1985), "Optimal Employment Contracts with Bankruptcy Constraints", Journal of Economic Theory 35, 343-365.
- Klein, B., R. A. Crawford and A. A. Alchian (1978), "Vertical Integration, Appropriable Rents, and the Competitive Contracting Process", Journal of Law and Economics 21, 297-326.
- Klein, B. and K. B. Leffler (1981), "The Role of Market Forces in Assuming Contractual Performance", Journal of Political Economy 89, 615-41.

- Kovenock, D., and R. Sparks (1985), "An Implicit Contract Approach to Employee Stock Ownership Plans", Discussion Paper No. 276, University of Essex.
- Lindbeck, A. and D. Snower (1986), "Wage Setting, Unemployment and Insider-Outsider Relations", American Economic Review 76, Papers and Proceedings, 235-239.
- Lommerud, K. E. (1986), "Worksharing versus Layoffs in Labor Contracts", mimeo, Norwegian School of Economics and Business Administration.
- Lommerud, K. E. (1987), "Labour Contract Theory and the Insider-Outsider Dilemma", mimeo, Norwegian School of Economics and Business Administration.
- Moore, J. (1985), "Optimal Labour Contracts when Workers Have a Variety of Privately Observed Reservation Wages", Review of Economic Studies 52, 37-67.
- Mortensen, D. T. (1978), "On the Theory of Layoffs", Discussion Paper 332, Northwestern University.
- Oswald, A. J. (1985), "The Economic Theory of Trade Unions: An Introductory Survey", Scandinavian Journal of Economics 87, 160-193.
- Oswald, A. J. (1986), "Unemployment Insurance and Labor Contracts under Asymmetric Information: Theory and Facts", American Economic Review 76, 365-377.
- Rogerson, W. P. (1984), "Efficient Reliance and Damage Measures for Breach of Contract", Rand Journal of Economics 15, 39-53.
- Rosen, S. (1985), "Implicit Contracts: A Survey", Journal of Economic Literature 23, 1144-1175.

Rotschild, M. and J. E. Stiglitz (1976), "Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information", Quarterly Journal of Economics 91, 629-650.

Sargent, T. (1979), Macroeconomic Theory, New York: Academic Press.

Shapiro, C. (1983), "Premiums for High Quality Products as Returns to Reputations", Quarterly Journal of Economics 98, 659-679.

Shapiro, C. and J. E. Stiglitz (1984), "Equilibrium Unemployment as a Worker Discipline Device", American Economic Review 74, 433-444.

Shavell, S. (1980), "Damage Measures for Breach of Contract", Bell Journal of Economics 11, 466-490.

Shavell, S. (1984), "The Design of Contracts and Remedies for Breach", Quarterly Journal of Economics 99, 121-148.

Strand, J. (1983), "Layoffs, Labor Productivity, and Worker Seniority Rules: An Analysis of Some 'Micro-Micro' Institutional and Long-run Equilibrium Aspects of Implicit Contract Theory", Memorandum 20/1983, University of Oslo.

Strand, J. (1984), "The Structure of Implicit Contracts with Word-of-Mouth Reputational Enforcement", mimeo, University of Oslo.

Williamson, O. (1975), Markets and Hierarchies: Analysis and Antitrust Implications, New York: Free Press.

Williamson, O. (1985), The Economic Institutions of Capitalism, New York: Free Press.



**Chapter 2:**

**LABOUR CONTRACT THEORY AND THE INSIDER-OUTSIDER DILEMMA**

October 1986

Revised August 1987

LABOUR CONTRACT THEORY AND THE INSIDER-OUTSIDER DILEMMA

ABSTRACT

This paper brings together some central elements both of labour contract theory and so-called insider-outsider theory. We study labour contracts with two generations of workers where there is a ban on differentiating wages between otherwise homogenous workers of different seniority. Two basic results emerge: First, there might be a region where wages fall as the state improves. Crudely put, this originates from the old workers being reluctant to share with new entrants the insurance indemnities which in some states constitute a part of the wage. Moreover, even if we assume the old workers to be ex post immobile (as in the original implicit labour contract theory), we get a contract structure of the same form as if we had assumed them to be ex post mobile. Again crudely put: New entrants are by necessity ex post mobile. If we tie the wages of old and new workers to each other, we will then, when any new entrants are hired, get a situation as if all workers were mobile. My conclusion is that within this framework the scope for the firm acting as an insurer of labour income is much narrower than in standard contract theory.

## LABOUR CONTRACT THEORY AND THE INSIDER-OUTSIDER DILEMMA

### 1. Introduction.

This paper attempts to bring together some of the central elements from standard implicit contract theory and so-called insider-outsider theory<sup>1)</sup>.

Early work in what rather misleadingly is called "implicit" contract theory<sup>2)</sup>, as e.g. Azariadis (1975) and Baily (1974), focuses on how the firm can play a role as an insurer against variations in the workers' labour income. The basic premise is that the firm can overcome the moral hazard problems which prevent the workers from buying such insurance from an ordinary insurance company.

This line of work has not very much to say about the relationship between different vintages of workers. Among the classical papers in the tradition only Holmström (1983) allows for more than one vintage of workers. However, in Holmström's model there are no limits to differentiating wages between different worker vintages. This is in contrast with the standard assumptions in insider-outsider theory. Also, Tracy and Woglom (1984) study unionized contracts with more than one vintage of workers. But again there is no limit to differentiating wages according to seniority.

Let us now turn to the insider-outsider literature. Here the focus is on the possible conflict of interest between old, seasoned workers and new entrants to the labour market. This in contrast to the main body of labour economics, which takes the conflict between the firm and its workers as the central clash of interest. Several pioneering articles in this field have been written by Assar Lindbeck and Dennis Snower (hereafter sometimes denoted LS). (LS (1984 a and b), (1985 a, b), (1986).) Other relevant work include Shaked and Sutton (1984), Oswald (1984),

Tracy and Woglom (1984), Sampson (1985), Solow (1985), Carruth and Oswald (1987), and Blanchard and Summers (1987). Also, the relationship between the original workers of a labour-managed enterprise and possible new hands might be seen as an example of an insider-outsider dilemma. (See Ireland and Law (1982) and the references therein).

The Lindbeck-Snowier story has the following main building blocks: The old, seasoned workers have some "edge" over possible new entrants, due to some kind of labour turnover costs. These costs could e.g. be firing/hiring/training costs (LS (1984a)), costs connected with high labour turnover affecting insider morale negatively (LS (1984b)), or that the insiders can withdraw cooperation from or directly harass new workers who replace old ones (LS (1985a)). The old workers are capable of extracting at least some of the rent arising from this bilateral monopoly situation between the firm and old workers. In addition, there are strict limits to wage discrimination against new workers. Lindbeck and Snowier's particular argument here is that after a short period the new entrants cannot be distinguished from the insiders. Further, the entrants cannot sign credible long-term contracts where they promise to work permanently for less than the insiders. That is, the entrants cannot restrain themselves in advance from using the bargaining power they will get as soon as they themselves become seasoned workers. Note that even if the Lindbeck-Snowier story explicitly rules out long-term contracts, this is not an essential assumption per se, but only a link in establishing limits to seniority wage differentiation.

The stated assumptions lead to a situation where the old workers use their bargaining power to push up wages, and as wage discrimination against new entrants is possible only to a very limited extent, these high wages lead to involuntary unemployment among new workers.

As contract theory is weak on the relationship between insiders and outsiders, the insider-outsider theory is mute about the possible insurance elements in a contract between the firm and

the old workers. It should therefore be some scope for merging these two bodies of literature.

Crudely put, the main argument of this paper is the following: In contract theory of the Azariadis-Baily type, wages are not only payment for work done, but includes an insurance premium or indemnity. When the wage for new workers is tied to that for old workers, this might distort the old workers' possibility to buy insurance from their employer. Exactly how this happens depends on the particular assumptions made.

As there is a wide variety of labour contract models, and also several versions of the insider-outsider theory, it is not immediately clear how these theories should best be merged. I have chosen to develop two versions of my model. The two versions differ in that in the first model senior workers are assumed to be ex post mobile, whereas they in the second are ex post immobile. This turns out to be an interesting distinction.

## 2. A Modified Holmström Model

I will here present a model which follows the framework of Holmström (1983) rather closely, with the main exception that there is a ban on seniority wage differentiation.

The model is a two period one, and production takes place in both periods.

The firm's goal is to maximize profit, expressed as

$$V = \theta_0 f(l_0) - w_0 l_0 + \sum_{s=1}^S \{ \theta_s f(\delta l_0 + l_s^+) - w_s (\delta l_0 + l_s^+) \} \phi_s \quad (1)$$

Here,

- $V =$  expected profit (the firm is risk neutral)  
 $\theta_0 =$  a known productivity parameter in period 1  
 $f =$  the production function, we assume  $f' > 0$ ,  $f'' < 0$ .  
 $\lambda_0 =$  the number of workers employed in period 1 (all workers work an institutionally fixed workday)  
 $\theta_s =$  the productivity parameter in period 2, given state  $s$   
(There are  $S$  mutually exclusive states of the world which can occur in period 2, numbered such that  $\theta_1 < \theta_2 < \dots < \theta_S$ )  
 $(1-\delta) =$  the percentage of first period workers who, for exogenous reasons, quit between periods 1 and 2. Every worker has the same probability of being one of those who quit.  
 $\lambda_s^+ =$  the number of new workers hired in state  $s$   
 $w_0 =$  the wage in period 1  
 $w_s =$  the wage in period 2, given state  $s$ , to both old, retained workers and to new ones  
 $\phi_s =$  the probability of state  $s$  occurring

Note that we assume that the firm's discount factor is 1. We will also assume that this is the workers' discount factor.

I restrict attention to those states where at least one new worker is hired,  $\lambda_s^+ > 0$ . The new insights of this paper are only relevant in these states, so I have chosen this assumption to simplify the exposition. Moreover, this might even be a realistic assumption if  $(1-\delta)$  is relatively large. Below, I will make some informal remarks about what can be expected to happen in deep slumps, when no new hirings are made. Further, I assume that old workers never are fired to be replaced by new ones. This seems reasonable, as long as we have specified that old and new workers receive the same wage. (This assumption about equal wages is this paper's most important assumption, and I will defend it in the concluding remarks.) Then it only takes the

slightest turnover cost to make the firm prefer seasoned workers to new entrants.

To attract workers in the first period the firm must offer an expected utility level at least equal to that which can be obtained elsewhere,  $\bar{U}$ :

$$U = u(w_0) + \delta \sum_{s=1}^S u(w_s) \phi_s + (1-\delta) x > \bar{U} \quad (2)$$

$U$  = a representative worker's expected utility

$u$  = ex post worker utility ( $u' > 0$ ,  $u'' < 0$ , implying risk aversion)

$x$  = the utility of a worker who has quit for exogenous reasons

Condition (2) contains the presuppositions that workers and firm have symmetric information as regards the state of the world, that there can be established a money measure of the various benefits and costs accruing to unemployed workers, and that all workers have the same probability of quitting for exogenous reasons between periods.

Like Holmström, I will assume that firms are able to enter into enforceable contracts. Workers, however, are ex post mobile and cannot commit themselves not to leave the firm in period 2. We must therefore have:

$$u(w_s) > u(R_s) \quad \forall s \quad (3)$$

$R_s$  = benefits accruing to workers not employed in the firm in question, measured in money terms. This can e.g. both be unemployed benefits or money earned in a spot market for labour.

We assume that there are no necessary connection between  $R_s$  and  $x$ . Condition (3) also ensures that the firm will be able to attract as many new workers as it wants in period 2. (Both old and new workers have outside opportunities  $R_s$ .)

Above we argued that the slightest turnover costs would make the firm prefer seasoned workers to new entrants when all workers must be paid the same wage. But the possibility remains that the firm can achieve a cut in wage costs by replacing all old workers by new ones. (With the recent incidents at the Murdoch papers in mind, this cannot altogether be dismissed as unrealistic). However, I will block this possibility by assuming that no firm can be founded in period 2 only with new workers. This could be because there is a shortage of entrepreneurial talent, capital goods etc. in the relevant run. In the very long run this is clearly an unsatisfactory assumption. However, there could - also in the long run - be a social convention forbidding a firm with only new workers to pay a different wage than that paid by a firm with both old and new workers; but this is speculative. If old and new workers could be segregated in different firms, this would mean that a ban on seniority wage differentiation to some extent could be evaded.

#### First-order Conditions

The firm's problem now is to maximize (1) subject to (2) and (3). We assume that  $V$  and  $U$  are twice differentiable and strictly concave. We let  $\lambda$  denote the Lagrangian multiplier associated with (2), and define  $\lambda_s \equiv \delta \lambda_0 + \lambda_s^+$ . The first-order conditions are:

$$\theta_0 f'(\lambda_0) - w_0 + \delta \sum_{s=1}^S (\theta_s f'(\lambda_s) - w_s) \phi_s = 0 \quad (4)$$

$$-\lambda_0 + \lambda u'(w_0) = 0 \quad (5)$$

$$\theta_s f'(\lambda_s) - w_s = 0 \quad \forall s \quad (6)$$

$$w_s = \max [w_s^*, R_s] \quad \forall s$$

where  $w_s^*$  is the solution to

$$-(\delta \lambda_0 + \lambda_s^+) + \lambda \delta u'(w_s^*) = 0 \quad \forall s \quad (7)$$



### Interpretation

Our main interest will lie with condition (7), characterizing the wage policy in period 2, as this is the main departure from standard contract theory. This being so, I will start out by briefly commenting on what standard contract theory predicts as regards wage setting.

In the original implicit labour contract literature (the Azariadis-Baily model) it was optimal that workers received a constant wage across states of the world. But such a wage scheme would have to imply that the wage in some states would be lower than the workers' reservation wage. If the workers are ex post mobile and cannot commit themselves to stay on with the firm even when  $w_s < R_s$ , this wage scheme would not be incentive compatible.

If we assume that senior workers are ex post mobile, but allow the firm to pay a wage  $w_s^+$  to new entrants which does not necessarily coincide with that of the older workers, we have a set-up rather similar to that in Holmström (1983). Within such a framework newcomers would be paid  $w_s^+ = R_s$ , and the first-order condition (7) would be replaced by

$$w_s = \max [w_o, R_s] \quad \forall s \quad (8)$$

In states where  $R_s$  is low, workers will receive a flat wage  $w_o$ . In states with a high  $R_s$ , workers must be paid a wage equal to the reservation wage, to prevent them from defecting from the firm. Holmström speaks of this as wages being "downward sticky and upward mobile". This might be a somewhat unfortunate usage of terms: In a multiperiod model, whether wages are sticky or mobile over time depends on how the state in one period compares with the state in the preceding period. If a good state is followed by yet another good state, wages might be both upward and downward flexible. If a bad state is followed by another bad state, wages might be upward and downward immobile. Therefore, I prefer to speak of such a wage scheme as a "safety net" insur-

ance arrangement. Moreover, as the wage in Holmström's model never can be lower than  $R_s$ , this insurance arrangement must be paid for by the workers accepting a lower wage in the first period than they would otherwise have done.

In our setting, however, where senior workers are ex post mobile and seniority wage differentiation is not allowed, the period 2 wage scheme is characterized by (7). Let us first approach this issue intuitively.

Now, the firm knows on the one hand that the cheapest way to buy labour from old workers is to offer these workers insurance against wage fluctuations. On the other hand, the new workers are uninsurable - as they arrive at the scene after the true state of nature is known. This means that the cheapest way of buying labour from this group is to pay them exactly their reservation wage. When wage differentiation is not possible, these two considerations cannot be reconciled. A balance must be struck between them.

Alternatively, it might be instructive to see things from the old workers' viewpoint. If these old workers buy insurance from their firm, their wage will in some states supercede their reservation wage by an "implicit insurance indemnity". But when there is no wage differentiation between old and new workers, also new workers must receive this insurance indemnity. But the new entrants cannot participate in paying for the insurance arrangement, simply because they are not present in the first period. This cools the old workers' desire to buy insurance.

Let us now turn to investigate to what extent these intuitive arguments are supported by the first-order conditions of the problem. Substituting (5) into (7) and rearranging, we obtain:

$$\frac{u'(w_s^*)}{u'(w_o)} = \frac{\delta l_o + l_s^+}{\delta l_o} \quad (8)$$

Remember that  $w_s^*$  is the wage that will be paid as long as the

restriction  $w_s > R_s$  does not bind. We see that as  $\lambda_s^+ \rightarrow 0$ ,  $w_s^+ \rightarrow w_0$ . If no new workers had been hired, we would have had a complete insurance of the workers' wages, as long as the restriction (3) does not block this. This corresponds to Holmström's result. When  $\lambda_s^+ > 0$ , we have that  $w_s^* < w_0$ . And as  $\lambda_s^+$  increases,  $w_s^*$  will be decreasing. It also follows immediately that the larger the number of remaining old workers and the more risk averse they are, the less  $w_s^+$  will be falling in  $\lambda_s^+$ . From the discussion above these results seem rather intuitive.

But condition (8) is not satisfactory as the end result of our analysis. First, (8) only relates one endogenous variable to another. Second, we are primarily interested in studying  $w_s$ , the wage actually paid in a state, not in  $w_s^*$ , what the wage would have been if (3) is not binding.

We now proceed to study how the optimal contract prescribes different values of  $w_s$  and  $\lambda_s$  for different ex post realizations of  $s$ . I will start out with the assumption that  $\theta_s/R_s$  and  $R_s$  are increasing in  $s$ . This means that as a state improves both "inside" productivity  $\theta_s$  and "outside" productivity  $R_s$  increases, but  $\theta_s$  increases relatively more. This is restrictive, and will be relaxed below.

We will look at two cases, in the first  $w_s^* < R_s$ , in the second  $w_s^* > R_s$ . Conditions (6) and (7) describe optimal  $\lambda_s$  and  $w_s$  for a given ex post realization of  $s$ . In case 1, (6) and (7) can be written as:

$$f'(\lambda_s) = \frac{R_s}{\theta_s} \quad \forall s \quad (9)$$

$$w_s = R_s \quad \forall s \quad (10)$$

As both  $\theta_s/R_s$  and  $R_s$  increase in  $s$ , it follows immediately that

both  $w_s$  and  $l_s$  are increasing in  $s$ .

Now look to the somewhat more difficult case 2. In this case, (6) and (7) can be stated as:

$$\theta_s f'(l_s) - w_s^* = 0 \quad \forall s \quad (11)$$

$$-l_s + \lambda \delta u'(w_s^*) = 0 \quad \forall s \quad (12)$$

Totally differentiating, we obtain:

$$\theta_s f''(l_s) \frac{\partial l_s}{\partial s} ds + f'(l_s) \frac{\partial \theta_s}{\partial s} ds - \frac{\partial w_s^*}{\partial s} ds = 0 \quad (13)$$

$$-\frac{\partial l_s}{\partial s} ds + \lambda \delta u''(w_s^*) \frac{\partial w_s^*}{\partial s} ds = 0 \quad (14)$$

Note that  $\lambda = l_0/u'(w_0)$  here is a constant, as  $w_0$  and  $l_0$  of course cannot be influenced by which state of the world occurs ex post.

Rearranging (14) we get:

$$\frac{\partial w_s^*}{\partial s} = \frac{1}{\lambda \delta u''(w_s^*)} \frac{\partial l_s}{\partial s} \quad (15)$$

Substitution into (13) yields:

$$\frac{\partial l_s}{\partial s} = \frac{-f'(l_s) \frac{\partial \theta_s}{\partial s}}{\theta_s f''(l_s) - \frac{1}{\lambda \delta u''(w_s^*)}} \quad (16)$$

We see that if we assume that

$$\theta_s f''(l_s) - \frac{\partial w_s^*}{\partial s} \frac{\partial l_s}{\partial s} < 0 \quad (17)$$

then  $\frac{\partial l_s}{\partial \lambda} > 0$  and  $\frac{\partial w_s^*}{\partial s} < 0$ .

Now, remember that we have assumed that the problem is "well behaved", in that the first-order conditions really do describe a unique and interior global optimum. Let us assume that we are in such an optimum. If we change  $l_s$  and  $w_s^*$  from these optimal values, and still abide by the contract (represented by the first-order conditions), this should not be profitable. But increasing  $l_s$  in such a way both brings about a fall in marginal productivity and in wages. In order for this move not to be profitable, the first effect must dominate the latter. This implies that if the problem is well behaved in the sense described, (17) must hold.

But still, (17) is restrictive. If for instance both the  $f$  and  $u$  functions are comparatively flat, the sign in (17) would be reversed. This would be the case when a new worker would not be very much less productive than the older ones, and that the workers are not very risk averse. It might then be that it would always be optimal, if new workers are taken on at all, not to include any insurance in the wage at all, and to take on many enough workers to equate marginal productivity  $\theta_s f'(l_s)$  and outside opportunity  $R_s$ . However, in my formal analysis I have chosen to focus on the case where the problem is "well behaved".

Let us now bring together the two cases. A diagram of optimal wage setting policy might look as follows:

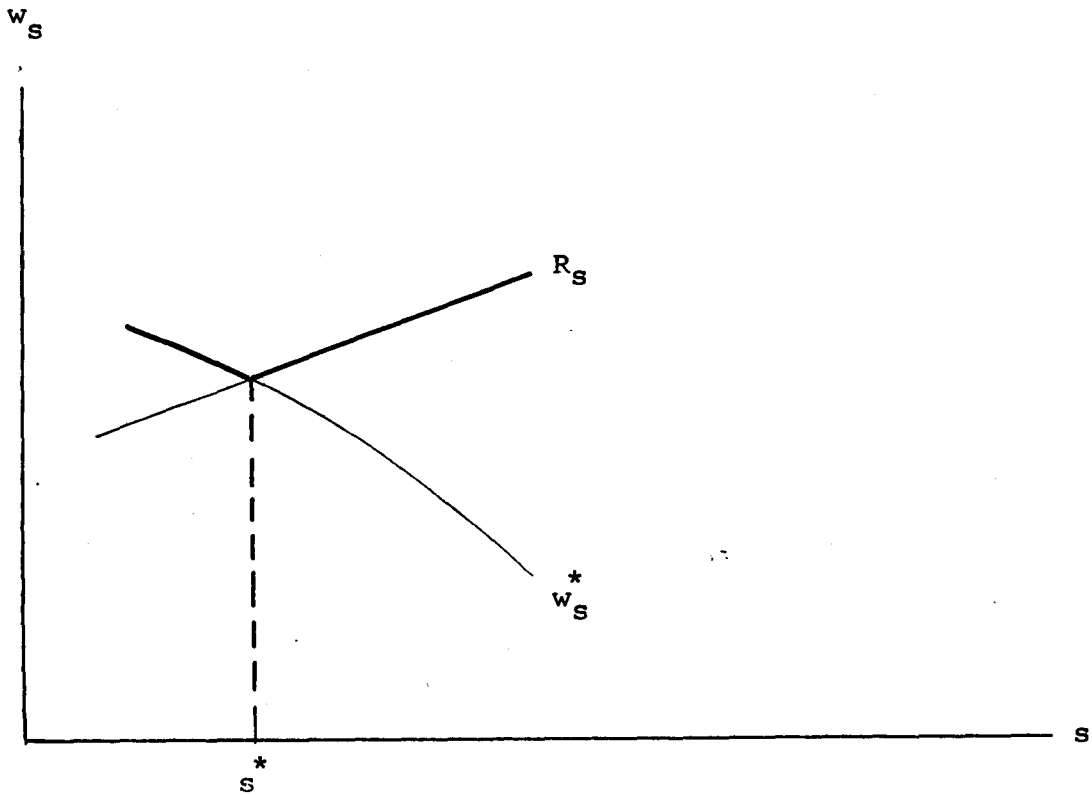


Fig. 1.

The thickly drawn line represents optimal  $w_s$  across different ex post realizations of  $s$ .  $s^*$  denotes the state where (3) starts to bind.

To the left of  $s^*$ , the wage contains an implicit insurance indemnity, in that  $w_s > R_s$ . We see that this indemnity becomes smaller, and that  $w_s$  is falling in  $s$ . From  $s^*$  onwards  $w_s = R_s$ , meaning that no insurance indemnity at all is included in the wage.

The novelty of this wage scheme compared to that found in the Holmström paper, is that when  $w_s > R_s$  does not bind, wages will be falling when the state improves.

We have assumed that states are numbered such that  $\theta_s/R_s$  is increasing in  $s$ . But then there is no reason to expect that the

absolute level of  $R_s$  to be increasing in  $s$ . In the more general case where no restrictions are placed on  $R_s$ , the optimal wage policy might look something like this:

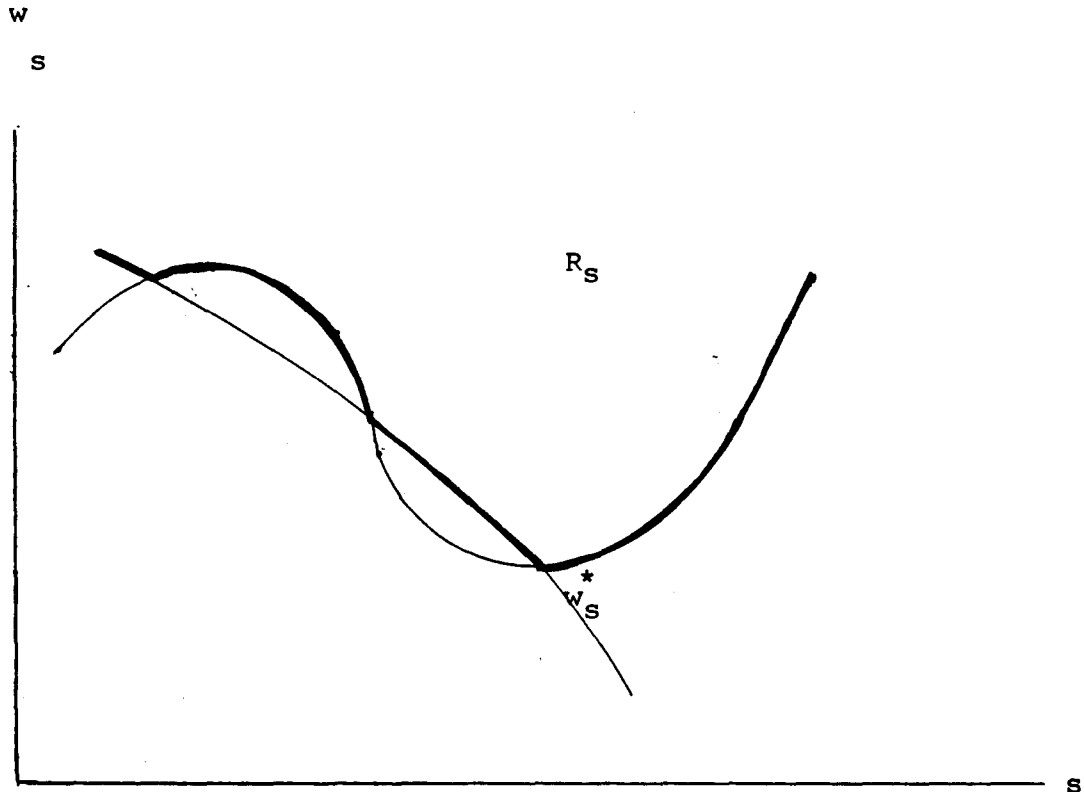


Fig. 2.

Again, the thickly drawn line represents optimal wage policy  $w_s = \max [w_s^*, R_s]$ .

A few words might be in order on what can be expected to happen in deep slumps, when no new workers are being hired: With no new workers distorting the old workers' possibility to buy insurance from their employer, the result will be a constant wage across these very bad states. Employment will be efficient/inefficiently high, depending on whether or not private unemployment insurance is allowed. These results are standard in labour contract theory.

Now we will turn to the case where the old workers are ex post immobile. It proves that in this case a ban on seniority wage differentiation have an even more limiting effect on a firm's possibility to act as an insurer of its employees' wage income.

### 3. Ex Post Immobile Workers.

Are workers best modelled as being ex post mobile or immobile? Presumably the truth lies somewhere in between. I believe, as Holmström, that bans on involuntary servitude block the possibility for workers to enter into contracts which specify that they shall remain in a firm, even when it is not in their best interest to do so. But on the other hand, this does not mean that workers ex post can move at no cost to another firm, as there are various kinds of mobility costs. Modelling workers either as being totally mobile or totally immobile are two extreme cases, and I think it might be fruitful to investigate both of them.

In the previous model condition (3) demanded that  $u(w_s) > u(R_s)$  if the firm should be able to attract new workers and keep its old ones.

Now, if the firm shall be able to attract new workers we must have that

$$(w_s - R_s)l_s^+ > 0 \quad \text{for } l_s^+ > 0 \quad \forall s \quad (18)$$

As we have restricted attention to the states where  $l_s^+ > 0$ , condition (18) is equivalent with (3). This means that the model with the old workers being ex post immobile is identical to the one with mobile old workers. Hence, the solution to the firm's problem, including the optimal second period wage policy, is the same. In other words: With a ban on seniority wage differentiation, the wage scheme will be of the "safety net insurance" type, even when old workers are ex post immobile.



I consider this as the present paper's most interesting result. Usually in labour contract theory, the more immobile workers are after having entered a firm, the larger the scope for writing incentive compatible contracts. Here we see that even when the old workers are completely immobile, the limitation on which contracts can be entered, is exactly the same as with ex post mobile workers. The intuition is straightforward. Even if old workers are immobile, the younger ones are not. In states where alternative opportunities are good, the firm pays a wage equaling  $R_s$  in order to attract new entrants. But due to the ban on seniority wage discrimination, also old workers must receive this wage. We are therefore in a situation as if all workers were mobile, even though this only applies for the new entrants. In turn, this implies that within our framework the scope for the firm acting as an insurer of its workers' labour income is much narrower than in standard contract theory.

Note that there is one slight difference between the model with ex post mobile old workers and that with immobile ones. Restricting attention to the states where  $\lambda_s^+ > 0$  does not mean the same in the two instances. With ex post mobile older workers, (3) binds always. With ex post immobile older workers, (3) binds only when new workers are actually being hired. This means that in the latter case the firm will hesitate more (meaning that the state must be better) before it takes on the first new workers.

#### **4. Seniority Wage Differentiation.**

The major assumption both in this paper and in the insider-outsider literature is that both incumbent workers and new entrants are paid the same wage. I will here discuss this assumption.

There is no denial that in real life we do observe seniority wage differentials. However, to me it seems that there are limits to such differentiation, although these limits perhaps

are narrower in Europe than in the U. S. (where the practice of so-called "two-tier contracts" seems to have spread recently). For some casual empiricism and further discussion of this point, see Drèze (1986).

Of course, this question is hard to settle by casual empiricism. But at least in my own country Norway, seniority wage differentiation seems to be rather small, and apparently not state-contingent. Note that unlimited seniority wage differentiation, as assumed in Holmström's model, not only requires that old and new workers can be paid differently, but that this difference can be made to vary across states of the world. (A word of caution: Even if there seems to be small seniority wage differentials between workers holding the same kind of job, promotion to better paid jobs could be based on seniority rather than ability. But also this is a rather imperfect way of differentiating wages between worker vintages.)

What can be the reason for these limits to seniority wage differentiation? It could be because of social conventions, people's ideas of justice etc. Alternatively, the reason might be that informational problems and transaction costs limit the possibilities of writing enforceable contracts. E.g. if the firm could circumvent the ban on replacing an old worker with a new one - it might be difficult to verify whether a worker is fired for shirking or for some other reason, then it might be in the interest of the old workers to demand that the new workers be paid the same wage as themselves.

My approach has been - in line with the insider-outsider literature - simply to assume that there are no possibilities to differentiate wages between generations of workers, but of course knowing that we then study a polar case.

## 5. Concluding Remarks.

In this paper we have investigated optimal labour contracts between firm and workers when there are two generations of workers and a ban on seniority wage differentiation. We found that this did not only lead to involuntary unemployment among young workers, as is the traditional message of insider-outsider literature, but the possibility of older workers being able to buy insurance from their employer against income fluctuations might also be distorted.

Two main results emerged. First, there might be regions where the wage falls as the state improves. The old workers will be reluctant to let their wage in a state include an insurance indemnity when they have to share this benefit with new workers. When an improved state implies more new workers, this might lead to a falling wage.

Second, we find that the scope for including insurance elements in the workers' wage is as narrow when the old workers are ex post immobile, as if they had been ex post mobile. Even when the old workers are immobile, the younger ones are not. Therefore, in a good state a wage equalling outside opportunities must be paid to attract new entrants - and due to the ban on seniority wage discrimination, also old workers must receive this wage. When new workers are hired, we are therefore in a situation as if all workers were mobile. In turn, the insurance arrangement between the old workers and the firm will therefore have the same form as if the old workers were ex post mobile, even though they are not. We get what I here have denoted a "safety net" insurance arrangement. This means that the scope for the firm acting as an insurer of its workers' labour income is considerably narrower than in standard contract theory.

Especially the first of these results, namely that wages fall as the state gets better, seems unrealistic. This might mean that the basic premise of this paper, that there are limits to seniority wage differentiation, is not a good assumption. But it

could also imply that some of the assumptions that this paper shares with the whole body of contract theory are inappropriate. I believe that contract theory often predicts too complex contracts compared with the real world. It is my opinion that problem solving costs are not dealt with properly in contract theory, and relatedly, that we do not understand well enough issues as commitment and renegotiation. (For some work along these lines, see Hart and Moore (1985) and Oswald (1985).) Perhaps these shortcomings have been clouded by early implicit contract theory predicting rather simple contracts with a constant wage across states. Here we see that rather small changes of assumptions from these early labour contract models bring about a much more complex contract structure. One possible interpretation of the present analysis is therefore that it highlights some shortcomings of the original labour contract literature. However, if labour contract theory were brought closer to reality when it comes to problem solving costs, commitment, renegotiation etc., I would still expect the clash of interest between different worker vintages to influence the structure of contracts. Exactly how is of course difficult to tell.

By way of conclusion I would like to point out the significance of one still uncommented, major assumption, namely that there are only two periods and two generations. This specification seems appropriate when the risk facing the firm is a major, one-shot risk (will the new production technique work, will the new product find a market?). If we are studying recurring risks (will productivity or product price go up or down this year?), a multi-period model with overlapping generations of workers would seem a more natural framework. In such models a given generation would not only have conflicting interests with the generation less senior to it, but perhaps also with the even more senior generation. In the context of our two models: Here the "old" generation wants to post a bond in the first period (implicitly, by accepting lower wages). This, however, might collide with the "even older" generation's preferred wage profile in this period. A closer study of such overlapping generations models I believe is an interesting topic for future research. However, one objection against such overlapping generations models would be that

our assumption that it is not possible to form a firm only with new workers seems very taxing in this long run.

Notes

- 1) It has been pointed out to me (by Andrew Oswald and John Moore) that Black and Bulkley (1985) deal with some of the same issues as here. There are substantial differences between the papers; but my result (to be presented) that wages might be falling as the state gets better has a close relative in Black and Bulkley's paper. Perhaps the most important difference between Black and Bulkley's paper and my own is that they employ a single period model, whereas I use a two period one. As the reader will discover, prepayment of insurance will here play an essential role in the optimal contract, even when the old workers are ex post immobile. This is not captured in a one period setting. Also, a recently published article by Horn and Svensson (1986) deals with the integration of contract theory and trade union theory, with emphasis on risk-shifting. However, this article does not focus on the insider- outsider dilemma.
  
- 2) The term "implicit" is not clearly defined in this work, and seems to mean only "unwritten". Recently it has become common to think of an implicit contract as one which cannot be enforced by a third party. In contrast, in early "implicit" labour contract theory contracts are assumed to be fully enforceable.

## References

- Azariadis, C. (1975) "Implicit Contracts and Underemployment Equilibria", Journal of Political Economy 83, 1182-1202.
- Baily, M. N. (1974), "Wages and Employment under Uncertain Demand", Review of Economic Studies 41, 37-50.
- Black, J. M. and Bulkley, G. (1985), "Do Trade Unions Reduce Job Opportunities of Non-Members?", Discussion Paper, University of Exeter.
- Blanchard, O. J. and L. H. Summers (1987), "Hysteresis in Unemployment", European Economic Review 31, 288-295.
- Carruth, A. A. and A. J. Oswald (1987), "On Union Preferences and Labour Market Models: Insiders and Outsiders", Economic Journal 97, 416-340.
- Drèze, J. H. (1986), "Work Sharing: Some Theory and Recent European Experience", Economic Policy, 561-601.
- Eaton, B.C. and W.D. White (1982), "Agent Compensation and the Limits of Bonding" Economic Inquiry 20, 330-343.
- Geneakopulos, J. and T. Ito (1981), "On Implicit Contracts and Involuntary Unemployment". Discussion Paper No. 81-155, University of Minnesota.
- Hart, O. and J. Moore (1985), "Incomplete Contracts and Renegotiation". Discussion Paper 85/115, ST/ICERD, London School of Economics.
- Holmström, B. (1983), "Equilibrium Long-term Labor Contracts", Quarterly Journal of Economics 98, Supplement, 23-54.

Horn, H. and L.E.O. Svensson (1986), "Trade Unions and Optimal Labour Contracts", Economic Journal 96, 323-341.

Ireland, N. and P. Law (1982), The Economics of Labour-Managed Enterprises, Croom Helm, London

Ito, T. (1984), "Implicit Contracts with Costly Search: The Incentive Constraint in Severance Payments". Discussion Paper No. 202, University of Minnesota.

Kahn, C. M. (1985), "Optimal Severance Pay with Incomplete Information", Journal of Political Economy 93, 435-451.

Lindbeck, A. and D. Snower (1984a), "Involuntary Unemployment as an Insider-Outsider Dilemma", Seminar Paper No. 282, Institute for International Economic Studies, University of Stockholm.

Lindbeck, A. and D. Snower (1984b), "Labor Turnover, Insider Morale and Involuntary Unemployment", Seminar Paper No. 310, Institute for International Economic Studies, University of Stockholm.

Lindbeck, A. and D. Snower (1985a), "Cooperation, Harassment, and Involuntary Unemployment", Seminar Paper No. 321, Institute for International Economic Studies, University of Stockholm.

Lindbeck, A. and D. Snower (1985b), "Explanations of Unemployment", Oxford Review of Economic Policy 1, No. 2, 34-59.

Lindbeck, A. and D. Snower (1986), "Wage Setting, Unemployment and Insider-Outsider Relations", American Economic Review 76, Papers & Proceedings, 235-239.

Moore, J. (1985), "Optimal Labor Contracts When Workers Have a Variety of Privately Observed Reservation Wages", Review of Economic Studies 52, 37-67.



- Oswald, A.J. (1984), "On Union Preferences and Labour Market Models: Neglected Corners". Seminar Paper No. 296, Institute for International Economic Studies, University of Stockholm.
- Oswald, A.J. (1985) "A Theory of Non-Contingent Wage Contracts", Mimeo, Centre for Labour Economics, London School of Economics
- Oswald, A.J. (1986), "Unemployment Insurance and Labor Contracts Under Asymmetric information: Theory and Facts", American Economic Review 76, 365-377.
- Oswald, A.J. and P. J. Turnbull (1985), "Pay and Employment Determination in Britain: What Are Labour 'Contracts' Really Like?", Oxford Review of Economic Policy 1, 80-97.
- Sampson, A.A. (1985), "Unionized Contracts With Fixed Wage Rates and State-Contingent Employment". Mimeo, University of Sheffield.
- Shaked, A. and J. Sutton (1984), "Involuntary Unemployment as a Perfect Equilibrium in a Bargaining Model", Econometrica 52, 1351-1364.
- Solow, R.M. (1985), "Insiders and Outsiders in Wage Determination", Scandinavian Journal of Economics 87,
- Tracy, J. and G. Woglom (1984), "Seniority Rules and the Gains from Union Organization", Working Paper No. 690, Centre For Labour Economics, London School of Economics.

**Chapter 3:**

**WORKSHARING VERSUS LAYOFFS IN LABOR CONTRACTS**

October 1985

Revised July 1986

A slightly different version of this chapter is issued as Discussion paper 1/1986, Norwegian School of Economics and Business Administration, January 1986.

**WORKSHARING VERSUS LAYOFFS IN LABOR CONTRACTS\***

**ABSTRACT**

Labor contract theory, like much other work in labor economics, focuses often only on the determination of the employment level. Whether worksharing or layoffs are used to effectuate fluctuation in employment, is a question mostly left unanswered. Exceptions are for instance Mortensen (1978) and Rosen (1985), who point out that indivisibilities of some sort are necessary for layoffs to be strictly preferred. This paper examines the implication of one such indivisibility: it is not possible to have more than one job at the same time. If we assume that the monetary value of leisure shows positive but diminishing marginal returns with respect to leisure, the following result can be obtained: When the "net wage" in alternative employment is below a certain value, worksharing will always be chosen. When it is above another, higher critical value, layoffs will always be chosen. Between these critical values it is possible that a combination of both instruments will be used to effectuate a given reduction of employment.

## WORKSHARING VERSUS LAYOFFS IN LABOR CONTRACTS

### 1. Introduction.

It has been common practice in economic reasoning about labor markets to focus on the determinants of the employment level, and to leave aside the question of how unemployment is distributed among workers. This holds true also for standard contract theory, as in the original treatments by Azariadis (1975), Baily (1974) and Gordon (1974): A result from this theory is that a firm might want to reduce its employment level in adverse states of nature, even when the firm acts as an insurer of its workers. However, labor contract theory by itself has little to say about whether worksharing or layoffs will be used to bring about such a reduction, i.e., whether a reduction in employment will be distributed equally among all workers or carried only by some, totally unemployed workers. Note however that this is a weakness labor contract theory shares with much other labor market theory. Within a contract framework Mortensen (1978) and Rosen (1985) among others have pointed out that nonconvexities of some sort - for instance in the production function or in the utility function - are necessary for layoffs to be strictly preferred.

Here I present an example of a nonconvexity different from Mortensen's examples: The key assumption is that only totally unemployed workers can find alternative employment in other industries, i.e., it is not possible to work in two firms at the same time. I assume that the marginal monetary value of leisure is positive but diminishing. Now the following result can be obtained: I show that when the "net wage" in alternative employment is below a certain critical value, worksharing will always be chosen. When it is above an other, higher critical value, layoffs will always be chosen. In the mid-zone between these critical values it is possible that a combination of both instruments will be used to effectuate a given reduction of employment <sup>1)</sup>.

The remainder of this paper is organized as follows. In section 2 I present the basic model framework, and also derive a complete insurance result, quite similar to the insurance results of standard contract theory. In section 3 I use this result to reformulate the firm's optimization problem. Given that the firm shall provide its workers full insurance against variations in utility, and taking total employment level as given, what is the optimal use of reduced hours and layoffs? Section 4 summarizes the discussion in section 3. Section 5 contains some final remarks on some of the restrictive assumptions of the preceding analysis.

## 2. Preliminaries.

Consider a firm that has a pool of  $\bar{N}$  identical workers. These workers are non-unionized. Production takes place in an uncertain environment: In period one a contract between firm and workers is entered, then in period two one of a finite set of states of the world occurs. A state  $\theta$  is characterized by a productivity coefficient  $s_\theta$ , which enters the production function in a multiplicatively separable way. The probability of a state  $\theta$  occurring is  $\pi_\theta$ .

Each worker supplies inelastically one unit of labor. The contract between the firm and the workers specifies  $w_\theta$ ,  $b_\theta$ ,  $n_\theta$  and  $N_\theta$ :  $w_\theta$  denotes the wage level in state  $\theta$  for retained workers,  $b_\theta$  is a severance pay to laid-off workers,<sup>2)</sup>  $n_\theta$  is the work intensity of a retained worker (the fraction of his unit of supplied labor actually purchased by the firm), and  $N_\theta$  is the number of retained workers.

We assume that the firm maximizes expected profit,  $V$ . This implies that the firm is risk-neutral.

$$V = \sum_{\theta} \pi_{\theta} [s_{\theta} f(n_{\theta} N_{\theta}) - w_{\theta} n_{\theta} N_{\theta} - b_{\theta} (\bar{N} - N_{\theta})] \quad (1)$$

Note that for simplicity we have abstracted from discounting. The expression  $s_{\theta} f(n_{\theta} N_{\theta})$  is a multiplicatively separable production function, expressed in value terms. We assume  $f' > 0$  and  $f'' < 0$ .

We assume that the firm must offer each of the workers in its pool at least a market determined expected utility level  $\bar{u}$  :

$$u \geq \bar{u} \quad (2)$$

Workers are assumed to maximize expected utility. The ex post utility function of a representative worker will be written as

$$u = u(M + r(\lambda)). \quad (3)$$

$u$  is ex post utility.  $M$  is money income, whereas  $r$  is the pecuniary value of leisure,  $\lambda$ . Leisure is a quantity between 0 and 1, denoting the fraction of a period which a worker does not work. As is well known, it presupposes restrictive assumptions to assign a pecuniary value to leisure<sup>3)4)</sup>. We will assume that marginal utility is positive but diminishing, implying risk aversion.

I will also assume that  $r' > 0$  and  $r'' < 0$ , and that  $r(0) = 0$ . I.e., the marginal money value of leisure is positive and diminishing. I hold this to be a realistic assumption, but admittedly it is a special case. Below it will turn out that the functional form of  $r$  is crucial for the results.

The ex post utility of a worker retained in period 2 can now be written:

$$u = u(w_{\theta} n_{\theta} + r(1-n_{\theta})) \quad \text{for all } \theta \quad (4)$$

A laid-off worker can choose between leisure or finding alternative employment according to what yields the highest utility. I will here assume that a worker always works full time in the alternative job, but this assumption can easily be relaxed. Let

$\bar{w}_\theta$  denote the "net wage" in the alternative job. We might for instance imagine that  $\bar{w}_\theta$  represents the alternative wage less mobility costs. We will further discuss the interpretation of  $\bar{w}_\theta$  towards the end of this paper. The ex post utility of a laid-off worker can now be stated as:

$$\max [ u (b_\theta + r(1)), u (\bar{w}_\theta + b_\theta) ] \quad \text{for all } \theta \quad (5)$$

By using (4) and (5), we can now restate (2) as:

$$\begin{aligned} \bar{U} \leq \sum_{\theta} \pi_{\theta} \left\{ \left[ \frac{N_{\theta}}{\bar{N}} u (w_{\theta} n_{\theta} + r(1 - n_{\theta})) \right] \right. \\ \left. + \left( 1 - \frac{N_{\theta}}{\bar{N}} \right) \max [ u (b_{\theta} + r(1)), u (\bar{w}_{\theta} + b_{\theta}) ] \right\} \quad (6) \end{aligned}$$

This formulation implies that workers are picked out to be laid off in an arbitrary fashion.

When we solve the firm's optimization problem, it will be convenient to distinguish between two cases. We define:

$$A \equiv \bar{w}_{\theta} - r(1) \quad (7)$$

Now, let case 1 be given by  $A < 0$ . By inspection of (5) we see that this implies that a laid-off worker will prefer leisure to finding a new job. Case 2 is defined by  $A > 0$ . In this case new employment will be chosen.

The firm's maximization problem in the two cases, respectively, will be to maximize (1) given (6), and subject to  $n_{\theta} \leq 1$ ,  $N \leq \bar{N}$  and that the choice variables be non-negative.

There are potential problems with applying standard techniques here as neither  $V$  nor  $U$  are necessarily concave. This is a common problem to many labor contract models, but rarely of any

real significance (see Holmström (1983), Appendix A). I will here apply a step-wise solution procedure: First I assume that  $n_\theta$  and  $N_\theta$  are exogenous variables ( $\neq 0$ ), and find then that the risk-neutral firm will provide its workers any quantity of fair insurance. (This problem is well behaved.) Then, in Section 3, I use this result to restate the firm's problem in a more "manageable" form.

As mentioned, let us for the time being consider  $n_\theta$  and  $N_\theta$  as exogenous variables. We will start off by considering case 1, i.e.,  $A < 0$ . We denote by  $\lambda$  the Lagrangian multiplier associated with the constraint (6). Obviously, condition (6) will always hold with equality. The first-order conditions with respect to  $w_\theta$  and  $b_\theta$  will be:

$$-\pi_\theta n_\theta N_\theta - \lambda \pi_\theta \left( \frac{N_\theta}{N} n_\theta u'(w_\theta n_\theta + r(1-n_\theta)) \right) = 0$$

for all  $\theta$  (8)

$$-\pi_\theta (\bar{N} - N_\theta) - \lambda \pi_\theta \left( \left(1 - \frac{N_\theta}{N}\right) u'(b_\theta + r(1)) \right) = 0$$

for all  $\theta$  (9)

As  $u$  is a strictly concave function of only one variable, we know that equal marginal utilities imply equal utility levels. In turn, equal utility levels imply equal money or money equivalent income. Considering this, conditions (8) and (9) imply:

$$w_\theta n_\theta + r(1-n_\theta) = b_\theta + r(1) = K \quad \text{for all } \theta \quad (10)$$

For case 2 a corresponding condition can be found:

$$w_\theta n_\theta + r(1-n_\theta) = \bar{w}_\theta + b_\theta = L \quad \text{for all } \theta \quad (11)$$



The state-independent constants  $K$  and  $L$  can be seen from (6) to be equal, and given by:

$$u(K) = u(L) = \frac{\bar{U}}{\sum_{\theta} \pi_{\theta}} .$$

Conditions (10) and (11) describe complete insurance in the sense that a worker's ex post utility will be constant, independent of the state of the world and of the values of  $n_{\theta}$  and  $N_{\theta}$ . As the complete insurance conditions are derived assuming  $n_{\theta}$  and  $N_{\theta}$  to be exogenous, they hold for any  $n_{\theta}$  and any  $N_{\theta}$ . In the remainder of this paper,  $n_{\theta}$  and  $N_{\theta}$  will be considered to be choice variables.

### 3. Worksharing versus Layoffs.

We are now ready to turn to our main issue: When will it be optimal to use worksharing or alternatively layoffs in order to obtain a desired reduction of the employment level?

I will now use the complete insurance result of section 2 to reformulate the firm's optimization problem. To simplify, I will take the desired total employment level in a state as given, and only investigate how a cost minimizing firm will choose between worksharing and layoffs to effectuate a given reduction in employment. (Naturally, if we compare the profit yielded by different levels of total employment, given the complete insurance condition and a cost minimizing choice between worksharing and layoffs, the optimal total employment level can be found.)

Conditions (10) and (11) tell us that a worker will not care about whether a reduction in employment is effectuated by the use of worksharing or layoffs: He knows that his utility in one basic time unit will be held constant anyway. He therefore leaves this decision with the firm, as long as the complete insurance condition is honored. Further, by summarizing conditions (10) and (11) for all workers, we see that the firm's total

outlay on wages and severance pay plus the value of spare time to the workers shall equal a constant. (In this paper "spare time" will mean the time a worker does not work in his primary firm, whereas "leisure" denotes the time he does not work either in the primary firm nor in any alternative job. The value of the workers' spare time is therefore the monetary value of leisure for the retained workers plus income net of mobility costs in alternative employment for the laid-off workers.) Obviously, when total labour costs plus the value of the workers' spare time shall equal a constant, to maximize the value of spare time implies to minimize labour costs. And as we have taken the total employment level as given, minimization of labour costs imply that profits are maximized. Hence, the firm's problem can be stated as one of distributing a given amount of additional spare time so as to maximize the value of this spare time.

The value of spare time can be expressed as:

$$VST \equiv \sum_i \{r(1-n_{\theta i}) + \bar{w}_{\theta i}\} \quad (12)$$

where subscript  $i$  refers to an individual worker  $i$ .  $n_{\theta i} = n_{\theta}$  for retained workers,  $n_{\theta i} = 1$  for laid-off workers who choose to find a new job, and  $n_{\theta i} = 0$  for laid-off workers who choose leisure rather than employment. Note that here we interpret  $\bar{w}_{\theta i}$  as the wage actually (as opposed to potentially) earned in alternative employment. That means that  $\bar{w}_{\theta i} = 0$  for all workers but for those who have chosen  $n_{\theta i} = 1$ .

We want to consider changes in  $n_{\theta}$  and  $N_{\theta}$  which bring about an equal reduction in employment from an arbitrary initial employment level  $\bar{n}_{\theta} \bar{N}_{\theta}$ . If we want to reduce employment by, say,  $dH$ , using only  $dN_{\theta}$  or  $dn_{\theta}$  as our policy variable, we must have

$$dH = \bar{n}_{\theta} dN_{\theta} = \bar{N}_{\theta} dn_{\theta} \quad \text{for all } \theta \quad (13)$$

Now, the problem is so simplified that it can be solved by a rather intuitive approach. We simply compare the value of an incremental increase in spare time if worksharing or a layoff is used, from any given starting point. Finite changes in the employment level will be viewed as composed of many such incremental changes.

Let us start by considering case 1, the case where the value of alternative employment was so low that a laid-off worker would choose to spend his spare time at leisure activities. Suppose the firm wants to bring about a small reduction in the employment level. Considering (12) and (13), the relevant comparison will be:

$$B_1 \equiv [r'(1-\bar{n}_\theta) \cdot \bar{n}_\theta] - [r(1) - r(1-\bar{n}_\theta)] \stackrel{?}{>} 0 \quad \text{for all } \theta \quad (14)$$

The first bracketed term is the value of extra spare time if worksharing is chosen, whereas the second bracketed term represents the value of additional spare time in the case of a layoff. We have multiplied by  $\bar{n}_\theta$  in the first term to make the employment reduction in the two cases equal, i.e., in order to satisfy condition (13). Obviously  $B_1 > 0$  implies that worksharing will be chosen, while  $B_1 < 0$  will lead to layoffs.

As  $r' > 0$  and  $r'' < 0$ , we know:

$$r'(1-\bar{n}_\theta) > \frac{1}{\bar{n}_\theta} \int_{1-\bar{n}_\theta}^1 r'(\lambda) d\lambda = \frac{1}{\bar{n}_\theta} (r(1) - r(1-\bar{n}_\theta))$$

for any  $\bar{n}_\theta$ , for all  $\theta$  (15)

From condition (15) we know that  $B_1$  always will be positive, accordingly worksharing will always be chosen in case 1. The intuition is rather simple: With the present formulation all workers have identical utility functions. Further, this common utility function implies that the marginal monetary value of

leisure decreases as one gets more leisure. In case 1 we know that all spare time will be used on leisure activities. It is then rather obvious that the value of a given spare time is maximized if it is distributed equally among the workers, rather than the unemployment being carried by only a few. Note that - by the same line of reasoning as above - if  $r$  is linear, the firm will be indifferent between worksharing and layoffs, and if  $r$  is strictly convex, layoffs will be the preferred choice.

We now turn to case 2, defined by  $A > 0$ . This means a laid-off worker will take up alternative employment. Once again, a cost minimizing firm will choose to maximize the value of the workers' spare time, as that minimizes total labor costs. When a small reduction in employment is warranted, the relevant comparison is:

$$B_2 \equiv [r'(1-\bar{n}_\theta) \cdot \bar{n}_\theta] - [\bar{w}_\theta - r(1-\bar{n}_\theta)] \gtrless 0 \quad \text{for all } \theta \quad (16)$$

The first bracketed term is, as before, the value of extra spare time in the case of worksharing. The second bracketed term expresses the value of additional spare time when a layoff is used.

A priori we cannot tell which one of the bracketed terms in (16) is the larger. If  $B_2 > 0$  for any  $\bar{n}_\theta$  worksharing will always be chosen, if  $B_2 < 0$  for any  $\bar{n}_\theta$  layoffs will always be the preferred instrument. But as both bracketed terms depend on  $\bar{n}_\theta$ , it might be that  $B_2 > 0$  for some values of  $\bar{n}_\theta$  and  $B_2 < 0$  for other values.

Let us differentiate  $B_2$  with respect to  $\bar{n}_\theta$ :

$$\frac{\partial B_2}{\partial \bar{n}_\theta} = -r''(1-\bar{n}_\theta) \cdot \bar{n}_\theta + r'(1-\bar{n}_\theta) - r'(1-\bar{n}_\theta) = -r''(1-\bar{n}_\theta)\bar{n}_\theta > 0 \quad \text{for all } \theta \quad (17)$$

This implies that if  $B_2$  has different sign for different values of  $\bar{n}_0$ , it will be positive for high values of  $\bar{n}_0$  and negative for low values.  $B_2$  will never switch sign more than once.

If a finite change in employment is desired, we can imagine that this finite change is decomposed into a sum of incremental changes. For each incremental change the firm must use condition (16) to determine whether worksharing or a layoff shall be used. As  $B_2$  might change sign for different values of  $\bar{n}_0$  a finite change of employment could be effectuated by the use of both instruments. Say that we start from the full employment level, and that we have the case where  $B_2$  changes sign. We then know from (17) that a relatively small decrease in employment will be brought about by worksharing, whereas for a relatively large decrease in employment both instruments will be taken into use. The intuition is simple: At a high employment level the value of leisure is high. It might then pay to choose worksharing in order to give all workers a little more leisure, rather than having one worker move to alternative employment. At lower employment levels the value of leisure is lower. Then it might be that the wage net of mobility cost in alternative employment for a laid-off worker is higher than the monetary value of a little more leisure to all workers. Consequently, a layoff would be the optimal choice. A finite reduction in employment might include both of these situations: We should then observe both shorter working time and layoffs simultaneously.

We have seen that in case 2 the firm will use layoffs to reduce the employment level if the alternative employment opportunity is of a "high enough" quality, i.e., when  $B_2 < 0$ . The idea is roughly that a big chunk of spare time sometimes can be better utilized than many small periods of time off. One of Mortensen's (1978) examples of indivisibility leading to layoffs being chosen, was when the workers' utility function was convex in leisure. Then one long spell of leisure is more valuable than the

same amount of leisure divided on many short spells. However, I find it hard to believe that the frequent use of layoffs is due to the workers having a preference for going on one long fishing trip rather than many short ones. In the present framework we have retained the idea that a long spell of spare time can be more valuable than many short ones, but this is not due to the utility function being convex in leisure, but rather that people cannot work in two firms at the same time. When a totally unemployed person can find new employment, whereas workers working reduced hours must use their spare time on leisure activities, this might be a rationale for the use of layoffs, provided that the net wage in alternative employment is high enough.

#### 4. Critical Values of the "Net Wage" in Alternative Employment.

Above we have studied case 1 and case 2 separately. Let us now see these two cases in conjunction. Both the quantity A, which delineates case 1 and case 2, and  $B_2$ , which determines whether worksharing or layoffs will be used in case 2, depend on  $\bar{w}_\theta$ . It might be instructive to compare the values of  $\bar{w}_\theta$  which make A and  $B_2$  exactly equal zero, respectively. Let us denote by  $\bar{w}_\theta^*$  the value of  $\bar{w}_\theta$  which makes A equal zero, whereas  $\bar{w}_\theta^{**}(\bar{n}_\theta)$  denotes the value of  $\bar{w}_\theta$  which makes  $B_2$  equal to zero for a given  $\bar{n}_\theta$ . Further,  $\bar{w}_{\theta L}^{**}$  and  $\bar{w}_{\theta H}^{**}$  denote the lowest and the highest value of  $\bar{w}_\theta^{**}(\bar{n}_\theta)$ , respectively, when we allow  $\bar{n}_\theta$  to vary.

We have:

$$\bar{w}_\theta^* = r(1) \quad (18)$$

$$\bar{w}_\theta^{**}(\bar{n}_\theta) = r'(1-n_\theta) \cdot \bar{n}_\theta + r(1-\bar{n}_\theta) \quad (19)$$

It is easy to show that  $\bar{w}_\theta^{**}$  is increasing in  $\bar{n}_\theta$ . This implies

that  $\bar{w}_\theta^{**} (0) = \bar{w}_{\theta L}^{**}$  and  $\bar{w}_\theta^{**} (1) = \bar{w}_{\theta H}^{**}$ . From (19),  $\bar{w}_\theta^{**} (0) = r(1)$ , i.e.

$$\bar{w}_{\theta L}^{**} = r(1) = \bar{w}_\theta^* \quad (20)$$

Hence, we have

$$\bar{w}_\theta^* = \bar{w}_{\theta L}^{**} < \bar{w}_{\theta H}^{**} \quad (21)$$

Condition (21) can be viewed as summarizing the findings in Sections 3 and 4:

$$\bar{w}_\theta < \bar{w}_\theta^* :$$

This is case 1, in which work-sharing will be chosen, since  $B_1$  always is positive.

$$\bar{w}_\theta^* = \bar{w}_{\theta L}^{**} < \bar{w}_\theta < \bar{w}_{\theta H}^{**} :$$

As  $\bar{w}_\theta > \bar{w}_\theta^*$ , we are in case 2, i.e., a laid-off worker will choose to find a new job. When  $\bar{w}_\theta$  lies in this area, layoffs can be taken into use in combination with worksharing to bring about a reduction in employment. We have seen above that this happens when the desired employment reduction is relatively large. The marginal value of leisure is decreasing in leisure. When the marginal value of leisure is low enough relative to  $\bar{w}_\theta$ , it will pay to take layoffs into use. For higher  $\bar{w}_\theta$ 's, the size of a finite reduction necessary for layoffs to be taken into use, is reduced.

$$\bar{w}_\theta > \bar{w}_{\theta H}^{**}$$

Now the value of  $\bar{w}_\theta$  has become so large that any reduction in employment will be brought about by layoffs.

We have said that  $\bar{w}_\theta$  denotes "net wage" in alternative employment, i.e., the gross alternative wage, say,  $\hat{w}_\theta$ , less mobility costs,  $C_\theta$ . Let us now, rather tentatively, try to introduce "duration" of a state in the analysis. Assume that a state  $\theta$  is characterized not only by a productivity coefficient  $s_\theta$ , but also by how many time units  $m_\theta$  it lasts. In this case, if there is some fixed cost element in the mobility cost, the per period mobility cost will be a declining function of  $m_\theta$ . If all mobility costs were independent of the length of the state, we would have:

$$\bar{w}_\theta = \hat{w}_\theta - \frac{C_\theta}{m_\theta} \quad (22)$$

We see immediately that a high  $m_\theta$  increases  $\bar{w}_\theta$ , as it in a sense will undermine the importance of mobility costs. However, when  $m_\theta$  has reached a certain value, we will have  $\bar{w}_\theta \approx \hat{w}_\theta$ , and further increases in  $m_\theta$  will not influence  $\bar{w}_\theta$ , and therefore not the choice between worksharing and layoffs either.

I think that these last remarks are fairly obvious, but such issues are rarely focused on, as duration of states, or more generally, correlation of productivity coefficients across periods, never seems to have been modelled explicitly in a labor contract framework<sup>5)</sup>.

## 5. Final Remarks.

Here I will comment on some of the simplifying assumptions underlying the present model.



One such restrictive assumption is that the choice between work-sharing and layoffs can be made contingent on the realization of  $\bar{w}_0$ . If off-the-job search is important, a worker's alternative employment opportunities and his mobility and search cost cannot be known before after search has been undertaken. If we assume that  $\bar{w}_0$ , at the time when an employment reduction is considered, is known only up to a probability distribution, it could be true that the workers' expected utility from alternative employment is high enough to justify layoffs. However, those workers actually laid off could be unlucky and for instance only find jobs with wages below their reservation wage. But this cannot by itself explain involuntary unemployment: The firm could insure the workers against this uncertainty, for instance by letting unsuccessful searchers return to the firm. However, several authors, (e.g. Arnott, Hosios and Stiglitz (1983), Geanakoplos and Ito (1981), Ito (1984), Kahn (1985), Moore (1985) and Strand (1985)) have focused on adverse selection and moral hazard problems of such insurance when search intensity and search outcome are unobservable by the firm. First when we incorporate incomplete insurance against uncertainty about  $\bar{w}_0$ , we have a model which allows both for layoffs and involuntary unemployment - as opposed to the present model, where all laid-off workers are guaranteed to find alternative employment.

Another important aspect of the present model is that all other non-convexities except the one focused on here, are assumed away. It is obvious that there also could be important nonconvexities in the production technology. Further, there could be nonconvexities in the tax system. In some countries the rules for public unemployment insurance are such that it is easier to "extract" public unemployment benefits with layoffs than with worksharing. (For instance, this is the case in U. S. A. and Canada.) Also, as pointed out by FitzRoy and Hart (1985), payroll taxes are in some countries more or less a fixed cost per worker, obviously this creates a bias towards layoffs.

Also, the model in the previous sections assumes that layoffs take place in a completely random manner. This is not necessarily so, and there could be some scope for the firms managing to lay off the relatively less productive among equally paid workers. Layoffs therefore could yield a higher productivity per remaining unit of labor than worksharing. (Johansen (1982a), (1982b), Thurow (1981).)

Still another explanation of layoffs left out in the present framework, is that unions could have a decisive influence on whether and how layoffs are to be used: If seniority rules are imposed on the firm, the median voter of the union might not be affected by layoffs, whereas he would be in the case of worksharing.

### Notes

- \* I thank Jan Erik Askildsen, Kåre Petter Hagen, Andrew Oswald, Erling Steigum, Bjørn Sandvik and Jon Strand for useful comments. The usual disclaimer applies.
- 1) Baily (1977) obtains results somewhat similar to the present ones when discussing the choice between worksharing and layoffs. However, he used a framework with risk-neutral workers.
  - 2) It is an old discussion in labor contract theory whether one should or should not include private severance pay among the firm's policy variables. I have rather arbitrarily chosen to do so. This assumption simplifies the results somewhat.
  - 3) In the labor contract literature utility functions of this type have been used by Hall and Lillian (1979), Chan and Ioannides (1982), and Grossman and Hart (1981). More general specifications can be found for instance in Chari (1983), Green and Kahn (1983), and Brown and Wolfstetter (1985). My personal belief is that the use of more general utility functions has confused the debate more than it has clarified it - see the comments on this issue in Holmström and Hart (1985). But, of course, the reader must bear in mind that our results would be somewhat more complicated with a more general utility function.
  - 4) The joint assumptions that the firm has the possibility to insure its workers also against layoff risk (i.e., that  $b_0$  is one of the choice variables) and that the value of leisure can be expressed in money terms, gives our insurance result (10) and (11) a particularly neat form. Different assumptions on these accounts would of course alter our results, but not in a way I consider substantial.

- 5) This simple modelling of "duration" of booms and slumps has its shortcomings. I think the most serious one is the tacit assumption that the length of the second period (the period where production takes place) is the same both in the "primary" firm and in alternative employment.

References

- Akerlof, G. A. (1980), "A Theory of Social Custom, of Which Unemployment May Be One Consequence", Quarterly Journal of Economics 95, 749-775.
- Arnott, R., A. Hosios and J. Stiglitz (1983), "Implicit Contracts, Labor Mobility and Unemployment", Queen's University, Discussion Paper No. 543.
- Azariadis, C. (1975) "Implicit Contracts and Underemployment Equilibria", Journal of Political Economy 83, 1182-1202.
- Baily, M. N. (1974), "Wages and Employment under Uncertain Demand", Review of Economic Studies 41, 37-50.
- Baily, M. N. (1977), "On the Theory of Layoffs and Unemployment", Econometrica 45, 1043-1063.
- Brown, M. and E. Wolfstetter (1985), "Under- and Overemployment in Optimal Layoff Contracts", Zeitschrift Für Nationalökonomie 45, 101-114.
- Chan, K. S. and Y. Ioannides (1982), "Layoff Unemployment, Risk Shifting and Productivity", Quarterly Journal of Economics 97, 212-229.
- Chari, V. V. (1983), "Involuntary Unemployment and Implicit Contracts", Quarterly Journal of Economics 98, Supplement, 107-123.
- FitzRoy, F. R. and R. A. Hart (1985), "Hours, Layoffs and Unemployment Insurance Funding: Theory and Practice in an International Perspective", Economic Journal 95, 700-713.

- Geanakoplos, J. and T. Ito (1981), "On Implicit Contracts and Involuntary Unemployment". Discussion Paper No. 81-155, University of Minnesota.
- Gordon, D. F. (1974), "A Neo-Classical Theory of Keynesian Unemployment", Economic Inquiry 12, 431-459.
- Green, J. and C. M. Kahn (1983), "Wage-Employment Contracts", Quarterly Journal of Economics 98, Supplement, 173-188.
- Grossman, S. and O. Hart (1981), "Implicit Contracts, Moral Hazard and Unemployment", American Economic Review 71, Papers and Proceedings, 301-307.
- Hall, R. and D. M. Lillien (1979), "Efficient Wage Bargains Under Uncertain Supply and Demand", American Economic Review 69, 868-879.
- Holmström, B. (1983), "Equilibrium Long-Term Labor Contracts", Quarterly Journal of Economics 98, Supplement, 23-54.
- Holmström, B. and O. Hart (1985), "The Theory of Contracts", mimeo. Paper presented at the Fifth World Congress of The Econometric Society.
- Ito, T. (1984), "Implicit Contracts with Costly Search: The Incentive Constraints in Severance Payments". Discussion Paper No. 202, University of Minnesota.
- Johansen, L. (1982a), "Some Notes on Employment and Unemployment with Heterogeneous Labour", Nationaløkonomisk Tidsskrift.
- Johansen, L. (1982b), "Arbeidsløsheten: Lettere opp enn ned?" ("Unemployment: Easier Up than Down?"), Sosialøkonomen 36.
- Kahn, C. M (1985), "Optimal Severance Pay with Incomplete Information", Journal of Political Economy 93, 435-451.

Moore, J. (1985), "Optimal Labour Contracts when Workers Have a Variety of Privately Observed Reservation Wages", Review of Economic Studies 52, 37-67.

Mortensen, D. T. (1978), "On the Theory of Layoffs", Discussion Paper No. 332, Northwestern University.

Rosen, S. (1985), "Implicit Contracts: A Survey", Working Paper No. 1687, National Bureau of Economic Research.

Strand, J. (1985), "Work Effort and Search Subsidies with Long-run Equilibrium Contracts", European Economic Review, forthcoming.

Thurow, L. (1981), The Zero-Sum Society, Penguin Books: Harmondsworth.

**Chapter 4:**

**MARITAL DIVISION OF LABOR WITH RISK OF DIVORCE:  
THE ROLE OF "VOICE" ENFORCEMENT OF CONTRACTS**

March 1985

Revised September 1986



**MARITAL DIVISION OF LABOR WITH RISK OF DIVORCE:  
THE ROLE OF "VOICE" ENFORCEMENT OF CONTRACTS**

**ABSTRACT**

The popular view seems to be that a rising probability of divorce leads to less pre-divorce marital specialization between market work and domestic work. People "learn by doing", and in the face of a high risk of divorce one becomes vulnerable if one is too specialized when building up human capital.

Here I investigate one particular "story" about why the probability of divorce might influence pre-divorce allocation of time. I assume that it is not possible to write an enforceable marital contract. I therefore picture that people must rely on emotional ties to "self-enforce" contracts (which I denote "voice enforcement"). As a divorce must be expected to weaken the possibility of such voice enforcement, divorce might influence pre-divorce allocation of time. The direction of this influence is (perhaps surprisingly) ambiguous even when we assume that the spouses' (identical) ex post utility functions are homothetic. However, it remains a distinct possibility that the popular idea of a higher probability of divorce leading to less specialization is correct.

Voice enforcement has no role to play when contracts are enforceable, or when transactions in perfect credit markets can act as substitutes for such contracts. One implication of this is that the "legal infrastructure", i.e. the costs of writing and enforcing contracts, matters for the extent to which the probability of divorce influences pre-divorce allocation of time.

MARITAL DIVISION OF LABOR WITH RISK OF DIVORCE: THE ROLE OF  
"VOICE" ENFORCEMENT OF CONTRACTS\*

1. Introduction.

Traditionally economists have considered the household to be the basic decision unit. In recent years, however, there has been a growing interest in how decisions are reached within a household, and in the institution of marriage. A pioneer in this field has been Gary Becker (e. g. (1973), (1974), (1981)). It is a tendency in this line of work to study economic aspects of marriage isolated from the emotional ones. The "economic aspects" of marriage are most often the gains from the spouses' division of labor between domestic work and market work, or the gains from some homeproduced goods being family-specific public goods.

This paper focuses on why and how the probability of divorce might influence pre-divorce allocation of time. I investigate one particular "story" in which I suppose that it is not possible to write an enforceable marital contract. The spouses must therefore rely on emotional ties to "self-enforce" contracts. As a divorce must be expected to weaken the emotional ties between the spouses and thereby the scope for using these for informal enforcement of contracts, the probability of divorce might influence the pre-divorce allocation of time.

In this paper we concentrate on the case where the gains from marriage arise due to specialization in different lines of work. Suppose that there are some goods which must be produced at home and which do not have perfect market substitutes. Normally, people's comparative advantages would be best utilized if they could trade with the whole market. But as this is blocked, the second best is to "trade" with one's spouse.

A central assumption of this paper is that people "learn-by-doing". If somebody in one period specializes in market work or domestic work he or she will build up human capital of the one kind or the other. We would therefore perhaps expect that the possibility of divorce would influence the value of different kinds of human capital investments, and thereby also the pre-divorce decisions about division of labour. E.g., if one spends much time on domestic work, one builds up human capital in domestic work but not in market work, and this could be a disadvantage after a divorce.

In section 2 I start out with a benchmark model in which it turns out that the probability of divorce does not influence pre-divorce allocation of time. However, the key assumptions of this model are carefully chosen to produce this result: I assume, among other things, that contracts are costless to write and fully enforceable by a third party. Further, I assume that there are no credit markets and that the homeproduced good is a private good. Within this set-up the spouses can insure each other against the risk connected with building up specialized human capital, and a divorce will have no real economic consequences.

What happens when contracts are not fully enforceable? People must then resort to informal ways of enforcing contracts. In section 3 I introduce the concept of "voice enforcement" of contracts, meaning that people rely on emotional ties to implicitly enforce contracts. We would expect marital contracts to be more easily "voice" enforceable if the marriage continues than if the couple is divorced.

However, in section 4 I demonstrate that if credit markets are complete (in a very strong sense) they can substitute for the transfers contingent on divorce specified by an enforceable contract - and therefore voice enforcement has no role to play.

In section 5 I comment upon the third key assumption of the benchmark model, namely that the homeproduced good is an ordi-

nary private good, transportable at no cost. This assumption is made in order to isolate the effect of "voice enforcement": When the homeproduced good is a family-specific public good, or there in some sense are economies of scale from living together, a divorce means that these economies of scale will no longer be utilized: In turn, this means a divorce has real economic consequences no matter whether or not contracts can be enforced. I would like to stress that the models of this paper are tailored to introduce the concept "voice enforcement of contracts" and to investigate under what circumstances this enforcement mechanism might be important. It has not been my aim to provide an all-embracing and fully realistic account of marital life.

The main model of this paper, found in section 6, investigates the case where contracts are not third-party enforceable and where credit markets do not substitute for this lacking enforceability. Here voice enforcement will be taken into use - and this leads to a situation where the probability of divorce does matter for pre-divorce time allocation. However, the direction of this influence is ambiguous even when we assume that the spouses' (identical) ex post utility functions are homothetic. Finally, section 7 concludes the paper.

As already hinted at: I think one basic insight from this analysis is that the economic and emotional sides of a marriage cannot be studied separately, as the emotional aspects might be important for the enforcement of implicit contracts and therefore have economic consequences. This is in contrast to almost the whole body of economic literature on marriage.<sup>1)</sup>

It should be clear from this introduction that parts of the economic theory of marriage bears a close resemblance to the theory of international trade. A married couple exchange market goods and homeproduced goods between themselves - two countries specialize in different lines of production and export and import goods. Moreover, the theory of relationship-specific capital (e.g. Grout (1984), Crawford (1983)) is somewhat parallel to this theory of marriage, as it concentrates on the vulnerability

arising from relationship-specific investments in the absence of long-term contracts. However, the economic theory of marriage is not a mere application of international trade theory or the theory of relationship-specific investment. The main difference is that the spouses, as portrayed here, not only transfer market goods and homeproduced goods to "trade" at an implicit price. They also transfer goods for distributional reasons. E.g., if there were no comparative advantages, but one spouse had absolute advantages in both lines of work, we would expect to see transfers from the more able spouse to his or her partner. Marriage is so to speak not only a trading relationship, but also an institution of redistributive taxation (and many other things).

## 2. A Benchmark Model.

I will here develop a basic model which will be used as a point of reference throughout the paper. This model is quite similar to those of Gronau (1973, 1977), except for the possibility of divorce being allowed. The key assumption of this version of the model will be that contracts are fully enforceable, that there are no credit markets, and that homeproduced goods are private goods. These assumptions will produce the result that divorce does not matter for the family allocation of time.

The model is a two-period one. In the beginning of period one two persons, denoted A and B, enter into marriage for emotional reasons, exogenous to the model. At the start of their marriage the spouses decide on their present and future allocation of time. They also agree upon present and future money transfers and transfers in kind of homeproduced goods between themselves. These decisions, "the marital contract", will in this section be assumed to constitute a binding and enforceable contract. Further, these decisions are assumed to be taken "in harmony", i. e., to maximize a family welfare function. (An alternative approach would be models of "conflict", where the gains from marriage are split according to bargaining strength (see e. g. Manser and Brown (1980) and McElroy and Horney (1981))).

But albeit this marital contract is entered in harmony, the spouses take into consideration that they will be divorced at the end of the first period with subjective probability  $q$ . Divorce occurs for emotional reasons, exogenous to the analysis. The reader might object that in real life divorces do not simply happen - they are the results of people's choices, perhaps undertaken to maximize utility. I.e., the divorce probability ought to be made endogenous to the analysis. In such a broader model, whether a divorce occurred or not would depend on the economic consequences of divorce, which in turn would depend on the spouses's degree of pre-divorce specialization in market work and domestic work. Divorce could also be due to exogenous factors, such as emotions. Comparative statics then could be performed with respect to these exogenous factors. I have felt that my main points are best expressed within the present, simpler model, where the probability of divorce is taken to be exogenous. It might even be realistic that divorces occur due to the emotional breakdown of the marriage, rather than being due to economic calculations?<sup>2)</sup>

The family welfare function,  $W$ , will be assumed to be additively separable, with the two spouses' ex ante utilities entering with equal weights:<sup>3)</sup>

$$W = U^A + U^B \quad (1)$$

$U^A$  and  $U^B$  are the interpersonally comparable ex ante utilities of A and B.

Let  $u^{js}$  denote the ex post utility of spouse  $j = A, B$ , in time-state  $s = 0$  (first period), 1 (second period, given continued marriage), 2 (second period, given divorce). I will assume that  $U^j$  can be written as:

$$U^j = u^{j0} + D[(1-q) u^{j1} + q u^{j2}] \quad \text{for } j = A, B \quad (2)$$

$D$  is a constant discount factor, equal for the two spouses. The time additivity of the utility function will make comparative

statics exercises somewhat simpler. Apart from the comparative statics in section 6 this assumption will not be restrictive.

Further we have the relationship:

$$u^{js} = u(M^{js}, H^{js}, \bar{L}^s) \quad \text{for } j = A, B \quad s = 0, 1, 2 \quad (3)$$

M denotes consumption of market goods, H of homeproduced goods, and  $\bar{L}$  of leisure. Marginal utilities are positive, and u is a concave function. Note that the function u is taken to be the same for A and B. To simplify, leisure will be taken to be given exogenously, at the same level for both spouses. It would be more realistic to treat leisure as a choice variable. A further step toward realism would be to incorporate that a married couple might have a liking for having leisure together (which indicates a consumption externality).

As there are no credit markets in this model, people spend their entire money income on market goods in the period when this income accrues. Money income consists of wage income  $w^{js} \cdot (1-h^{js})$  (where w denotes wage, and h the fraction of the time available for market work and domestic work which is spent on domestic work) and of a possible money transfer, a, which is received by one spouse from the other.<sup>4)</sup> We have:

$$M^{js} = w^{js}(1-h^{js}) + a^{js} \quad j = A, B \quad s = 0, 1, 2 \quad (4)$$

$$a^{As} = -a^{Bs} \equiv a^s \quad s = 0, 1, 2 \quad (5)$$

(The price of the market good is normalized to one.)

The homeproduced good is assumed to be a private good - this is an assumption that will be discussed below. We assume:

$$H^{js} = d^{js} f(h^{js}) + b^{js} \quad j = A, B \quad s = 0, 1, 2 \quad (6)$$

We assume that the spouses' production functions for homeproduced goods only differ by the "productivity coefficients"  $d^{js}$ .

We assume  $f' > 0$  and  $f'' < 0$ .  $b^{js}$  denotes transfers of home produced goods to spouse  $j$  from the other spouse, in time-state  $s$ . Clearly,

$$b^{As} = -b^{Bs} \equiv b^s \quad s = 0, 1, 2 \quad (7)$$

We will further assume that people "learn by doing". That is, the spouses' productivities in different lines of work in period 2 are dependent on how many hours they spent at these types of work in the first period. We write:

$$w^{j1} = w^{j2} = w^j(-h^{j0}) \quad \text{for } j = A, B \quad (8)$$

$$d^{j1} = d^{j2} = d^j(h^{j0}) \quad \text{for } j = A, B \quad (9)$$

We assume  $w^j$  and  $d^j$  to be strictly increasing, concave functions. The productivities at the outset of the marriage, measured by  $w^{j0}$  and  $d^{j0}$ , will be taken to be exogenous.

The learning-by-doing conditions (8) and (9) are crucial to the analysis. These conditions imply that the decisions at time 1 have consequences in period 2. Therefore, the possibility of divorce in period 2 may influence the couple's allocation of time in period 1. More specifically: The choice between working at home or in the market in period 1 is (also) an investment in future ability. This might be a risky investment, as divorce might alter the relative valuation of period 2 abilities. It is this link between periods 1 and 2 which makes the possibility of divorce influencing pre-divorce allocation of time. This holds for all the models in this paper.

The conditions (2) - (9) now can be substituted into (1). The welfare maximization problem of the family then reduces to the unconstrained maximization of (1).

It is now easy to state the first-order conditions of this problem and show that they are independent of  $\alpha$ . This is done in Appendix A. The first-order conditions simply say that in opti-



mum the marginal utility from time spent on domestic work shall equal the marginal utility from time spent on market work in all time-states, and that the transfers shall be driven to the point where the marginal utility gain of the receiving spouse is equal to the marginal utility loss of the other spouse.

What is the intuition behind this result that the family allocation of time is independent of  $q$ ? In the model marriage can be viewed as having an emotional and an economic side. The economic aspects of a marriage consist of two people being able to realize gains from specialization. With the described marital contract, this specialization survives the dissolution of the marriage: At the start of the marriage the couple binds itself to transferring the same amount of money and H-goods irrespective of a divorce having occurred or not. In this sense divorce implies the end of the emotional side of the marriage, but not the economic one. Therefore it is not surprising that the probability of divorce has no influence on the allocation of the spouses' time between market work and domestic work.

But as already underlined, this version of the model is a benchmark one: Its key assumptions were chosen not for their realism, but to produce the irrelevance-of-probability-of-divorce result. In the next three sections we will try to get a better understanding of the three central assumptions of this version of the model: That contracts are fully enforceable, that there are no credit markets and that the homeproduced good is a private good.

Note that if we should take this benchmark setting seriously, (which I think we should not) such that the economic and emotional aspects of a marriage really could be separated, there should be no reason why one should not have one "economic spouse" and one "emotional spouse". (By "economic aspects" of the marriage is here meant the transfer agreements.) One could for instance marry "economically" somebody which was very different from oneself - in order to maximize gains from specialization. Emotionally one would perhaps prefer to choose otherwise.

### 3. Voice Enforcement of Marital Contracts.

Several authors have focused on the problems of enforcement of contracts by a third party, due to the costs of specifying and verifying the contract elements. When third-party enforcement is not possible the parties to the contract must rely on "self-enforcement", i. e., they must limit themselves to contracts which no party has an incentive to breach. E. g. Klein and Leffler (1981), Shapiro (1983) and Strand (1983) focus on the threat of terminating a relationship as a discipline device - and characterize the resulting consequences. In Hirschmann's (1970) terminology: These authors focus entirely on "exit" as a discipline device, but neglect enforcement by "voice".

In many contexts people care for their esteem in the eyes of a contract partner: People fear a loss of reputation, not only because a bad reputation may lead to some relationships being terminated, but because they care about their reputation in itself. This kind of reputational enforcement I will denote "voice enforcement".<sup>5)</sup> Presumably, voice enforcement will be most important in cases where people are strongly emotionally attached to each other. Marital contracts would then be a prime example of contracts where voice enforcement is important, but not necessarily the only one.

In this paper I will model voice enforcement in a simplified manner: I will assume that as long as a couple remains married their marital contract will be fully enforceable by voice enforcement. As soon as they are divorced, however, no contract is enforceable. Realistically, even when husband and wife care about what the other thinks of them, this does not necessarily imply that contracts between them are fully enforceable. Conversely, people may also care about what their former spouse thinks of them after a divorce (but presumably to a lesser extent than had they remained married): This implies that contracts contingent on divorce might not be completely unenforceable. To sum up: Our assumption that contracts are fully enfor-

ceable when a couple remains married and not at all enforceable after a divorce may be seen as dramatizing the difference between the states divorce/no divorce.

Note that in this paper it is the possibility of exit enforcement that is assumed away, as divorce is taken to occur exogenously. But note further that if divorce somehow had been endogenized in the model, exit enforcement still would not play an important role here. This is because we have assumed voice enforcement to be very forceful. Contracts are perfectly enforced through voice enforcement as long as the couple stays together, only after a divorce do problems occur. These problems, however, cannot be resolved by the threat of divorce (exit enforcement) as the couple already is divorced.

In relation to our benchmark model, our enforceability constraint can be stated formally as:

$$a^2 = b^2 = 0 \tag{10}$$

But, as we soon shall see, whether or not this enforceability constraint really is binding, depends heavily on the credit market structure.

I think a remark on alimony is due at this point. By alimony I will here mean transfers between the spouses after a divorce, established by a judge or by administrative rules, as opposed to transfers which originate from voluntary contracts. Clearly, alimony in this sense to some extent substitutes for contracts contingent on divorce. But there are limits to such substitution: First, an alimony set by a judge will only by coincidence be equal to the optimal contracted transfer of  $a^2$ . This is both because a judge does not know the parameters and functional forms of the problem as well as the spouses themselves and because his aim may not be to maximize family welfare. His goal could for instance be to secure a minimum level of welfare for the least well-to-do spouse, or make the parent who does not live together with the children pay a fair part of the expenses

of child-rearing. Second, such an alimony will normally be a money transfer - and when homeproduced goods to some extent are transferable the optimal contract might entail transfers of homeproduced goods between divorced spouses.

Carl Shapiro (1983) has drawn attention to the importance of "standards" when there are problems of enforcing contracts. A "standard", e.g. a minimum quality standard, will on the one hand limit the freedom of contracting, but on the other hand reduce the enforceability problem. The optimal standard strikes a balance between these two considerations. I think alimony in the sense described here functions in a rather parallel manner to such "standards".

#### 4. Credit Market Transactions as Substitutes for Transfers Contingent on Divorce.

In this version of the model I will assume that the enforceability constraint (10) holds, but that there at the same time are perfect credit markets. I will assume that there exist both a competitive credit market for money (i. e., for market goods) and a competitive credit market for non-marketed homeproduced goods. Especially the latter part of this assumption is of course dramatically unrealistic. However, these very strong assumptions bring out very clearly how credit markets can be used as substitutes for enforceable contracts. Perhaps one insight from this section is precisely that credit markets must be perfect in a very strong sense if they shall fully substitute for contracted transfers. The realism of the complete credit markets assumption will be further discussed below.

From the benchmark model in section 2 conditions (1) - (3), (5), (7) - (9) still apply. In addition we have the enforceability constraint (10). However, the budget constraints for market goods and homeproduced goods, (4) and (6), must be replaced by the following conditions:

$$M^{js} = w^{js} (1-h^{js}) + a^{js} + m^{js} \quad \text{for } j = A, B \quad (11)$$

$$s = 0, 1, 2$$

$$H^{js} = d^{js} f(h^{js}) + b^{js} + n^{js} \quad \text{for } j = A, B \quad (12)$$

$$s = 0, 1, 2$$

Here  $m^{js}$  denotes money loans for person  $j$  in time-state  $s$ . (Savings can be viewed as negative loans.) Correspondingly,  $n^{js}$  denote loans in the credit market for homeproduced goods. Clearly, we must have the following restrictions:

$$m^{j1} = m^{j2} = - (1+r)m^{j0} \quad \text{for } j = A, B \quad (13)$$

$$n^{j1} = n^{j2} = - (1+i)n^{j0} \quad \text{for } j = A, B \quad (14)$$

Here  $r$  and  $i$  denote the interest rates in the two credit markets.

Now, by studying the first-order conditions of this problem we can see that the optimal solution is the same as in the benchmark model. This is shown in Appendix B. In other words, the credit markets have rendered the use of transfers contingent on divorce superfluous.

I will here outline the intuition behind this result: Say that in the optimal solution, B transfers  $a^{2*}$  to A. When transfers contingent on divorce are not enforceable, B can increase his or her transfer to A in period 1 by  $a^{0*}$ . This transfer can be financed by a loan at interest rate  $r$ . A saves this money, and will in period 2 receive  $(1+r)a^{0*}$ . If the marriage has survived, A returns the money to B, who uses them to repay his debt, which now amounts exactly to  $(1+r)a^{0*}$ . In case of divorce, however, A keeps this money, and B must repay the debt out of his or her own pocket. With an appropriate size of  $a^{0*}$ , this would exactly correspond to B paying A  $a^{2*}$  contingent on divorce. The credit market for homeproduced goods can be used in a parallel manner to render transfers of H-goods contingent on divorce superfluous.

However, perfect credit markets in the sense above do not exist. First of all, there is no credit market for homeproduced goods. Why is this? Generally, we must expect that for the same reasons as there are no spot markets for homeproduced goods (high costs of transportation and marketing), there will not exist credit markets for these goods either. How serious this is, depends on how close substitutes market goods and homeproduced goods are: If they are close substitutes, the loss from not having access to a credit market for homeproduced goods or from not being able to transfer H-goods contingent on divorce might not be very large.

But also the assumption of a perfect credit market for money is a strong assumption. In many economies consumer loans are rationed. But even with an unregulated credit market, people may feel that they are rationed. With uncertainty and asymmetric information people may experience rationing because they cannot convince the banks that their future earnings prospectives are as promising as they themselves believe. (For one study of endogenous credit rationing, see Stiglitz and Weiss (1980).)

Even with perfect credit markets, our "story" depends on the spouses having separate ownership to bank accounts, and a willingness to use credit markets as prescribed here. It might be argued that such visible preparations for a possible divorce would involve emotional strains and might carry social stigma. The reader must be his own judge about the relevance of this argument.<sup>6)</sup>

Credit markets may be incomplete in many ways: In the model in Section 6 I have chosen to examine the extreme case with no credit markets at all. However, we turn first to a discussion of the third key assumption of the benchmark model, namely that the homeproduced good is an ordinary private good.

## 5. The Nature of the Homeproduced Good.

In this paper we have mostly assumed that the homeproduced good is an ordinary private good, transferable at no cost. This assumption has not been chosen for its realism, but rather to bring more clearly out that one reason why divorce might influence the family's allocation of time is because it influences the possibility of enforcing implicit contracts. Clearly, there are public good aspects of many homeproduced goods, as child rearing, house cleaning etc. Moreover, many homeproduced goods are services which can be rather costly to transfer to somebody you do not live together with. There might also be emotional costs connected with performing services for a former husband or wife.<sup>7)</sup>

A key aspect of all models presented here is that homeproduced goods are not marketed, nor do they have perfect market substitutes. The underlying explanation for this must be that there are high costs of transportation and marketing. This does not go very well together with our assumption that homeproduced goods are ordinary private goods, transportable at no cost. This must be kept in mind while interpreting the results. In Appendix C I show how the benchmark model is altered when an assumption about the homeproduced good being a family-specific public good is added. We see that now the probability of divorce influences the allocation of time irrespective of whether contracts are enforceable. A divorce has real economic effects in that the economies of scale from living together is no longer utilized. (Appendix C should be read after Section 6, as the analysis is somewhat parallel.)

The analysis of Appendix C might perhaps be viewed as the economics of sharing a flat, rather than the economics of a family. On the other hand, our theory of voice enforcement demands emotional ties between the spouses, and can therefore be viewed more specifically as a theory of the family.<sup>8)</sup>

## 6. Enforceability Constraints/No Credit Markets.

I will now turn to a model where there are enforceability constraints with regard to transfers contingent upon divorce and where there are no credit markets. Homeproduced goods will still be considered private goods. The discussion in the previous sections suggests that within this setting divorce probability will influence time allocation. We now turn to study how time allocation will be influenced, and will especially be interested in the effect on pre-divorce time allocation.

The natural model choice would now perhaps be equations (1) - (9), with addition of the enforceability constraints (10). However, this constitutes a model with ten choice variables, which in turn would render the comparative statics rather messy. I have therefore chosen to work within a simplified model.

It will now be necessary to say something about the relative productivities of the spouses. I will assume that spouse A always has an absolute advantage in market work, and that spouse B always has an absolute advantage in domestic work. Formally, irrespective of choice of  $h^{j0}$ , I assume:

$$w^A > w^B, \quad d^A < d^B \quad (15)$$

Remember that  $w^j$  denotes the common value of  $w^{j1}$  and  $w^{j2}$ .

### A simplified model

I will assume that the state of affairs initially can be described as the equilibrium of the model consisting of equations (1) - (10), for a given  $q$ . However, in the short run after a change in  $q$  only the time allocation of spouse B will be variable. The time allocation of spouse A and the transfers will be treated as constants.

A justification for these assumptions could be that we considered spouse B as the marginal worker of the family. The labor



supply of spouse B, which we in a sexist manner might envisage to be the female, would then be more easily adjusted than that of her husband. We could also picture that transfers are more sluggishly adjusted than the labor supply of the wife, for instance because varying transfers requires communication and consensus between two persons.

These assumptions may or may not be deemed realistic. In any case, they grossly simplify the problem. I further believe that this simplified version of the model brings out insights that are valid also in more realistic specifications of the problem.

A further simplifying assumption will turn out to be convenient: I assume that the utility function  $u$  is homothetic. The homotheticity assumption implies that relative demand for the two commodities depends only on relative prices (measured in time cost). This excludes income effects from playing a role. This is important to have in mind when interpreting the results.

Treating  $h^{A0}$ ,  $h^{A1}$ ,  $h^{A2}$ ,  $a^0$ ,  $a^1$ ,  $a^2$ ,  $b^0$ ,  $b^1$  and  $b^2$  as constants, the first-order conditions of the problem will be as follows:

$$\begin{aligned} \frac{\partial W}{\partial h^{B0}} = & \{ u_H^{B0} d^{B0} f'(h^{B0}) + D[(1-q) u_H^{B1} d^{B1} f(h^{B1}) \\ & + q u_H^{B2} d^{B2} f(h^{B2})] \} - \{ u_M^{B0} w^{B0} + D[(1-q) u_M^{B1} w^{B1} (1-h^{B1}) \\ & + q u_M^{B2} w^{B2} (1-h^{B2})] \} = 0 \end{aligned} \quad (16)$$

$$\frac{\partial W}{\partial h^{B1}} = (1-q) D[ u_H^{B1} d^{B1} f'(h^{B1}) - u_M^{B1} w^{B1} ] = 0 \quad (17)$$

$$\frac{\partial W}{\partial h^{B2}} = q D[ u_H^{B2} d^{B2} f'(h^{B2}) - u_M^{B2} w^{B2} ] = 0 \quad (18)$$

In all three time states the marginal utility derived from domestic work shall equal that derived from market work.

**The Time Allocation at Date 2.**

Our ultimate concern will be how  $h^{B0}$  is influenced by a change in  $q$ . However, as a step towards this end it is necessary to find out how  $h^{B1}$  and  $h^{B2}$  compare. In the Benchmark Model we had the result that  $h^{j1} = h^{j2}$  for  $j = A, B$  which lead to the irrelevancy-of-divorce-probability result.

State 1 and state 2 differ only in that transfers are allowed in the first but not in the latter. It is therefore rather intuitive that the comparison of  $h^{B1}$  and  $h^{B2}$  depends on the directions of transfers in state 1: Our first task will be to establish these. When it comes to comparative statics we have assumed that the transfers are constants. Remember however that the sign and magnitude of these transfers originally was established as the optimal solution to the model consisting of (1) - (10), for the original value of  $q$ . It can be established that in state 1 spouse A transfers market goods to B, whereas B transfers home-produced goods to A, i.e.,  $a^1 < 0$ ,  $b^1 > 0$ . This is shown formally in Appendix D. Note that I had to use assumption (15) about absolute advantages to obtain this. People familiar with Ricardian trade theory might think that an assumption of comparative advantages should suffice. But as underlined in the introduction, transfers within a family is not only "trade", but also redistribution to maximize family welfare. Therefore, the theorems of trade theory are not directly applicable on this arena.

Now we can turn to the main issue of this subsection, namely the comparison of  $h^{B1}$  and  $h^{B2}$ . Using the first-order conditions (17) and (18), we obtain:

$$\frac{u_H^{B1}}{u_M^{B1}} f'(h^{B1}) = \frac{u_H^{B2}}{u_M^{B2}} f'(h^{B2}) \tag{19}$$

We define  $p^{B1}$  and  $p^{B2}$  implicitly by rewriting (19) as:

$$p^{B1} f'(h^{B1}) = p^{B2} f'(h^{B2}) \quad (20)$$

As we have assumed  $u$  is a homothetic function,  $p^{B1}$  is a decreasing function of the ratio  $H^{B1}/M^{B1}$ , and  $p^{B2}$  is the same decreasing function of  $H^{B2}/M^{B2}$ .  $p^{B2}$  is therefore a decreasing function only of  $h^{B2}$  (as there are no transfers in this state), whereas  $p^{B1}$  is a function of  $h^{B1}$ ,  $a^1$  and  $b^1$ .

Let us start out with the assumption that  $h^{B1} = h^{B2}$  (at any level). We know that the two states differ by the transfers  $a^1 < 0$  and  $b^1 > 0$ . This implies:

$$\frac{H^{B1}}{M^{B1}} < \frac{H^{B2}}{M^{B2}} \quad \text{for } h^{B1} = h^{B2} \quad (21)$$

Then, obviously

$$p^{B1} > p^{B2} \quad \text{for } h^{B1} = h^{B2} \quad (22)$$

When we keep  $h^{B2}$  at the same level as  $h^{B1}$ , going from state 1 to state 2 only implies the removal of the transfers. This reduces the relative value of H-goods to M-goods, according to condition (22). In turn, this seems to give B incentives to work less at home and more in the market. And, indeed, condition (20) and (22) together imply:

$$h^{B1} > h^{B2} \quad (23)$$

The spouse which is most able in domestic work will work less at home in case of divorce than if the marriage had continued.

It was the difference between  $p^{B1}$  and  $p^{B2}$  which encouraged spouse B to work relatively less at home after a divorce. It seems therefore natural to believe that B will never work so much at home as to reverse the inequality of (22). This turns out to be correct. Taking (23) together with (20) we have that in optimum:

$$p^{B1} > p^{B2} \tag{24}$$

**The Probability of Divorce and the Pre-Divorce Time Allocation.**

A standard classical comparative statics exercise, which can be found in Appendix E, yields the following result:

$$\text{sign } \frac{dh^{B0}}{dq} = \text{sign } \beta_1,$$

$$\text{where } \beta_1 = [u_H^{B2} f(h^{B2}) - u_H^{B1} f(h^{B1})] d^{B'} - [u_M^{B2} (1-h^{B2}) - u_M^{B1} (1-h^{B1})] w^{B'} \tag{25}$$

Note that the assumption of  $u$  being homothetic underlies the particular simple form of (25).

Let us try to interpret  $\beta_1$ . We know that both market work and domestic work lead to a period 2 productivity increase in the respective line of work. As one works more at home and less in the market in the one state than the other, the probability of the two states clearly affects the value of such productivity increases due to learning-by-doing. Through this channel first period time allocation is influenced. If one increases  $h^{B0}$  this increases productivity in domestic work and reduces the productivity in market work in period 2. The quantity  $\beta_1$  is the derivative of this combined productivity increase and decrease with respect to  $q$  (with the positive constant  $D$  removed). Therefore, a positive  $\beta_1$  will make domestic work in period 1 more attractive relative to market work after an increase in  $q$ , and vice versa for a negative  $\beta_1$ .

However, it turns out that the sign of  $\beta_1$  cannot be decided unambiguously. First, we know from condition (23) that  $h^{B1} > h^{B2}$ . This means that after a divorce spouse B works relatively less at home. This implies that the (expected) value of increased

productivity in domestic work and decreased productivity in market work will fall if  $q$  rises. And we see from (25) that  $h^{B1} > h^{B2}$  tends to make  $\beta_1$  negative.

But the expected value of productivity increases and decreases in the two lines of work does not only depend on how many units of H-goods and M-goods that are produced in the two states, but also on the value of these goods. We see from (25) that the relevant expressions of value are the marginal utilities of the two goods in the two states. We know from condition (24) that  $p^{B1} > p^{B2}$ . This is compatible with three cases

$$(a) \quad u_H^{B1} > u_H^{B2} \quad \text{and} \quad u_M^{B1} < u_M^{B2}$$

$$(b) \quad u_H^{B1} < u_H^{B2} \quad \text{and} \quad u_M^{B1} < u_M^{B2}$$

$$(c) \quad u_H^{B1} > u_H^{B2} \quad \text{and} \quad u_M^{B1} > u_M^{B2}$$

If case (a) holds, we see from (25) that with  $h^{B1} > h^{B2}$  the sign of  $\beta_1$  is unambiguously negative.

Let us now look at case (b). Here the marginal utilities of both goods are higher in state 2 than in state 1. This implies that spouse B has less of both goods after a divorce, i.e., she is "poorer". (As both spouses lose from less specialization, whereas one loses and one gains from redistributive transfers being stopped, we can either have that both spouses are poorer in state 2 than 1, or that one spouse is better off and the other is worse off.) As  $u_H^{B2} > u_H^{B1}$  this might make the first bracketed expression positive. In turn, if  $d^{B'}$  is large enough relative to  $w^{B'}$ , we might get that  $\beta_1$  is positive.

To get some sort of a grip on the intuition behind this result, I think it is useful to consider the special case where  $w^{B'} = 0$  and  $d^{B'} > 0$ . Now, after a divorce spouse B works less at home

and homeproduced goods are less worth compared with market goods (conditions (23) and (24)). This tends to make a productivity increase in domestic work less valuable. On the other hand, spouse B is poorer after a divorce, which implies that her expected utility in the second period is less the higher the probability of divorce. The higher the probability of divorce, the more spouse B wants to "move utility" from the first to the second period. But as she in this model has no access to credit markets or state-contingent markets, and as  $w^{B'} = 0$ , this can only happen by her building up much productivity in domestic work. This effect taken alone indicates that spouse B works more domestically if the divorce probability rises. And we cannot know which effect dominates.

However, it seems very unrealistic that it should be very much easier to build up human capital by learning-by-doing in domestic work than in market work. On the contrary, we would expect the reverse to hold. If we assume  $d^{B'} = 0$  and  $w^{B'} > 0$ , we see from (25) that  $\beta_1$  is unambiguously negative also in case (b).

Finally, let us turn to case (c). Here the marginal utility of both goods are lower after a divorce, implying that spouse B has more of both goods, i.e., she is "richer". As  $u_M^{B1} > u_M^{B2}$ , it might be that the second bracketed term in (25) is negative. With  $w^{B'}$  large relative to  $d^{B'}$  (which I find more realistic than the reverse),  $\beta_1$  might be positive. The intuitive explanation follows the same lines as in case (b).

To sum up: We have not been able to sign  $\beta_1$  unambiguously. However, there are strong effects that point to the "intuitive" result that  $\beta_1$  is negative. At least it is a fair chance that the common belief that an increase in the probability of divorce leads to less pre-divorce specialization is correct within the setting of this section.

## 7. Concluding Remarks.

The main message of the paper has been that one possible route through which divorce probability can influence time allocation is that divorce weakens the scope for "voice enforcement" of implicit contracts between spouses. The direction of this influence is ambiguous, but it is not unlikely that a higher probability of divorce would lead to less specialization due to productivity differences between the spouses. This could be one possible explanation for the joint rise of the number of divorces (and thereby people's subjective probability of divorce) and of female labor market participation, observed in many countries. This stylized fact, of course, could have many other explanations as well: For instance the causality might be the reverse: Less specialization makes the economic consequences of divorce less drastic, and therefore the probability of divorce rises. Further, the joint rise of divorces and female labor market participants might not be linked by direct causality, but both be caused for instance by changing attitudes to female roles.

If it is true that the enforcement problem is of significant practical importance, a policy implication might be to make explicit contracting cheaper. With low probabilities of divorce, voice enforcement may have rendered explicit contracting superfluous. In the face of present divorce rates, this might no longer be so. Measures towards making explicit contracts cheaper could include the design of standard contracts, or simplifying the system of legal enforcement.

Alternatively, the analysis could be given a normative interpretation. It might well be that most families do not maximize family welfare as described here. Many actual choices tend to be far less "rational" than economists perceive them to be. But if a couple really wants to maximize family welfare, they should make use of contracting possibilities and credit market transactions as hinted at here. The existing contracting possibilities could be rather good, were they only taken into use. Especially, the use of credit markets to substitute for contracts, I suspect is rather unfamiliar to most people. Perhaps a married couple

has better possibilities of insuring themselves against falls in the value of different types of human capital than one might think, and these possibilities should be taken advantage of.



**APPENDIX A.**

In this appendix we state the first-order conditions of the benchmark model in Section 2, and show that the optimal solution is independent of  $q$ . We will assume that the first-order conditions characterize an interior maximum. Subscripts H and M will denote partial derivatives with respect to H and M in the respective time-state. The first-order conditions are as follows:

$$\begin{aligned} \frac{\partial W}{\partial h^{j0}} = & \{u_H^{j0} d^{j0} f'(h^{j0}) + D[(1-q)u_H^{j1} d^{j1} f'(h^{j1}) \\ & + q u_H^{j2} d^{j2} f'(h^{j2})]\} - \{u_M^{j0} w^{j0} + D[(1-q) u_M^{j1} w^{j1} (1-h^{j1}) \\ & + q u_M^{j2} w^{j2} (1-h^{j2})]\} = 0 \quad \text{for } j = A, B \end{aligned} \quad (A1)$$

$$\begin{aligned} \frac{\partial W}{\partial h^{j1}} = & (1-q) [D u_H^{j1} d^{j1} f'(h^{j1}) - D u_M^{j1} w^{j1}] = 0 \\ & \text{for } j = A, B \end{aligned} \quad (A2)$$

$$\begin{aligned} \frac{\partial W}{\partial h^{j2}} = & q [D u_H^{j2} d^{j2} f'(h^{j2}) - D u_M^{j2} w^{j2}] = 0 \\ & \text{for } j = A, B \end{aligned} \quad (A3)$$

Before we look at the first-order conditions which characterize optimal transfers, we will try to interpret conditions (A1) - (A3). Even though these conditions may seem messy, their interpretations turn out to be straightforward. Condition (A1) characterizes the choice of  $h^{j0}$  in optimum. In optimum, the marginal utility derived from time spent on domestic work shall equal the marginal utility derived from time spent working in the market. The marginal utility from spending time on domestic work in the first period consists of the extra utility from having more homeproduced goods in the first period, but also from having more H-goods in both states in the second period, due to the learning effects. The marginal utility loss from working less in

the market in the first period can be split into corresponding terms. And as mentioned, in optimum the marginal utility from having more homeproduced goods in all time-states shall equal the marginal utility loss from having less market goods.

Conditions (A2) and (A3) have similar interpretations: that the marginal utility derived from time spent on domestic work in optimum shall equal the marginal utility derived from time spent on market work. However, conditions (A2) and (A3) have a somewhat simpler structure than (A1), as there are no learning effects in the last period.

Before we state the last six first-order conditions, it will be convenient to introduce some new notation. We define:

$$V_M^{js} = \frac{1}{p_s} \cdot \frac{\partial U^j}{\partial M^{js}} \quad \text{with } p_s = \begin{cases} 1 & \text{if } s = 0 \\ (1-q) & \text{if } s = 1 \\ q & \text{if } s = 2 \end{cases} \quad (\text{A4})$$

$$V_H^{js} = \frac{1}{p_s} \cdot \frac{\partial U^j}{\partial H^{js}} \quad \text{with } p_s = \begin{cases} 1 & \text{if } s = 0 \\ (1-q) & \text{if } s = 1 \\ q & \text{if } s = 2 \end{cases} \quad (\text{A5})$$

The V's can be interpreted as the (family and individual) marginal utilities of homeproduced goods and market goods in different periods and states - except for the probability of the relevant state being separated out.

The last six first-order conditions are:

$$\frac{\partial W}{\partial a^0} = V_M^{A0} - V_M^{B0} = 0 \quad (\text{A6})$$

$$\frac{\partial W}{\partial b^0} = V_H^{A0} - V_H^{B0} = 0 \quad (\text{A7})$$

$$\frac{\partial W}{\partial a^1} = (1-q) V_M^{A1} - (1-q) V_M^{B1} = 0 \quad (\text{A8})$$

$$\frac{\partial W}{\partial b^1} = (1-q) V_H^{A1} - (1-q) V_H^{B1} = 0 \quad (A9)$$

$$\frac{\partial W}{\partial a^2} = q V_M^{A2} - q V_M^{B2} = 0 \quad (A10)$$

$$\frac{\partial W}{\partial b^2} = q V_H^{A2} - q V_H^{B2} = 0 \quad (A11)$$

Conditions (A6) - (A11) say that transfers - both of money and of homeproduced goods - shall be driven to the point where the marginal utility gain of the receiving spouse is equal to the marginal utility loss of the other spouse.

Within the present setting, how is the family allocation of time influenced by the probability of divorce? First, note that conditions (A2) - (A3) and (A6) - (A11) are independent of  $q$ , except for the possible effect through  $q$  influencing pre-divorce allocation of time, which in turn influences post-divorce allocation of time through the learning-by-doing effects. Note also that as long as  $u$  and  $f$  are state-independent, as they here are assumed to be, conditions (A2) and (A3) imply:

$$h^{j1} = h^{j2} \quad \text{for } j = A, B \quad (A12)$$

Further, when  $h^{j1} = h^{j2}$  for  $j = A, B$ , we also have  $M^{j1} = M^{j2}$  and  $H^{j1} = H^{j2}$  for  $j = A, B$ . Condition (A1) now can be rewritten as:

$$\{u_H^{j0} d^{j0} f'(h^{j0}) + D u_H^{j1} d^{j1} f(h^{j1})\} \\ - \{u_M^{j0} w^{j0} + D u_M^{j1} w^{j1} (1-h^{j1})\} = 0 \quad \text{for } j = A, B \quad (A1')$$

As we see, this condition turns out to be independent of  $q$ . This means that all the first-order conditions of the family's maximization problem are independent of  $q$ .

**APPENDIX B.**

The two new first-order conditions which characterize the optimal use of the credit markets are:

$$\frac{\partial W}{\partial m^{j0}} = v_M^{j0} - (1+r)[(1-q) v_M^{j1} + q v_M^{j2}] = 0 \quad (B1)$$

for  $j = A, B$

$$\frac{\partial W}{\partial n^{j0}} = v_H^{j0} - (1+i)[(1-q) v_H^{j1} + q v_H^{j2}] = 0 \quad (B2)$$

for  $j = A, B$

It follows from the first-order conditions (A6) and (A8), which also apply here, that:  $v_M^{A0} = v_M^{B0}$  and  $v_M^{A1} = v_M^{B1}$ . Condition (B1) then implies:

$$v_M^{A2} = v_M^{B2} \quad (B3)$$

Correspondingly: From (A7) and (A9) we have that  $v_H^{A0} = v_H^{B0}$  and  $v_H^{A1} = v_H^{B1}$ . Then condition (B2) implies:

$$v_H^{A2} = v_H^{B2} \quad (B4)$$

We see from the first-order conditions (A10) and (A11) that if transfers contingent on divorce had been enforceable, such transfers would have been driven to the point where  $v_M^{A2} = v_M^{B2}$  and  $v_H^{A2} = v_H^{B2}$ . But we see from conditions (B3) and (B4) that these marginal utilities are equalized even without the use of transfers when credit markets are perfect in the sense above. In other words, credit market have rendered the use of transfers contingent on divorce superfluous.

**APPENDIX C.**

I will here indicate how our analysis would change if we loosened the assumption that the homeproduced good is an ordinary private good. I will do this by investigating the benchmark model of section 2 with an added assumption that homeproduced goods are family-specific public goods. In this setting divorce does play a role, even when contracts are costless to write and to enforce.

The benchmark model consisted of equations (1) - (9). Equations (1) - (5) and (8) - (9) still apply. Equations (6) and (7), however, have to be replaced by (6'), (6'') and (7').

$$H^{js} = d^{As} f(h^{As}) + d^{Bs} f(h^{Bs})$$

for  $j = A, B \quad s = 0, 1 \quad (6')$

$$H^{j2} = d^{j2} f(h^{j2}) + b^{j2} \quad \text{for } j = A, B \quad (6'')$$

$$b^{A2} = - b^{B2} = b^2 \quad (7')$$

Conditions (6') and (6'') express that as long as a couple lives together they both enjoy their combined production of H-goods. After a divorce one consumes only those H-goods one produces oneself. As the homeproduced good is a family-specific public good, transfers of this good is meaningless except in state 2. This is expressed by (7').

To save space, I will only state the first-order conditions concerning A's allocation of time. The first-order conditions for B would be corresponding.

$$\frac{\partial W}{\partial h^{A0}} = \{ (u_H^{A0} + u_H^{B0}) d^{A0} f'(h^{A0}) + D[ (1-q)(u_H^{A1} + u_H^{B1}) d^{A'} f'(h^{A1}) + q u_H^{A2} d^{A'} f'(h^{A2}) ] \} - \{ u_M^{A0} w^{A0} + D[ (1-q) u_M^{A1} w^{A'} (1-h^{A1}) \}$$

$$+ q u_M^{A2} w^{A1} (1-h^{A2}) \}} = 0 \quad (C1)$$

$$\frac{\partial W}{\partial h^{A1}} = (u_H^{A1} + u_H^{B1}) d^{A1} f'(h^{A1}) - u_M^{A1} w^{A1} = 0 \quad (C2)$$

$$\frac{\partial W}{\partial h^{A2}} = u_H^{A2} d^{A2} f'(h^{A2}) - u_M^{A2} w^{A2} = 0 \quad (C3)$$

These three first-order conditions are admittedly longwinded, but so close parallels to conditions (A1) - (A3) that they should be self-explaining. The first-order conditions (A6), (A8), (A9) and (A11) still apply also within this framework.

Noting that here  $u_H^{A1} = u_H^{B1}$ , conditions (C2) and (C3) yield:

$$2 p^{A1} f'(h^{A1}) = p^{A2} f'(h^{A2}) \quad (C4)$$

As we have assumed  $u$  to be homothetic,  $p^{A1}$  and  $p^{A2}$  are functions only of the relative quantities of H-goods and M-goods consumed in the respective state.

Suppose that  $h^{A1} = h^{A2}$ . As the technology is the same in the two states, we know the total production will be the same in the two states. Further, we know that in state 2 transfers will be used so that A and B both enjoy half of the total production of both goods. In state 1 transfers will be used to give A and B each half of the total production of M-goods. Both of them, however, will enjoy the whole production of H-goods, as these are family specific public goods. These arguments imply:

$$\frac{H^{A1}}{M^{A1}} > \frac{H^{A2}}{M^{A2}} \quad \text{for } h^{A1} = h^{A2} \quad (C5)$$

Accordingly,

$$p^{A1} < p^{A2} \quad \text{for } h^{A1} = h^{A2} \quad (C6)$$

Unfortunately, combining conditions (C4) and (C6) we cannot

decide whether  $h^{A1} > h^{A2}$ . Two effects are at play. First, when both spouses can enjoy a homeproduced good when living together, the productivity in this line of work is in a sense doubled if the marriage continues in contrast to state 2. (This is represented by the number 2 in condition (C4).) This indicates  $h^{A1} > h^{A2}$ . But on the other hand, as one gets more homeproduced goods, the marginal value of H-goods falls. (This is captured by condition (C6).) If  $p$  is decreasing fast enough with respect to  $H/M$ , it could be optimal to spend less time at home in state 1 than in state 2. A couple living together can so to speak secure a "sufficient" level of homeproduced goods by working only a little each at home. Which of the two effects which will dominate, cannot be decided in general.

As it cannot be determined in general how  $h^{A1}$  and  $h^{A2}$  compare, of course the effect on  $h^{A0}$  of a rise in  $q$  also must be ambiguous. This should be clear from the discussion in Section 6.

**APPENDIX D**

In this appendix we will examine the directions of the transfers in state 1. As state 1 and state 2 are equal except for transfers being allowed in state 1 we can just as well examine the directions the transfers would have had in state 2, had they been allowed.

The relevant first-order conditions are (A1) - (A3) and (A6) - (A9). Slightly restated, condition (A3) is:

$$u_H^{A2} d^{A2} f'(h^{A2}) - u_M^{A2} w^{A2} = u_H^{B2} d^{B2} f'(h^{B2}) - u_M^{B2} w^{B2} = 0 \quad (D1)$$

This can be written as:

$$\frac{u_H^{A2} d^{A2} f'(h^{A2})}{u_H^{B2} d^{B2} f'(h^{B2})} = \frac{u_M^{A2} w^{A2}}{u_M^{B2} w^{B2}} \quad (D2)$$

Rearranging:

$$\frac{u_H^{A2}}{u_M^{A2}} f'(h^{A2}) \cdot \frac{w^{B2}/d^{B2}}{w^{A2}/d^{A2}} = \frac{u_H^{B2}}{u_M^{B2}} f'(h^{B2}) \quad (D3)$$

We define  $p^{A2}$ ,  $p^{B2}$  and  $K$  implicitly by restating (D2) as:

$$p^{A2} f'(h^{A2}) \cdot K = p^{B2} f'(h^{B2}) \quad (D4)$$

The assumption (15) about relative productivities implies that  $K < 1$ . Therefore, (D4) can be written as

$$p^{A2} f'(h^{A2}) > p^{B2} f'(h^{B2}) \quad (D5)$$

We have assumed that  $u$  is a homothetic function. This implies that  $p^{A2}$  is a (decreasing) function of the ratio  $H^{A2}/M^{A2}$ . This ratio, in turn, is increasing in  $h^{A2}$  (as there are no transfers



and  $h^{A2}$  therefore alone determines  $H^{A2}/M^{A2}$ ). Hence,  $p^{A2}$  is a decreasing function of  $h^{A2}$ . Correspondingly,  $p^{B2}$  is the same decreasing function of  $H^{B2}/M^{B2}$  as  $p^{A2}$  is of  $H^{A2}/M^{A2}$ . Also,  $p^{B2}$  is a decreasing function of  $h^{B2}$ , but this is not the same function as  $p^{A2}$  is of  $h^{A2}$ , as abilities differ.

From (D5) we cannot decide which of the spouses who works most at home. (D5) holds for  $h^{A2} < h^{B2}$ , but might also hold for  $h^{A2} > h^{B2}$ . This is intuitive: If B has a very high productivity in domestic work, this will be most utilized if much time is spent at this line of work. On the other hand, it also means that B can secure a given quantity of homeproduced goods by working very little at home. The functional forms and parameters of the problem decide which spouse spends the more time on domestic work.

However, (D5) will let us determine which spouse consumes more market goods and which one more homeproduced goods after a divorce. Suppose A consumes more homeproduced goods and B more market goods:

$$H^{A2} > H^{B2} \tag{D6}$$

$$M^{A2} < M^{B2} \tag{D7}$$

From the discussion above, we then know:

$$p^{A2} < p^{B2} \tag{D8}$$

Conditions (D5) and (D7) imply:

$$h^{A2} < h^{B2} \tag{D9}$$

(D9) says that B work longer hours at home. And as we know that B is also (absolutely) most productive in home production, we know:

$$H^{A2} < H^{B2} \tag{D10}$$

Correspondingly, we know:

$$M^{A2} > M^{B2} \tag{D11}$$

Conditions (D10) and (D11) clearly contradict the starting point (D6) and (D7). Hence we have proven:

$$H^{A2} < H^{B2}$$
$$M^{A2} > M^{B2} \tag{D12}$$

We know from the benchmark model that had transfers been used in state 2, we know that they would have been used to equate marginal utilities and also levels of consumption of the two goods. This implies that B will transfer H-goods to A, whereas A transfers M-goods to B.

$$a^1 < 0, \quad b^1 > 0 \tag{D13}$$

**APPENDIX E**

In this appendix we will look at the comparative statics exercise which leads to the expression (25). Let us start out by stating the three first-order conditions of the simplified problem stated in Section 5:

$$\begin{aligned} & \{ u_H^{B0} d^{B0} f'(h^{B0}) + D[ (1-q)u_H^{B1} d^{B1} f'(h^{B1}) + qu_H^{B2} d^{B2} f'(h^{B2}) ] \} \\ & - \{ u_M^{B0} w^{B0} + D[ (1-q)u_M^{B1} w^{B1} (1-h^{B1}) + qu_M^{B2} w^{B2} (1-h^{B2}) ] \} = 0 \end{aligned} \quad (E1)$$

$$p^{B1} d^B f'(h^{B1}) - w^{B1} = 0 \quad (E2)$$

$$p^{B2} d^B f'(h^{B2}) - w^{B2} = 0 \quad (E3)$$

By totally differentiating these equations, we obtain a system of this structural form:

$$\alpha_{11} \frac{dh^{B0}}{dq} + \alpha_{12} \frac{dh^{B1}}{dq} + \alpha_{13} \frac{dh^{B2}}{dq} = -\beta_1 \quad (E4)$$

$$\alpha_{21} \frac{dh^{B0}}{dq} + \alpha_{22} \frac{dh^{B1}}{dq} + \alpha_{23} \frac{dh^{B2}}{dq} = -\beta_2 \quad (E5)$$

$$\alpha_{31} \frac{dh^{B0}}{dq} + \alpha_{32} \frac{dh^{B1}}{dq} + \alpha_{33} \frac{dh^{B2}}{dq} = -\beta_3 \quad (E6)$$

(E4) is (E1) totally differentiated, and so on. All the coefficients  $\alpha$  and  $\beta$  are second-order derivatives of family welfare  $W$ . Some of these coefficients will be spelt out below.

Using Cramer's rule, we get:

$$\frac{dh^{B0}}{dq} = \frac{\begin{vmatrix} -\beta_1 & \alpha_{12} & \alpha_{13} \\ -\beta_2 & \alpha_{22} & \alpha_{23} \\ -\beta_3 & \alpha_{32} & \alpha_{33} \end{vmatrix}}{\begin{vmatrix} \alpha_{11} & \alpha_{12} & \alpha_{13} \\ \alpha_{12} & \alpha_{22} & \alpha_{23} \\ \alpha_{13} & \alpha_{32} & \alpha_{33} \end{vmatrix}} \quad (E7)$$

We have:

$$\beta_2 = \beta_3 = \alpha_{23} = \alpha_{32} = 0 \quad (E8)$$

Recognizing the determinant in the denominator as the Hessian of  $W$ , and using (E8), we obtain:

$$\text{sign } \frac{dh^{B0}}{dq} = \text{sign } \beta_1 \cdot \alpha_{22} \cdot \alpha_{33} \quad (E9)$$

We have:

$$\begin{aligned} \beta_1 &= [u_H^{B2} f(h^{B2}) u_H^{B1} f(h^{B1})] d^{B'} \\ &\quad - [u_M^{B2} (1-h^{B2}) - u_M^{B1} (1-h^{B1})] w^{B'} \end{aligned} \quad (E10)$$

$$\alpha_{22} = \frac{\partial p^{B1}}{\partial h^{B1}} d^{B1} f'(h^{B1}) + p^{B1} d^{B1} f''(h^{B1}) \quad (E11)$$

$$\alpha_{33} = \frac{\partial p^{B2}}{\partial h^{B2}} d^{B2} f'(h^{B2}) + p^{B2} d^{B2} f''(h^{B2}) \quad (E12)$$

Both  $\alpha_{22}$  and  $\alpha_{33}$  are negative. Note however that the homotheticity assumption about  $u$  is necessary to establish this unambiguous result. But now (E9) simplifies to

$$\text{sign } \frac{dh^{B0}}{dq} = \text{sign } \beta_1 \quad (E13)$$

This is precisely condition (25) in the main text.

Notes

- \* I thank Jan Erik Askildsen, Kristin Dale, Jan Haaland, Kåre Petter Hagen, Agnar Sandmo, Geir Helge Sjøtrø and Erling Steigum for comments and discussion.
- 1) Notions similar to mine, however, can be found in Pollak (1985) and Weiss and Willis (1985).
  - 2) In e. g. Becker, Landes and Michael (1977) and Landes (1978) models are developed where divorce occurs for reasons endogenous to the model.
  - 3) It would be simple enough to generalize the model to the case where the spouses' utilities carried unequal weight.
  - 4) I have implicitly assumed that there is symmetric information about productivities at home and in the labour market. Peters (1986) analyzes the consequences of asymmetric information.
  - 5) The concept of voice enforcement is similar, but not identical to Akerlof's concept "social customs" (Akerlof (1980)). In Akerlof's work a "social custom" is enforced by people's fear for loss of reputation - and people care for their reputations as such, not only about the possible consequences of such reputational losses. The difference is that Akerlof does not explain how a social custom comes about - in our setting it is an optimal contract that is to be enforced. Further, Akerlof pictures that people care for their reputation in general - here it is what your contract partner thinks of you that matters.
  - 6) A similar discussion about the role of credit market transactions as substitute for contracted transfers - in a somewhat different context - can be found in Gunning's (1984) comment on King (1982).

- 7) When homeproduced goods have perfect market substitutes, the family's problem reduces to determine for each spouse whether he or she can obtain most market goods by working in the market and buying them or by producing these goods himself (herself). If homeproduced goods were marketed, every individual could realize gains from specialization by trading with the market rather than with his or her spouse. See Dale Titlestad (1983).
  
- 8) Even though it does not fit completely in with the present model framework, the most important economies of scale from living together might be those stemming from the partly indivisibility of capital goods as cars, houses, refrigerators etc.

REFERENCES

- Akerlof, G.A (1980): "A Theory of Social Custom, of Which Unemployment May Be One Consequence", Quarterly Journal of Economics 95, 749-775.
- Becker, G.S. (1973): "A Theory of Marriage: Part I", Journal of Political Economy 81, 813-846.
- Becker, G.S. (1974): "A Theory of Marriage: Part II", Journal of Political Economy 82
- Becker, G.S. (1981): A Treatise on the Family, Cambridge, Massachusetts: Harvard University Press.
- Dale Titlestad, K: (1983): "A Model for the Allocation of Time Within the Family". Mimeo, University of Bergen.
- Gronau, R. (1973): "The Intrafamily Allocation of Time: The Value of Housewives' Time", American Economic Review 63, 634-681.
- Gronau, R. (1977): "Leisure, Home Production, and Work - The Theory of the Allocation of Time Revisited", Journal of Political Economy, 1099-1123.
- Gunning, J.P. (1984): "Marriage Law and Human Capital Investment: A Comment", Southern Economic Journal 51, 594-597.
- Hirschmann, A.O. (1970): Exit, Voice and Loyalty, Cambridge: Harvard University Press
- King, A.G., (1982): "Human Capital and the Risk of Divorce: An Asset in Search of a Property Right", Southern Economic Journal 49, 536-41.



- Klein, B. and K.B. Leffler (1981): "The Role of Market Forces in Assuming Contractual Performance", Journal of Political Economy 89, 615-641.
- Manser, M. and M. Brown (1980): "Marriage and Household Decision-Making: A Bargaining Analysis", International Economic Review 21, 31-44.
- McElroy, M.B. and M.J. Horney (1981): "Nash-Bargained Household Decisions: Toward a Generalization of the Theory of Demand", International Economic Review 22, 333-349.
- Peters, H.E. (1986): "Marriage and Divorce: Informational Constraints and Private Contracting", American Economic Review 76, 437-454.
- Pollak, R.A. (1985): "A Transaction Cost Approach to Families and Households", Journal of Economic Literature 23, 581-608.
- Shapiro, C. (1983): "Premiums for High Quality Products as Returns to Reputations", Quarterly Journal of Economics 98, 658-679.
- Stiglitz, J.E. and A. Weiss (1981): "Credit Rationing in Markets with Imperfect Information", American Economic Review 71, 393-410.
- Strand, J. (1983): "Structure and Efficiency of Reputational Labor Contracts", Stanford Workshop on Factor Markets Research Paper No. 46.
- Weiss, Y. and R.J. Willis (1985): "Children as Collective Goods and Divorce Settlements", Journal of Labor Economics 3, 268-292.

**Chapter 5:**

**PERSISTENT DISCRIMINATION WITH SOCIAL ABILITY AS A PRODUCTIVE FACTOR.**

September 1984

Revised June 1985

Reprinted from Journal of Institutional and Theoretical Economics/Zeitschrift für die gesamte Staatswissenschaft 143, 261-271, 1987.

## Persistent Discrimination with Social Ability as a Productive Factor

by

KJELL ERIK LOMMERUD\*

I

Traditionally, economists have assumed that discrimination reflects "tastes" against a definable group (e.g. BECKER [1957], ALCHIAN and KESSEL [1962], ARROW [1972], [1972a] and [1973].) Suppose that some employers had a "taste" against women – and therefore paid male workers more than equally able female workers. Suppose further that the employer in the short run is a monopsonist in the labor market. We then know that profit maximization would imply equal pay for equally productive workers (as long as their elasticities of the wage rate with respect to the employment level also are equal). Discrimination is a deviation from that policy: hence, an employer must pay the price of less profits for satisfying his taste for discrimination. Further, in the long run we would expect other employers – with less or no taste for discrimination – to hire the cheap female labor. In the end we would wind up with a situation where discriminating employers employed only males and non-discriminating employers all the females: the labor market would be segregated, but no discriminatory wage differentials would exist. (However, AKERLOF [1980] has pictured that the social stigmatization from breaking the "social custom" of discrimination can outweigh the pecuniary benefits of non-discriminatory behavior, such that discrimination persists.)

It could also be that some of the employees of a firm had a "taste" for not working together with a certain group: for instance, white workers could refuse to work together with black workers. But again, the long run implication of this would be segregated production, not discriminatory wage differentials.

Further, discrimination could be rooted in the customers having "tastes" against certain groups producing certain goods. Such discrimination could

---

\* I thank Jan Erik Askildsen, Kåre Petter Hagen, Astrid Ingeborg Kleppe, Agnar Sandmo and an anonymous referee for comments. This note is a much revised version of Discussion Paper 12/84, Norwegian School of Economics and Business Administration.

persist also under perfect competition, but is likely to be most important in lines of work with direct customer contact.

We see that under perfect competition, discrimination will have a tendency not to be sustained. But, as ARROW [1973] notes: "Since in fact racial discrimination has survived for a long time, we must assume that the model ... must have some limitation."

One approach to the stability of discrimination is the modelling of divide-and-rule policies. Some writers (REICH [1980], ROEMER [1979] and GINTIS [1976]) picture that discrimination between equally productive workers can reduce the bargaining power of the workers. These models, however, do not describe perfect competition in all markets, as that would leave no room for a concept as "bargaining power". BOWLES [1985] has suggested that divide-and-rule discrimination can render surveillance of worker productivity cheaper.

Also, this paper deals with persistent discrimination – in a competitive setting where employers have no "taste" for discrimination; neither is discrimination due to variables such as race or sex being proxies for technical ability.<sup>1</sup> The basic idea is that the ability to adapt socially to the relevant surroundings is an important productive factor for some jobs. For instance: a firm wants to employ a manager. Assume that in order to fit in with the "managerial culture", one has to be a white, male conservative. A non-conformist manager must be expected to get less informal information, business proposals etc. from "business friends". Let us focus on the role of informal exchange of information. Suppose that all managers are of one "kind", and that people tend to like others of their own kind. Suppose further that people exchange bits of information when they mix with people they like. A prospect manager of a different "kind" might be excluded from this network of information exchange. He would then have to obtain information at a cost from other sources, or perhaps have to do without this information. It would take a lot of extra "technical ability" in a strict sense to outweigh this disadvantage. Discrimination occurs in the model in the sense that people with the same "technical" ability, but who differ with respect to "social" ability, are treated differently. The concepts "technical" and "social" ability will be defined more precisely shortly. Note that when we take as given facts who likes who and that friends exchange information informally, no one stands to gain from breaking this discrimination. A main argument of this paper is that there is scope for government intervention against this type of discrimination. Private agents must take the social requirements for filling a job as given. However, we would expect non-marginal employment decisions, as for instance a large scale hiring of female managers, to change these social norms. Such non-marginal employment decisions might be efficiency improving if they lead to a better utilization of the technical abilities in the population.

---

<sup>1</sup> For instance, see SPENCE [1974].

Let us develop these ideas more formally. Consider an economy where everybody delivers inelastically one unit of labor. This one unit of labor corresponds to a certain number of efficiency units of labor, depending on an employee's "technical ability" and "social ability"<sup>2, 3</sup> for doing a job. For simplicity, assume there are two types of jobs, "managerial" and "unskilled". Some social abilities are productive for people holding a managerial job, but would have been less productive for people who work as unskilled. For other social abilities, it is the other way around. Formally:

$$(1 a) \quad l_{iu} = l_{iu}(x_i, y_i) \quad \text{for all } i.$$

$$(1 b) \quad l_{im} = l_{im}(x_i, y_i) \quad \text{for all } i.$$

Here  $l_{iu}$  stands for the efficiency units of labor which agent  $i$  would supply if employed as unskilled, whereas  $l_{im}$  denotes the efficiency units of labor supplied if he is employed as a manager. Further,  $x_i$  and  $y_i$  – which may be thought of as vectors – denote agent  $i$ 's technical and social ability, respectively.

It is important to note that the functions  $l_{iu}$  and  $l_{im}$  are postulated to be independent of how jobs are assembled into plants and firms. The productivity of a certain social talent in a specific type of job is independent of which firm one works in. If this shall hold true it must mean that "social ability" to a large extent refers to the ability to adapt socially to surroundings *external* to the firm. If "social ability" only meant the ability to adapt to co-employees of the same firm, we would basically be back to a model where employees have "tastes" for whom they prefer to work with. As noted above this implies segregated production, but not inequal pay.<sup>4</sup> Of course, different jobs differ greatly with respect to the importance of firm-external contact. "Social ability" may therefore be of much larger importance as a productive factor in "managerial" jobs than in "unskilled".

For expositional simplicity, I assume that the efficiency units of labor stemming from managerial and unskilled work are perfect substitutes. Further, we

<sup>2</sup> Terms as "social productivity" are often used in a quite different meaning, namely that certain social groups develop counter-productive life-styles and attitudes. See CAIN [1976] and the references cited therein.

<sup>3</sup> In this simple model we assume that people are born with certain social abilities. The analysis would not be much altered if we allowed for the possibility that social abilities could be acquired, but that different people had different costs of obtaining them. To begin with, we also assume that technical abilities are given facts, but we relax this assumption below.

<sup>4</sup> That people have tastes for whom they would like to work with, is an example of "hedonic clubs". See for instance DRÈZE and GREENBERG [1980]. Their starting point is that people's utilities from working in a coalition depend on who the other members in the coalition are and how big the coalition is. Drèze and Greenberg then investigate the optimality and stability of coalitions within this context.

assume that workers sell their efficiency units of labor at the wage  $w$ . Hence, if worker utility has wage income as its only argument, a worker will choose managerial occupation if  $l_{im} > l_{iu}$ , and unskilled if the inequality is reversed.<sup>5</sup> This implies:

$$(2) \quad l_i = \max [l_{iu}, l_{im}],$$

where  $l_i$  is the actual supply of efficiency units of labor from worker  $i$ .

The firm is assumed to maximize profit. If non-labor production factors are held constant, the firm's problem can be written as:

$$(3) \quad \text{Max}_l \pi = pf(l) - wl.$$

Here  $\pi$  denotes profit,  $p$  the price of the output good,  $l$  efficiency units of labor used as input and  $f$  the production function, assumed to meet standard conditions. The first order condition of the problem is of course:

$$(4) \quad pf'(l) = w.$$

The only noteworthy fact about this very simple maximization problem is that all that counts for the firm is  $l$ . Whether it is social or technical ability that brings about  $l$  is of no importance. Therefore, in this model employees are paid according to their productivity – but this productivity may just as well be caused by social as technical ability. And people are sorted to different types of jobs according to relative ability, but again their ability might just as well be social as technical.

An individual worker must take the structure of how different social abilities fit with different jobs as given. Therefore it may for instance be optimal for a worker with a high degree of technical qualification for managerial work, but who lacks in social qualifications, to choose unskilled work.

As already hinted at, this situation is only efficient if one takes the social requirements for filling a job as given. The government, which can bring about non-marginal changes in the labor market, e.g. a large flux of women into managerial occupations, might be expected to be able to influence these social requirements. To do this might be efficiency improving as it might lead to a better utilization of the technical abilities in the population.

---

<sup>5</sup> In the present formulation everybody can choose whether to work as a "manager" or as "unskilled". This is because everybody gets paid only for the efficiency units of labor they actually deliver in a specific occupation. This would imply very low – or even negative – wages for people with low productivity for instance as managers. In practice, we would not expect wages to be this flexible. Therefore some people would not find occupation as managers even if they were prepared to accept negative wages.

Let us try to capture this formally. We must start by rewriting (1 a) and (1 b):

$$(5a) \quad l_{iu} = l_{iu}(x_i, g_u(y_i, z_u(y_u, z_u^o))) \quad \text{for all } i.$$

$$(5b) \quad l_{im} = l_{im}(x_i, g_m(y_i, z_m(y_m, z_m^o))) \quad \text{for all } i.$$

$z_u$  is the "social norm" which decides how productive a certain social ability  $y_i$  will be in unskilled work.  $z_u$ , in turn, is a function of  $y_u$ , the vector of social abilities of those occupied in unskilled work, and "all other factors",  $z_u^o$ .  $g_u$  is a functional symbol, denoting "efficiency units" of social abilities.  $g_m$ ,  $z_m$ ,  $y_m$  and  $z_m^o$  have similar interpretation.

It is both beyond the scope of this paper and the competence of its author to try to pinpoint exactly what social norms apply to what jobs, how such norms have come about and how they change.

Assume that society is concerned only with efficiency, given by:

$$(6) \quad W = \sum_i l_i.$$

By inserting (5 a) and (5 b) into (2), and (2) into (6), we obtain:

$$(7) \quad W = \sum_i \max [l_{iu}(x_i, g_u(y_i, z_u(y_u, z_u^o))), \\ l_{im}(x_i, g_m(y_i, z_m(y_m, z_m^o)))].$$

Let us calculate two quantities,  $l_{iu}^s$  and  $l_{im}^s$ , given by.

$$(8a) \quad l_{iu}^s = l_{iu}^s(x_i, g_u(y_i, \bar{z}_u)) + \sum_n \Delta l_n^u \quad \text{for all } i.$$

$$(8b) \quad l_{im}^s = l_{im}^s(x_i, g_m(y_i, \bar{z}_m)) + \sum_n \Delta l_n^m \quad \text{for all } i.$$

$l_{iu}^s$  and  $l_{im}^s$  can be interpreted as workers  $i$ 's productivity in the two lines of work, seen from the social planner's viewpoint.  $l_i^s$  will denote the largest of these two quantities.  $\bar{z}_u$  and  $\bar{z}_m$  are the social norms as they would be when all workers are employed as they will be in optimum, except for the worker  $i$  which is not yet employed. The first terms of (8 a) and (8 b) then are the productivities of worker  $i$  in the two lines of work – if we assume that his employment decision will not affect the social norms.  $\Delta l_n^u$  is the productivity change for worker  $n$  caused by the change in the social norm  $z_u$ , following worker  $i$  taking up work as unskilled.  $\Delta l_n^m$  is interpreted similarly. The second terms of (8 a) and (8 b) therefore capture the sum of the productivity changes for all workers (including worker  $i$  himself) caused by the changes in the social norms. When a social norm changes, people's productivity can change because they become more or less productive in the type of work they would have chosen also given the initial

social norm – but it can also be that people choose another type of job than they else would have done. Note that the second terms in (8 a) and (8 b) may be of either sign. Also, note that as  $l_{iu}^s$  and  $l_{im}^s$  for one individual worker are calculated given that all other workers are employed optimally, these quantities must be calculated simultaneously for all workers. In practice, this would of course be a very messy task. When a worker chooses whether to work as unskilled or as manager, he will affect the social norms  $z_u$  and  $z_m$ . For the individual worker, the effect on his own productivity in the two lines of work will be negligible. His choice also affects the productivity of all the other workers – but this is not compensated for through a market. Hence, this is an example of a (positive or negative) externality.

And even if the productivity change for one individual employee induced by a small change in social norms may be negligible, the sum of these effects for all workers may be of considerable importance.

Now, the welfare function (7) will be maximized if a worker  $i$  is assigned to unskilled or managerial work according to a comparison of  $l_{iu}^s$  and  $l_{im}^s$ , instead of  $l_{iu}$  and  $l_{im}$ . However, so far we have not been very precise about *how* the government should intervene in the labor market. Here, as in many other contexts, the government has the choice between using taxes/subsidies or quotas. I believe direct assignment to jobs/quotas might be seen as a too serious interference with the freedom of contracting to be feasible in democratic societies. Also, the use of quotas could lead to further stigmatization of women/blacks and thereby to these groups being even more shut out from informal networks of information exchange. An alternative to quotas is to use taxes and subsidies to e.g. increase the cost of hiring white, male managers and to reduce the cost of managers from other groups. (In this simple, illustrative model labor supply is inelastic. Of course, in practice those who design a tax system should concern themselves not only with people's occupational choice, but also with labor supply.) The American system of demanding "affirmative action" against discrimination from federal contractors can – even though it might look like a system of quotas – be interpreted as a tax on the labor income of white, male workers (LEONARDS [1984]). From an economic theorist's viewpoint there seems to be no reason to limit such taxation to federal contractors, but political feasibility considerations might account for this limitation.

I think it is important to remember that the type of discrimination we are talking about here is important only in occupations where a firm-external network of contacts is important. There are potential problems of using the traditional measures against discrimination in such a setting. First, the relevant occupational categories might be so small or so ill defined that a system of quotas or of taxes/subsidies might be very difficult to implement. Second, as hinted at, the government intervention *itself* might create unwanted stigmatization. Personally I believe it would be a sound policy proposal to try to favor – by quotas or by incentives – e.g. women in those educational paths which lead to managerial positions. This should be implementable enough, and as the



intervention does not take place in the labor market itself, the risk of creating further stigmatization should be reduced.

## II

A complication arises if we consider that technical skills can be acquired by education. Suppose that workers are characterized by their "primitive technical abilities" and their costs of acquiring additional skills. The workers now face a two-step maximization problem. First they must decide on the level of education given unskilled and given managerial work. The levels of education will be chosen in order to equate the marginal wage increase from education with the marginal cost of education, except for those particular skills where we get corner solutions.

The second step in a worker's utility maximization problem is to calculate the utility he will obtain as unskilled or as manager, given optimal education. Naturally, he will choose the line of occupation which yields the highest utility.

Now, what does the situation look like from the viewpoint of a social planner which is only concerned about efficiency? It is here important to delineate between two different scenarios.

We could imagine (unrealistically) that the planner arrives at the scene *before* any committing educational choices have been undertaken. This situation would be a rather straightforward extension of the model above. In an alternative (realistic) scenario we could picture that people's occupational and educational choices are made before the social planner enters the picture, and that people do not expect any future interference from a planner. We will assume that the planner thereafter can assign people to further education and order them to change job (although in most economies incentive schemes would have to be used.)

When a planner enters the scene after the workers have acquired their "first-round" education, the cost of this education is sunk. If some people lack the relevant social ability for a job, they will also have less incentives for acquiring technical skills for such jobs. In this situation it might be that it would not be optimal to make a worker change occupation after he has undertaken education, even though this would have been optimal had the planner intervened before any irreversible educational choices had been made.

A policy implication could be the following one: As already mentioned, people who lack the relevant social ability for a job, will on average not acquire very much education relevant for this job. A "crash programme", as the sudden interchange of a group of managers with a group of unskilled workers, could subsequently lead to an important immediate efficiency loss, due to large costs of reschooling. A more gradual policy, aiming at influencing the occupational choices of new entrants to the labor market (who not yet have made irreversible

educational choices), could be a much cheaper way of changing for instance the social norms for managerial behavior.

### III

My main point has been that in some jobs social adaption and external contacts are important productive factors. If someone is excluded from such a network of contacts due to sex, race, beliefs, religion etc. this might be interpreted as discrimination. And as stressed, such discrimination will not be eroded by competitive pressures. By way of conclusion I would like to focus on a perhaps less obvious form of such discrimination than against women or blacks. More precisely, I would like to point out how the present line of reasoning can shed some light on a discussion of the relative efficiency of capital-managed and labor-managed firms in PUTTERMAN [1984].

Several writers (ALCHIAN and DEMSETZ [1972], WILLIAMSON [1980], JENSEN and MECKLING [1979]) have claimed that the superiority of the capital-managed over the labor-managed firm can be established only by noting that so few labor-managed firms exist in economies where both organizational forms are allowed. The argument is that a less productive firm must pay its workers less if it is to survive. By revealed preference we know that this decrease in pay for most workers outweigh the benefits of worker control of the firm. If not, workers would organize labor-managed firms instead of taking employment in capitalist firms. Another version of this argument is to say that a less efficient firm which has to reward its workers competitively, will not survive in a market economy. This form of the argument can be called "economic Darwinism", a term dating back at least to FRIEDMAN [1953].

PUTTERMAN [1984] argues that the arguments above are too simple: success in the marketplace does not directly imply economic efficiency. Putterman's most important argument is that the Darwinism of the marketplace single out *locally* efficient firms, which not necessarily are *globally* efficient. "... the economic survivability operator is not a *global* optimality operator, because it selects for viability within an existing environment, including its system of laws and property right" (PUTTERMAN [1984], p. 186).

Of course, Putterman is not the first one to have noted that a market equilibrium may be locally, but not globally efficient. Much work on incomplete markets for "package deals" either deals explicitly with this issue, or can be interpreted as dealing with it (e.g. DRÈZE [1974], GROSSMAN [1977], DRÈZE and HAGEN [1978], STIGLITZ [1982]). (A "package" can for instance be a package of contingent claims called a share, a package of quality characteristics called a consumption good, or a package of working conditions called a job). Other examples of literature making a point of the distinction between local and global efficiency are BARAN and SWEEZY [1966] and LINDBECK [1977].

I offer "discrimination" in the sense of this note as an example of how the market outcome may be locally, but not globally efficient - and which might

be of relevance for the non-viability of labor-managed firms in capitalist economies. Prejudice and discrimination in the sense above may not only hurt women or minorities, but also managers of labor-managed firms. The very fact that a manager is a hired manager in a labor-directed firm might make him less able to interact socially with the rest of the business community – and this could be very damaging<sup>6</sup>. The chance of survival for a labor-managed firm might increase if the firm were a member of a group of labor-managed firms, as this would make contacts external to the group less important. The apparent success of the Mondragon group of labor-directed firms might be an example of this. Also, it might be that if the workers of a labor-directed firm all were highly qualified “white collar” workers, rather than blue collar workers, labor management might carry less social stigma. And casual empiricism seems to suggest that organizational forms that look like labor management (but not necessarily are named as such) are more common in small, “white collar” firms.

But of course it is difficult to ascertain whether managerial norms in fact do stigmatize leaders of labor-managed firms (as “suspect communists”). Whether the outlined possibility is of practical relevance, is left to the reader’s judgement.

In a long survey on the non-viability question, FANNING and MCCARTHY [1983] claim to present “the more substantial” of the hypotheses on why so few labor-controlled firms exist in capitalist economies. A “substantial” hypothesis is defined as one that is not simply based on prejudice. I have tried here to show that prejudice and discrimination can be given an interpretation which make these phenomena persistent also under perfect competition. I feel therefore that it is somewhat out of place to call discrimination an insubstantial phenomenon – when the non-viability of labor-managed firms is discussed, and in other contexts.

#### *Summary*

We can think of an employee’s productivity as depending on his or her “technical” and “social” ability. By social ability is meant the ability to adapt to a firm’s external network of contacts, and to extract information, business proposals etc. from this network. Individual agents must take the social requirements for filling different jobs as given, whereas a government might be able to influence these. Government intervention might be efficiency improving, as there is no reason to expect the given existing set of such social requirements to lead to the best utilization of the technical abilities in the population.

---

<sup>6</sup> Of course, to be a manager in a labor-managed firm is not an inherent trait of a person, as race and sex are. So when people are socially stigmatized for being hired managers in labor-managed firms, this could lead to adverse selection problems not accounted for here.

*Zusammenfassung*

*Anhaltende Diskriminierung mit sozialer Fähigkeit als Produktionsfaktor*

Man kann sich vorstellen, daß die Produktivität eines Arbeitnehmers von dessen technischer und sozialer Fähigkeit abhängt. Mit sozialer Fähigkeit ist die Fähigkeit gemeint, sich auf das externe Kontaktnetz einer Unternehmung einzustellen und sich über dieses Netz Informationen, Angebote usw. zu verschaffen. Individuelle Handlungsträger müssen die sozialen Anforderungen für die Ausübung verschiedener Tätigkeiten als gegeben hinnehmen, wohingegen der Staat in der Lage sein könnte, diese zu beeinflussen. Staatliche Intervention könnte evtl. die Effizienz steigern, da nicht zu erwarten ist, daß die tatsächlich gegebenen sozialen Anforderungen die beste Nutzung der technischen Fähigkeiten der Bevölkerung bewirken.

*References*

- ALCHIAN, A. A. and DEMSETZ, H. [1972], "Production, Information Costs, and Economic Organization", *American Economic Review*, 62, 777-795.
- ALCHIAN, A. A. and KESSEL, R. A. [1962], "Competition, Monopoly, and the Pursuit of Pecuniary Gain", pp. 156-175, in: *Aspects of Labor Economics: A Conference of the Universities - National Bureau for Economic Research Committee*, Princeton, N.J.
- AKERLOF, G. A. [1980], "A Theory of Social Custom, of Which Unemployment May Be One Consequence", *Quarterly Journal of Economics*, 94, 749-775.
- ARROW, K. [1972], "Models of Job Discrimination", pp. 83-102, in: A. H. Pascal (ed.), *Racial Discrimination in Economic Life*, Lexington, Mass.
- [1972a], "Some Mathematical Models of Race Discrimination in the Labor Market", pp. 187-204, in: A. H. Pascal (ed.), *Racial Discrimination in Economic Life*, Lexington, Mass.
- [1973], "The Theory of Discrimination", pp. 3-33, in: Ashenfelter, O. and Rees, A. (eds.), *Discrimination in Labor Markets*, Princeton, N.J.
- BARAN, P. A. and SWEEZY, P. M. [1966], *Monopoly Capital*, Monthly Review Press.
- BECKER, G. S. [1957], *The Economics of Discrimination*, Chicago.
- BOWLES, S. [1985], "The Production Process in a Competitive Economy: Walrasian, Neo-Hobbesian, and Marxian Models", *American Economic Review*, 75, 16-36.
- CAIN, G. G. [1976], "The Challenge of Segmented Labor Market Theories to Orthodox Theory: A Survey", *Journal of Economic Literature*, 14, 1215-1257.
- DRÈZE, J. H. [1974], "Investment Under Private Ownership: Optimality, Equilibrium and Stability", pp. 129-166, in: Drèze (ed.), *Allocation under Uncertainty: Equilibrium and Optimality*, London and Basingstoke.
- and GREENBERG, J. [1980], "Hedonic Coalitions: Optimality and Stability", *Econometrica*, 48, 987-1003.
- DRÈZE, J. H. and HAGEN, K. P. [1978], "Choice of Product Quality: Equilibrium and Efficiency", *Econometrica*, 46, 493-513.
- FANNING, C. and MCCARTHY, T. [1983], "Hypotheses Concerning the Non-Viability of Labor-Directed Firms in Capitalist Economies", *Economic Analysis and Workers' Management*, 17, 123-154.
- FRIEDMAN, M. [1953], *Essays in Positive Economics*, Chicago.
- GINTIS, H. [1976], "The Nature of the Labor Exchange", *Review of Radical Political Economics*, 8, 36-54.

- GROSSMAN, S. J. [1977], "A Characterization of the Optimality of Equilibrium in Incomplete Markets", *Journal of Economic Theory*, 15, 1-15.
- JENSEN, M. C. and MECKLING, W. H. [1979], "Theory of the Firm: Managerial Behavior, Agency Costs and Ownership Structure", *Journal of Financial Economics*, 3, 305-360.
- LEONARDS, J. S. [1984], "The Impact of Affirmative Action on Employment", *Journal of Labor Economics*, 2, 439-463.
- LINDBECK, A. [1977], *The Political Economy of the New Left*, 2nd ed., New York.
- PUTTERMAN, L. [1984], "On Some Recent Explanations of Why Capital Hires Labor", *Economic Inquiry*, 22, 171-187.
- REICH, M. [1980], *Racial Inequality*, Princeton.
- ROEMER, J. [1979], "Divide and Conquer: Micro Foundations of the Marxian Theory of Discrimination", *Bell Journal of Economics*, 10, 695-705.
- SPENCE, A. M. [1974], *Market Signaling*, Cambridge, Mass.
- STIGLITZ, J. E. [1982], "The Inefficiency of the Stock Market Equilibrium", *Review of Economic Studies*, 49, 241-261.

*Kjell Erik Lommerud*  
*The Norwegian School of*  
*Economics and Business Administration*  
*Institute of Economics*  
*Helleveien 30*  
*N-5035 Bergen-Sandviken*  
*Norway*

**Chapter 6:**

**EDUCATIONAL SUBSIDIES WHEN RELATIVE INCOME MATTERS**

The germ of this paper's main idea can also be found in "Educational Subsidies when Workers are not Paid their Marginal Product", Discussion Paper 13/84, Norwegian School of Economics and Business Administration, October 1984.

**EDUCATIONAL SUBSIDIES WHEN  
RELATIVE INCOME MATTERS**

**ABSTRACT**

This paper postulates that relative income, or "status", enters people's utility functions, and that the total value of status in society is largest when inequality is minimized. I propose that a natural measure for "status" is relative labour income after tax in the years after a concluded education. Moreover, I suggest that educational decisions are undertaken in order to maximize lifetime utility. This leads to a situation where it might be optimal to tax away some of the differences in people's labour income - and to use educational subsidies to restore people's incentives to undertake education.

## EDUCATIONAL SUBSIDIES WHEN RELATIVE INCOME MATTERS

### 1. Introduction

Even though every country I know of subsidizes higher education, it is not completely clear why this should be so. It might be argued that the recipients of these subsidies are among the most resourceful people in the population and the last ones that ought to be subsidized. A counterargument could be that there are external effects from education. But even if some activities lead to more external effects than others, it is not obvious that the total value of these externalities necessarily increases with the total level of education. Another line of argument is that credit markets might be incomplete (see e.g. Kodde and Ritzen (1985)), e.g. because of the problems of putting up human capital as collateral. This could lead to underinvestment in education, especially among children from poor families which presumably would have less access to credit. Moreover, those who do choose to study will not be able to smooth their income over the life-cycle to the degree they would want. But this line of thought seems to point to loans for schooling, rather than subsidies. Another argument for subsidizing higher education is that the risk connected with human capital investments is very difficult to diversify. For the individual student the risk of not being able to complete his or her education successfully, or the risk of the market value of a certain skill dropping drastically is considerable. For society as a whole it only matters how many people that on average manage to increase their productivity through education. But this leaves unanswered the question why the government can handle the informational problems which rule out private insurance any better than the private sector. Moreover, distributional aspects of educational subsidies may not be as clear cut as one might think. Johnson (1984) has pointed out that with complementarity between labour with low and high skills, the relatively poor realize a portion of the gains from the more able receiving higher education.



I will here suggest an additional justification for educational subsidies. I will take as my starting point that people care for their relative income. This is an old idea in economics (e.g., see Veblen (1899), Duesenberry (1949) and Leibenstein (1950)), although it has never played an important role in the mainstream of the field. Recent work on the importance of relative income or "status" include Hirsch (1976), Boskin and Sheshinski (1978), Layard (1980), Sen (1983), Oswald (1983), Frank (1984a, 1984b, 1985a, 1985b), Orosel (1985), James (1987) and Ng (1987). Trade union models where relative wages matter include Oswald (1979) and Gylfason and Lindbeck (1984).

However, what exactly is it that yields status? It might not only be relative income, but for instance education per se. And what measure of relative income should be used? Income before or after taxes? Annual income or lifetime income? I will propose that a natural measure might be labour income after tax in the years after education is concluded. A fuller discussion of this suggestion will follow below. The choice of a measure of status is the present paper's main departure from Boskin and Sheshinski's (1978) analysis of optimal redistributive taxation when status matters. Further, I suggest that educational decisions are undertaken in order to maximize lifetime utility. This leads to a situation where it might be optimal for the government to tax away some of the differences in people's labour income - and to use educational subsidies to restore people's incentives to undertake education. I point out that one would get similar results in any other situation where inequality of per period income after tax for some reason were counterproductive, and where educational choices were guided by lifetime income considerations.

In section 2 I will illustrate this idea in a rather simple model. The most restrictive assumption in this model is constant labour supply, which is loosened in section 3. Section 4 contains some concluding remarks.

## 2. A Simple Model

Consider an economy with  $N$  individuals, who are identical, except that they have different costs of obtaining education. The model is a two period one. In the first period people have an exogenous income  $\bar{w}$ , which is untaxed.<sup>1)</sup> In this period they obtain education. The cost of obtaining units of education,  $e$ , is a function  $c_n(e_n)^2$ , where  $c'_n > 0$  and  $c''_n > 0$ . Subscript  $n$  denotes individual  $n$ . In the second period, people work. Every worker supplies  $\bar{L}$  units of labour inelastically. How many efficiency units this labour supply represents, depends on how much education has been undertaken. We assume, for simplicity, that  $y_n = e_n \bar{L}$  (where  $y_n$  denotes efficiency units of labour). The workers sell their efficiency units of labour at the price  $w$  to competitive, constant-returns-to-scale firms which use labour as their only input.

Let us begin by studying the workers' choice of education. We will assume that people maximize utility, and that they have a common (cardinal) utility function,  $U$ . As people have different costs of obtaining education, reflecting differing abilities, they will still have different utility levels. We write the utility of individual  $n$ :

$$U_n = u_{n1} + u_{n2} \quad (1)$$

where

$$u_{n1} = \bar{w} - (1-s) c_n(e_n)$$

$$u_{n2} = (1-t) w e_n \bar{L} - a + g(S_n)$$

$$S_n = (1-t) w e_n \bar{L} - a - \frac{1}{N} \{ \sum_n ((1-t) w e_n \bar{L}) - aN \}$$

$u_{n1}, u_{n2}$  = utility in periods 1 and 2

$s$  = a subsidy on the costs of education

$t$  = a tax on labour income

$a$  = a lump-sum tax element

$g$  = utility derived from relative income, or "status". We assume  
 $g' > 0$ ,  $g'' < 0$ .

$S_n$  = individual  $n$ 's status.

Some comments are in order on the implications of (1). The simplification that utility is linear in net income and additively separable in net income and status is used to ease the exposition, and is not material for the main insights. As an index for status I have chosen the difference between a worker's after tax income and the average after tax income. This particular measure of relative income will be convenient, as it implies that the lump-sum tax element,  $a$ , has no influence on status. Note also that we have presupposed that status depends on income relative to average income in the economy as a whole. This contrasts e.g. with Robert Frank's work on status, mentioned above. Frank mostly studies the case where it is relative wages within a firm which constitute the basis for status. We would expect the truth to lie somewhere in between. Those "others" who form the basis for interpersonal comparisons for a given individual are likely to be determined by spatial proximity and degree of interaction (Festinger (1954), Williams (1975)). My assumption has two important implications relative to Frank's setting. First, firms will here have no incentive to try to compress wage schemes - as an individual firm presumably has a negligible effect on the wage dispersion in the whole economy. (Such a wage scheme compression within a firm is at the center of interest in Frank (1984a).) Second, a firm trying to compress wage schemes could experience that those workers paid below their marginal product moved out. For a state, emigration due to tax pressure certainly will be a much lesser problem, and will here be abstracted from.

Perhaps the most important assumption behind (1) is that it is only differences in after tax income in the second period which determines status. Low income net of educational expenses does

not give students low status in period one. I think this is realistic. It is not irrelevant under which circumstances a low income is earned. In fact, as education per se might yield status (which is not modelled here), one might imagine that students have a higher status than non-students. As already mentioned, this assumption is a driving force behind my argument, and represents the main deviation from the analysis in Boskin and Sheshinski.

Note that we have restricted our attention to linear tax and subsidy schemes. I will further restrict the choices of  $s$  and  $t$  as follows:

$$-1 \leq s, t \leq 1 \quad (2)$$

Remember that we have assumed that  $g' > 0$  and  $g'' < 0$ . I.e., there is diminishing marginal utility of status. Also,  $g(0) = 0$ . The concavity assumption is crucial for the results - but it will be easy to see how results must be altered if there is constant or increasing marginal utility of status. A further tacit assumption behind (1) is that the workers cannot move income between period 1 and 2, other than investing in education, i.e., there are no saving/borrowing possibilities. This is also crucial, which will be discussed below.

A person chooses  $e_n$  in order to maximize (1). The first-order condition of this problem is:

$$(1-t) w \bar{L} \left( 1 + g' \left( S_n \right)^{\frac{N-1}{N}} \right) = (1-s) c'_n(e_n) \quad \forall n \quad (3)$$

I.e., in optimum the marginal gain from education should equal marginal cost. Note that we have tacitly assumed that workers are entertaining Nash conjectures vis-à-vis each others' choice of  $e_n$ . If we look at the case where status yield no utility, i.e., where  $g = 0$  always, (3) is reduced to the familiar condition:

$$(1-t) w \bar{L} = (1-s) c'_n(e_n) \quad \forall n \quad (4)$$

Let us now turn to the government's problem. We will here assume that the government is utilitarian, wishing to maximize the un-weighted sum of individual utilities:

$$\text{Max } \sum_n U_n \quad (5)$$

The government faces a public budget constraint: Net tax revenue must equal  $\bar{T}$ . Formally,

$$\sum_n (t w e_n \bar{L} - s c_n(e_n)) + a N = \bar{T} \quad (6)$$

The government's problem will now be to maximize welfare given the public budget constraint, given the restriction that  $|s|$  and  $|t|$  are less or equal to one, and given that workers choose their level of education as described. It turns out that for this very simple form of the problem, the easiest way to arrive at a solution will be to use intuition to suggest a solution, and then to show that this comes arbitrarily close to being first best optimal. A more formal analysis of both of this and an extended version of the problem will follow in the next section.

Let us first look at the case where status yields no utility. Clearly, the lump-sum tax element can be used to take care of the public budget constraint (6). Taxation will not distort labour supply decisions, since labour supply is taken to be in-elastic. It therefore remains for the government to give the workers proper incentives for acquiring education. In the case with no taxes or subsidies, a representative worker's first-order condition w.r.t. choice of educational length, will be

$$w \bar{L} = c'_n(e_n) \quad \forall n \quad (7)$$

By comparing (4) and (7) we see that there will be no distortion of the choice of education if

$$s = t$$

(8)

However, the joint level of  $s$  and  $t$  is indeterminate. (Strictly speaking, it should not be 1, because then a worker would be indifferent as regards his length of education.)

This means that in the case where only absolute income matters educational subsidies have no independent role to play. They are just an indirect way of lowering the tax on income. When the joint level of  $s = t$  is indeterminate, it might as well be zero, especially if there were some costs (not modelled) incurred when collecting taxes on labour income and handing the money out again as educational subsidies.

Let us now turn to the more general case where status might yield positive private utility. Again, the lump-sum tax element  $a$  can be used to take care of the public budget constraint. Now, the government has two additional choice variables,  $s$  and  $t$ . It wants to use these to balance the wish to minimize distortions of educational choices against the wish to maximize the total value of status in society.

The sum of utility in society derived from status will be largest when there is no inequality at all of after tax labour income in period 2. This follows from the fact that the marginal utility of status is diminishing. (Had we assumed  $g$  to be convex, social welfare would be largest if inequality was maximized.) A situation of no inequality arises when  $t = 1$ , i.e. when the authorities confiscate all wage income, and thereafter hands out an equal amount,  $a$ , to everyone. We approach this first best optimal situation when  $t \rightarrow 1^-$ .

Let us now look at the case where the distortions of educational choices are minimized. In this model there can be both over- and underinvestment in education, depending on the values of  $s$  and  $t$ . We define:

$$A_n \equiv (1-t) w\bar{L} \left( 1 + g'(S_n) \frac{N-1}{N} \right) / (1-s) c'_n(e_n) \quad (9)$$

$$B_n \equiv w\bar{L} / c'_n(e_n) \quad (10)$$

$A_n$  is the fraction of marginal benefits from education to marginal costs.  $B_n$  expresses what this fraction would have been in the absence of taxes and subsidies, and with all external effects internalized.<sup>3)</sup>

When  $A_n > B_n$ , we have overinvestment in education,  $A_n < B_n$  implies underinvestment, whereas  $A_n = B_n$  means a first best level of investments.

If we could individualize  $t$  or  $s$ , we could always ensure that people's investment in education were at the first best levels by setting:

$$\left( (1-s)/(1-t) \right)_n = 1 + g'(S_n) \frac{N-1}{N} \quad (11)$$

This follows immediately from comparing the definition of  $A_n$  and  $B_n$ . Unfortunately, as  $g'$  is generally evaluated at different points for different workers, we cannot reach the first best optimum when  $t$  or  $s$  cannot be individualized, as is the assumption of this paper. We must then set  $(1-s)/(1-t)$  so as to minimize the sum of the cost of distortions of investments in education. However, in one particular case we do not need individualized taxes: As  $t \rightarrow 1^-$ ,  $S_n \rightarrow 0 \forall n$ . Then  $g'(S_n) \rightarrow g'(0) \forall n$ , which in turn implies that  $(1-s)/(1-t)$  should be the same for all workers.

Special attention should be paid to the case when  $s = t = 1$ . In this case  $A_n$  is not defined. When  $s = t = 1$ , the state pays all education and gets all the returns from it. Hence, workers are indifferent as to their length of education. One possible assumption in this case would be that when people are indifferent, the state can instruct people what to do. Obviously, the

state would then instruct people to choose the first best levels of education. This means that the state reaches a first best optimum, with the value of status in society maximized (this maximal value is zero) and no distortions of people's choices of education, if it sets

$$s = t = 1 \quad (12)$$

However, I do not find it very attractive to assume that the state can instruct people how much to educate themselves. Instead, I will investigate what happens when  $s = t \rightarrow 1^-$ , when the lump sum tax element  $a$  at the same time is adjusted so as to meet the public budget constraint. By l'Hopital's rule we obtain:

$$\lim_{s=t \rightarrow 1^-} A_n = w\bar{L} (1+g'(0)\frac{N-1}{N})/c'_n(e_n) \quad (13)$$

This means that there is overinvestment in education when  $s = t \rightarrow 1^-$ , since  $\lim A_n > B_n$ .

Taking all these arguments together, we realize that the solution approaches a first best optimum as

$$t \rightarrow 1^- \quad (14)$$

$$\frac{(1-s)}{(1-t)} \rightarrow (1 + g'(0)\frac{N-1}{N})^+ \quad (15)$$

Note that (15) implies that as  $t \rightarrow 1^-$ , also  $s \rightarrow 1^-$ , but in such a way that we always have  $s < t$ . E.g., if  $N$  is a large number so that  $\frac{N-1}{N} \approx 1$ , if  $g'(0) = 2$ , and if we have set  $t = 0.99$ , then optimal  $s$  will be 0.97.

When  $t$  is arbitrarily close to one, the value of status in society is arbitrarily close to its maximum value. At the same time, when  $t \rightarrow 1^-$ , we know that it is possible to realize a



first best level of education without the use of personalized educational subsidies. We use  $t$  to tax away differences in status. Then we use  $s$  to correct people's incentives to undertake education, but due to the inherent tendency to overinvestment in education when status matters, we set  $s$  somewhat below  $t$ . But as  $t \rightarrow 1^-$ , so will  $s$ .

Here, as opposed to a model with no status, educational subsidies have an independent role to play. Subsidizing education is not merely another way of lowering the tax on the resulting labour income. This is so, because (by assumption) educational choices are governed by life-time utility considerations, whereas status depends on after tax labour income. This means that raising  $s$  and lowering  $t$  have asymmetric impacts on people's status.

Note the importance of the assumption of no lending/borrowing possibilities. A joint use of  $s$  and  $t$  moves income from period 2 to period 1, while at the same time maintaining incentives to undertake education. If individuals had access to credit markets, they could of course move income back to period 2, the period where income yields status.

In the present model inequality was "counterproductive" in the sense that it reduced the total value of status in society. Naturally, it could be that inequality was counterproductive also in the sense that it lowered the production of physical goods. It could e.g. influence worker motivation, cooperation among workers, informal information flows in firms. As long as it is inequality in after tax labour income which is counterproductive, whereas life-time considerations rule educational choice, the basic insights of the model will survive.

Of course, conditions (14) and (15) must by no means be taken literally. The full equalization of everybody's after tax labour income is only optimal because we have assumed labour supply to be inelastic. In this section, I have tried to present the gist of my argument, by means of a very simple model. In the next

section we analyze more formally an extended version of the model.

### 3. An Extended Model.

The most restrictive assumption of the preceding section is the assumption of constant labour supply. In the model of section 2 only educational decisions and not labour supply decisions was affected by taxes and subsidies. In real life we know that taxation of labour income will distort labour supply decisions. To increase  $t$  means to reduce inequality, but in this broader perspective inequality has both productive and counterproductive effects. It is productive to allow inequality, because labour income taxation distorts labour supply. The counterproductive effect of inequality is still that the total value of status in society is decreasing in inequality. Now,  $t$  must be set to strike a balance between these two considerations, and it would be highly unlikely that we would still want to drive taxation to the point where no inequality remains, as was the prediction of the preceding section. Let us now have a more formal look at this.

An individual's utility in period 2 now will be written as:

$$u_{n2} = (1-t)w_n L_n - a - h(L_n) + g(S_n) \quad (16)$$

$$\text{where } S_n = (1-t)w_n L_n - a - \frac{1}{N} \{ \sum_n ((1-t)w_n L_n) - aN \}$$

Now  $L_n$  is a choice variable, and not exogenously given. The disutility of work is denoted  $h(L_n)$ , where  $h' > 0$ ,  $h'' > 0$ . Note that we assume it is total after tax labour income which decides status, and not a measure of income per hour.

With these alterations from the basic model, a representative worker's first-order conditions will be:

$$(1-t)wL_n \left(1+g'(S_n)\frac{N-1}{N}\right) = (1-s)c'_n(e_n) \quad \forall n \quad (17)$$

$$(1-t)we_n \left(1+g'(S_n)\frac{N-1}{N}\right) = h'(L_n) \quad \forall n \quad (18)$$

These conditions simply say that the marginal after tax income increase from more education/longer hours plus the value of increased status due to this increase should equal the marginal after tax cost of obtaining more education/working longer hours.

Conditions (17) and (18) yield the demand function for education and the supply function for labour:

$$e_n = e_n(s, t) \quad \forall n \quad (19)$$

$$L_n = L_n(s, t) \quad \forall n \quad (20)$$

(Remember that we have assumed that the demand for efficiency units of labour is perfectly elastic.) The effect of  $s$  on  $L_n$  is due to the fact that  $s$  influences  $e_n$ , and that the level of education in turn is important for the returns from working longer hours.

By inserting these functions into the utility function, we obtain an indirect utility function, where utility is expressed as a function of after tax "prices". This indirect utility function will be denoted  $V_n$ .

The government's problem will be to maximize

$$\sum_n V_n \quad (21)$$

subject to the public budget constraint

$$\sum_n (t w e_n L_n - s c_n(e_n)) + aN = \bar{T} \quad (22)$$

and

$$-1 < s, t < 1 \quad (23)$$

Some notation:  $e_{ns}$ ,  $e_{nt}$ ,  $L_{ns}$  and  $L_{nt}$  are partial derivatives of  $e_n$  and  $L_n$  with respect to  $s$  and  $t$ .  $\gamma$  is the Lagrangian multiplier associated with the constraint. We assume that local optimality also imply global optimality, and that the optimum is interior. The first-order conditions are:

$$-N + \gamma N = 0 \quad (24)$$

$$\begin{aligned} \Sigma_n \{ & -(1-s) c'(e_n) e_{nt} + (1-t) w(L_n e_{nt} + e_n L_{nt}) - w L_n e_n \\ & - h'(L_n) L_{nt} + g'(S_n) [-w e_n L_n + (1-t) w(L_n e_{nt} + e_n L_{nt}) \\ & - \frac{1}{N} \Sigma_{k=1}^N (-w e_k L_k + (1-t) w(L_k e_{kt} + e_k L_{kt}))] \} \\ & + \gamma \Sigma_n \{ w e_n L_n + t w(L_n e_{nt} + e_n L_{nt}) - s c'(e_n) e_{nt} \} = 0 \quad (25) \end{aligned}$$

$$\begin{aligned} \Sigma_n \{ & c_n(e_n) - (1-s) c'(e_n) e_{ns} + (1-t) w(L_n e_{ns} + e_n L_{ns}) \\ & - h'(L_n) L_{ns} + g'(S_n) [(1-t) w(L_n e_{ns} + e_n L_{ns}) \\ & - \frac{1}{N} \Sigma_{k=1}^N (1-t) w(L_k e_{ks} + e_k L_{ks})] \} \\ & + \gamma \Sigma_n \{ t w(L_n e_{ns} + e_n L_{ns}) - s c'(e_n) e_{ns} - c_n(e_n) \} = 0 \quad (26) \end{aligned}$$

By making use of the individual first-order conditions (17) and (18), we can eliminate a whole series of terms from (25) and (26). Further by noting that (24) implies  $\gamma = 1$ , more terms can be eliminated. We can now restate (25) and (26) as:

$$\Sigma_n \{ g'(S_n) [-w e_n L_n - \frac{1}{N} \Sigma_{k \neq n} (-w e_k L_k + (1-t) w(L_k e_{kt} + e_k L_{kt}))] \}$$

$$+ (twL_n - sc'_n(e_n))e_{nt} + t w e_n L_{nt} = 0 \quad (27)$$

$$\sum_n \{ g'(S_n) [ - \frac{1}{N} \sum_{k \neq n} (1-t) w (L_k e_{ks} + e_k L_{ks}) ] \\ + (twL_n - sc'_n(e_n))e_{ns} + t w e_n L_{ns} \} = 0 \quad (28)$$

Condition (27) says that part of the effect an increase of  $t$  has on the value of status in optimum should equal the negative of the net increase in tax revenue due to changes in labour supply and educational length following an increase in  $t$ . Condition (28) can be given a parallel interpretation.

This sounds very cryptic, but it turns out that these conditions have rather straightforward intuitive interpretations.

Let us start with (27). Raising  $t$  has three effects: First it has a direct effect on status. Further it influences education, which in turn influences the productivity of workers, costs of schooling, and status. The third effect is the effect on labour supply, which also in turn influences worker productivities, disutilities of work, and status.

However, parts of these effects are of the second order, i.e., they cancel out when we take into account that individuals optimize. The elements in (27) represent the three effects mentioned above, after second-order effects have been eliminated.

So now we can interpret (27) as follows: Increased taxation reduces the negative external effects from status-seeking, but at the same time distorts  $e_n$  and  $L_n$ . These effects must be balanced in optimum. Condition (28) can be interpreted in a parallel way.

Anyone familiar with the theory of optimal taxation will recognize the structure of these results. Status-seeking is but an example of reciprocal negative externalities. The theory of taxation with status-seeking will therefore be nothing but an

example of the theory of taxation with externalities.<sup>4)</sup> In this light, the main point of the present paper is that a certain activity (here: consumption) might have different external effects at different points in time. This makes it optimal for the government to use taxes and subsidies to control the distribution over time of the externality-generating activity.

In order more fully to capture the intuition behind (27) and (28), I think the following thought experiment might be instructive: In the previous section we saw that in optimum there was, so to speak, a strict division of labour between  $t$  and  $s$ .  $t$  was used to minimize inequality of after tax labour income, whereas  $s$  was used to restore incentives to undertake education. We could envisage that we also in the present setting would want there to be such a strict "division of labour". Then,  $t$  would be used to strike a balance between the wish to minimize differences of status and the wish to minimize distortions of labour supply. Moreover,  $s$  could be used to restore educational incentives to the best level achievable, given that personalized subsidies cannot be taken into use. If e.g. optimal  $t$  is less than one, we would have to choose  $s < t < 1$ , because of the tendency to overinvest in education.

However, such a procedure would generally not be optimal. A subsidy on education does not only influence education, but also labour supply. This in turn influences the optimal trade-off between status considerations and distortions of labour supply. Moreover, when  $s$  is not set at the first best level, we have to take into account that  $t$  influences education when determining optimal  $t$ . Precisely these considerations are reflected in (27) and (28). We see that the interplay between  $s$  and  $t$  here is rather more complicated than in the previous model, there is no longer a strict "division of labour".

Let us have a brief look at the case where  $L_n = \bar{L}_n$ , i.e., we

return to the simpler version of the model we looked at in the previous section. Conditions (27) and (28) then can be restated as:

$$\sum_n \{ g'(S_n) [ -w e_n \bar{L} - \frac{1}{N} \sum_{k \neq n} (-w e_k \bar{L} + (1-t) w \bar{L} e_{kt}) ] + (t w \bar{L} - s c'_n(e_n)) e_{nt} \} = 0 \quad (29)$$

$$\sum_n \{ g'(S_n) [ -\frac{1}{N} \sum_{k \neq n} (1-t) w \bar{L} e_{ks} ] + (t w \bar{L} - s c'_n(e_n)) e_{ns} \} = 0 \quad (30)$$

We have already argued that when  $t \rightarrow 1^-$ , it follows from concavity of  $g$  that the total value of status approaches its maximum value. This means that as  $t \rightarrow 1^-$ , the derivatives of the total value of status with respect to  $t$  and  $s$  approaches zero. Therefore, the first main terms both of (29) and (30) are zero. Moreover, we know that if we set  $s$  according to (15), we approach the first best levels of education where  $w \bar{L} = c'_n(e_n) \forall n$ . Hence, the second of the main terms of (29) and (30) are zero. This means that when we choose  $t$  and  $s$  according to (14) and (15), the first-order conditions (29) and (30) are arbitrarily close to being satisfied.

Returning to the more general case, we remember the restriction (23). A natural alternative restriction might be

$$0 < s, t < 1 \quad (34)$$

In this case we have new possibilities of corner solutions. E.g. it might be that  $g$  is almost linear, so that there is little to be gained by reducing inequality. At the same time the cost of distorting labour supply might be very substantial. This could lead us to a corner solution where  $t = 0$ . When  $s = t = 0$ , there is a tendency to overinvestment in education, so what we really would want is to tax education. As this is blocked by (34), we set  $s = 0$ . We see that within such a setting the main points of this paper become irrelevant.

#### 4. Conclusions.

This paper postulates that relative income, or "status", enters people's utility functions, and that the total value of status in society is largest when inequality is minimized. I propose that a natural measure for "status" is relative labour income after tax in the years after a concluded education. Moreover, I suggest that educational decisions are undertaken in order to maximize lifetime utility. This leads to a situation where it might be optimal to tax away some of the differences in people's labour income - and to use educational subsidies to restore people's incentives to undertake education.

The presented model is admittedly speculative, in that its main point is based on a specification of people's utility functions for which I offer no empirical justification. Personally I feel the model captures a point of some importance. Admittedly, such simple models offer only the vaguest guidelines for public policy. But I think it is equally ad hoc and equally little supported by empirical evidence (although much more common) to base policy recommendations on the assumption that status does not enter people's utility functions. What seems to be needed is much more thorough empirical knowledge about what really governs people's choices.

Some might argue that status enters people's motivation structure in a more complicated way than described here. Status might not only depend on income per se, but also on the purchase of different goods. Some goods - "positional goods" - might yield more status than others, and precisely education might be one such positional good. Moreover, it might be that movements up or down a ladder of relative positions have different impact. I fully agree with these arguments. But as long as an increase in disposable income during one's education and a corresponding decrease in the years after a concluded education do not leave utility derived from status unchanged, my main point will hold in some modified form.



Notes.

\* I thank Geir Asheim and Kåre Petter Hagen for valuable comments.

- 1)  $\bar{w}$  can be thought of as gifts from parents, government grants, earning capacity in the first period, etc. The exact interpretation of  $\bar{w}$  determines the realism of the assumption that  $\bar{w}$  is untaxed.
- 2) I have here chosen to model educational costs as pecuniary costs only. Realistically, a large part of the costs in education is the alternative value of time. Including such time costs would not change the flavour of my results, but would be notationally awkward.
- 3) A first best solution arises when there are no distortions from taxes or subsidies, and when the negative externalities from consumption in period two are fully internalized. If this were the case, optimal choice of  $e_n$  is described by:

$$w\bar{L}(1+g'(S_n)) = \sum_n g'(S_n) \cdot \frac{w\bar{L}}{N} + c'(e_n)$$

If the first best situation should imply, as it turns out that it does within the present model, that everybody has the same income and status in optimum, the relationship can be rewritten as:

$$w\bar{L} = c'(e_n)$$

This implies condition (10).

- 4) Some contributions to the theory of optimal taxation with externalities are Baumol and Oates (1975), Diamond and Mirrlees (1973), Hagen (1978) and Sandmo (1980).

## REFERENCES

- Baumol, W. J. and W. E. Oates (1975), The Theory of Environmental Policy, Princeton: Prentice-Hall.
- Boskin, M. J. and E. Sheshinski (1978), "Optimal Redistributive Taxation when Individual Welfare Depends upon Relative Income", Quarterly Journal of Economics 92, 589-601.
- Diamond, P. A. and J. A. Mirrlees (1973), "Aggregate Production with Consumption Externalities", Quarterly Journal of Economics 87, 1-24.
- Duesenberry, J. (1949), Income, Savings and the Theory of Consumer Behavior, Harvard University Press.
- Festinger, L. (1954), "A Theory of Social Comparison Processes", Human Relations 7, 117-140.
- Frank, R. H. (1984a), "Are Workers Paid Their Marginal Products?", American Economic Review 74, 549-571.
- Frank, R. H. (1984b), "Interdependent Preferences and the Competitive Wage Structure", Rand Journal of Economics 15, 510-520.
- Frank, R. H. (1985a), "The Demand for Unobservable and Other Nonpositional Goods", American Economic Review 75 101-116.
- Frank, R. H. (1985b), Choosing the Right Pond: Human Behavior and the Quest for Status. New York: Oxford University Press.
- Gylfason, T. and A. Lindbeck (1984), "Union Rivalry and Wages: An Oligopolistic Approach", Economica 51, 129-139.

- Hagen, K. P. (1980), "Consumption Externalities and Direct versus Indirect Corrective Pricing", European Economic Review 11, 59-73.
- Hirsch, F. (1976), Social Limits to Growth. Cambridge: Harvard University Press.
- James, J. (1987), "Positional Goods, Conspicuous Consumption and the International Demonstration Effect Reconsidered", World Development 15, 449-462.
- Johnson, G. E. (1984), "Subsidies for Higher Education", Journal of Labor Economics 2, 303-318.
- Kodde, D. A. and J. M. M. Ritzen (1985), "The Demand for Education under Capital Market Imperfections", European Economic Review 28, 347-362.
- Layard, R. (1980), "Human Satisfaction and Public Policy", Economic Journal 90, 737-750.
- Leibenstein, H. (1950), "Bandwagon, Snob and Veblen Effects in the Theory of Consumer Demand", Quarterly Journal of Economics 64, 183-207.
- Ng, Y.-K. (1987), "Real Income Effects and the Appropriate Level of Public Expenditure", Oxford Economic Papers 39, 293-300.
- Orosel, G. O. (1985), "Tentative Notes on Prestige Seeking and Pareto-Efficiency", Working Paper No. 8509, University of Vienna.
- Oswald, A. J. (1979), "Wage Determination in an Economy with Many Unions", Oxford Economic Papers 31, 369-385.

Oswald, A. J. (1983), "Altruism, Jealousy and the Theory of Optimal Nonlinear Taxation", Journal of Political Economy 20, 77-87.

Sandmo, A. (1980), "Anomaly and Stability in the Theory of Externalities", Quarterly Journal of Economics 94, 799-807.

Sen, A. (1983), "Poor, Relatively Speaking", Oxford Economic Papers 35, 153-169.

Veblen, T. (1899), The Theory of the Leisure Class.

Williams, R. M. (1975), "Relative Deprivation". In L. A. Coser (ed.), The Ideal of Social Structure: Paper in Honor of Robert K. Merton. New York: Harcourt, Brace, Jovanovich.