The Determinants of Workers' Effort: Theory and Evidence

Steven McIntosh

PhD Thesis

London School of Economics

Abstract

The opening chapter of the thesis reviews efficiency wage literature. After considering the theoretical justifications that have been offered for a link from wages to effort, which is at the heart of efficiency wage theory, the chapter continues with an account of the attempts made so far in the literature to test this theory.

The first of two empirical chapters examines how effort varies across individual characteristics. The efficiency wage link between individuals' earnings and their effort is investigated in a two-stage least squares framework, to allow for the endogeneity of earnings. Further tests check for the existence of a reverse causality argument. The chapter ends with an investigation into the effect of local unemployment rates on effort.

A second empirical chapter examines the effect of workplace characteristics on effort. The influence of trade unions is considered by examining whether management policies to elicit effort from their employees, for example raising the cost of job loss or human resource management techniques, have differential effects on effort according to the degree of unionisation of the workplace.

The final chapter is theoretical, considering group norms for effort, whereby the effort choice of individuals is influenced by their colleagues at work. Under the assumption that if a firm observes a single worker shirking, all members of that individual's workgroup will receive a sanction or lose a reward, the effort equilibrium of the workgroup is derived as a non-cooperative Nash equilibrium in expectations and effort choices. Given this equilibrium, the analysis continues by examining whether or not it would be in the workgroup's interest to establish a group norm for supplying effort. Thus the chapter provides a theoretical justification for the existence of, and adherence to, group norms for effort, which has been absent from previous analyses which have included such a concept.

Table of Contents

List of Tables

List of Figures

Acknowledgements

My first acknowledgement must go to my supervisor, Andrew Oswald, for all the help that he has provided during the course of writing this thesis. This has taken many forms, principally the advice he has given concerning the research plan to be followed, and later the content of the written results. Of almost equal importance has been the constant support that he has offered, and his faith during my times of doubt. For all of this I am very grateful.

I have also benefitted from the help and advice of colleagues at the London School of Economics, and in particular members of the Centre for Economic Performance. I would like to express my gratitude to Donna Brown, Jan Eeckhout, Steve French, Francis Green, Adam Lubanski, Steve Machin, Marco Manacorda, Alan Manning, Lucy Matthew, Dave Metcalf, Kaspar Richter, Jonathan Wadsworth, Jane Waldfogel and Dave Wilkinson. I am very fortunate to have such friends.

The three data sets in this thesis were all supplied by the ESRC Data Archive at the University of Essex. The 1989 British Social Attitudes Survey was collected by Social and Community Planning Research. The 1989 International Social Survey Programme data was documented and originally made available by the Zentralarchiv Fuer Empirische Sozialforschung in Koeln, the data having been collected by various national independent institutions. The principal collectors of the 1990 Workplace Industrial Relations Survey were M. Stevens and N. Millward. None of these individuals or institutions bear any responsibility for the analyses or interpretations presented here.

I would also like to gratefully acknowledge financial support from the Economic and Social Research Council (PhD Research Award number R00429234214).

Finally, thanks to my family, and friends not mentioned above, for their support.

Chapter 1 : Introduction

Effort is an important economic variable, but one which economists know little about. Even with all the capital and labour in the world, if all workers input zero effort, nothing will be produced. Effort is therefore a vital part of the production process. The fact that only a limited amount of economic research has considered this crucial variable is the primary motivation for this study.

If effort is a constant, then there is little to investigate. This is unlikely to be the case, however. Surely everyone knows somebody who cares little about their job and will do the least amount of work they can get away with. Equally, everyone knows a perfectionist who throws themselves into a job and will not rest until a task is completed. Within these extremes there is likely to be a whole range of possible effort choices. The aim of this thesis is to investigate the determinants of such choices.

Implicit in these opening paragraphs is an assumption that will be maintained throughout the thesis: that individuals *can* choose their own effort levels. Such an idea can be traced back to the old Marxian principle that firms buy units of labour *time*, rather than units of labour *effort*. Thus, an employment contract will typically detail the number of hours worked, but it will not describe how hard the employees have to work in that time. There are probably a number of reasons why contracts do not contain specific effort requirements, such as the huge informational requirement needed to determine an optimal effort level, the degree of supervision required to enforce such a contract, the demotivational aspects of completely pre-determined work patterns and constant supervision, and finally the question of who sets effort levels for the supervisors.

Given all of these problems, labour contracts are likely to be incomplete, in the sense that they will not totally specify the effort required of the worker, who is then left with some discretion in choosing an effort level. What factors will influence this choice? Few areas of economics have considered this question.

The largest economic literature to be accumulated on the concept of workers' effort is that of efficiency wage theory. The literature review concentrates on this theory, since it will be an important point of reference in all of the analysis that follows. The basic premise is that workers' effort responds positively to the wage that they are paid. Theoretical work in this field has concentrated on why such a link should exist. Four main ideas have been put forward, all of which are reviewed in Chapter 2. The basic arguments are that a high wage reduces an individual's motivation to shirk for fear of dismissal, reduces the likelihood of productive workers quitting, attracts higher quality, and hence more productive, workers, or leads to the offer of a 'gift' of high effort from employees. As far as testing these theories is concerned, the principal problem is a lack of data that accurately measures effort. Various attempts have been made to offer empirical tests of efficiency wage theory, however, and these are also reviewed in Chapter 2.

The literature review is followed by two empirical chapters that attempt to gather evidence on the determinants of workers' effort. Both use self-reported effort in their analysis, one at an individual level and the other at an establishment level. The first therefore describes how effort varies with individual characteristics, in an attempt to answer the question 'who works the hardest?' Is it men or women, the young or the old, the well educated or the less well educated, the skilled or the unskilled? The analysis continues with a test of the efficiency wage prediction that an individual's effort will respond positively to the wage that he or she receives. The possibility of reverse causality, whereby wages respond to effort, has often been ignored in work done so far in this area, but will be a central part of the analysis offered here. A final section considers whether an individual's effort choice is affected by the level of unemployment in their local area.

The chapter based on establishment-level data takes a broader view by considering how workplace characteristics affect average effort levels. The results provide answers to questions such as; how do the average characteristics of the workforce affect the average effort level, how is effort related to the degree of unionism in a workplace, does organisational change affect effort choices, what is the impact of management schemes to elicit more effort from their employees, and how do such schemes interact with the presence of strong trade unions? The empirical evidence that exists on effort determinants has usually focused on the characteristics of individuals, and does not consider such questions. The schemes under consideration include the use of human resource management techniques, and the efficiency wage mechanism of raising the cost of job loss to dissuade shirking.

The final chapter of the thesis is an attempt to improve the theoretical efficiency wage literature. Of the four types of efficiency wage models described above, the one that is the least formally developed is the sociological 'Gift Exchange' model. This theory argues that a high wage will lead to the formation of a group norm for supplying high effort among employees. However, this is merely proposed, and the mechanisms at work are not derived in a formal model. The aim of Chapter 5 is therefore to model the formation of, and adherence to, group norms for effort, within the typical economic framework of individual utility maximisation. It will also be shown how the theoretical hole in the 'Gift Exchange' model can be filled by the group norm for effort thus derived, so providing that theory with a greater degree of rigorous formality. With this improvement to the theoretical efficiency wage literature, together with the evidence describing the relevance of this theory, as well as other effort determinants, it is hoped that this thesis can improve the existing knowledge about the important topic of workers' effort choices.

Chapter 2

Literature Review

2.1 : Introduction

The thesis that follows this literature review is devoted to the concept of workers' effort. Compared with other inputs to the production process, effort has not received a great deal of attention in economic literature. Other social sciences, such as psychology, have had more to offer in this area. In particular, the idea of worker motivation, or persuading workers to supply more effort in their jobs, has long been found in the industrial psychology literature. Economics has begun to concern itself with such matters rather later, although some ideas have begun to appear. A review of economic models of motivation and worker effort choices is provided by Lazear (1995). Perhaps the most considered topic in this area is that of variable pay, whereby a worker's income is based on some measure of output produced, rather than a fixed amount per time period worked. The most obvious example of such a compensation scheme is a piece rate, in which a worker is paid a set amount for each unit of output produced, for example:

$$pay = \alpha + \beta q$$

In this example, the worker is paid β for each unit of output, q, produced, as well as a flat rate, α. For ease of argument, normalise the measurement of effort so that one unit of effort produces one unit of output. Then, it can be shown that the profit maximising choice of piece rate chosen by the firm will be 1, so that the worker is paid the full value of her effort. Since she will choose to supply effort where the marginal increase in compensation equals the marginal disutility of effort, and that the value of her compensation equals the value of her output, this means that the marginal increase in output, that is the social value of the effort, just equals the marginal cost of that effort to the individual. Thus, such a piece rate leads to an efficient effort choice. The fixed part of the worker's compensation, α, can be set to ensure that total compensation is large enough relative to outside alternatives to attract the worker to the firm.

There are a number of problems with piece rates, which ensure that they are only witnessed in a minority of employment relationships. Perhaps the most obvious is that it is necessary to be able to measure the number of units of output produced by the worker, in order to calculate the payment to which she is due. The more difficult or expensive such measurement is, then whatever the motivating impact of the piece rate scheme, the measurement costs involved mean that it is less likely to be used. A related point is that even in cases where output can easily be measured in terms of physical units, it may be less than straightforward to assess its value. In this case, if workers know that they are to be paid only for the number of physical units produced, they may sacrifice quality for quantity. The more difficult the firm finds it to fully assess the output produced, along the quality, or any other, dimension, the less likely are piece rates to be used.

Risk aversion amongst workers is another reason why piece rates are often not used. The problem is caused by the fact that the worker is being paid according to a measure of output, rather than the effort that she inputs, and variation in the former may be caused by variation in factors other than the latter. Thus, a worker's output could be low through no fault of her own, for example a faulty machine. If the variation in such other factors is high, then risk averse workers may be unwilling to accept a piece rate contract.

A final problem, known as the ratchet effect, is described graphically by Roy (1952). If employees work hard following the introduction of a piece rate system, and achieve high earnings, then the firm may consider that the work is too easy, and adjust the piece rate accordingly. Knowing this, employees may not work hard in the first place, and so the system does not have the desired effect on effort. Roy describes in great detail the informal mechanisms used within work groups to stop anyone working too hard, for fear that a piece rate will be cut in future periods.

Related to piece rates are merit pay schemes, whereby workers are paid not according to objective measures of output, but subjective measures of performance. The advantage of such schemes over piece rates is that firms do not have to be able to measure a particular individual's output exactly, although clearly some monitoring is still required, if fair performance appraisals are to be supplied. The main disadvantage would seem to be the subjective nature of the assessment. If workers fear that they may be unfairly treated in the assessment procedure, then they may not co-operate with the scheme in the first place.

The schemes described so far have been individual in nature, but incentive plans can also be collective, whereby all employees are rewarded on the basis of their collective performance. Perhaps the most well-known example of such compensation schemes is profit-sharing. As has been frequently pointed out, such compensation schemes face the '1/n' problem. This problem is a result of the fact that if an individual raises the profits of the firm by £1, she will receive only γ.£1/n as a profit-sharing reward herself, where n is the number of workers involved in the scheme, and γ is the profit-sharing coefficient. Individual workers are therefore not receiving the correct incentives, and may decide simply to free-ride on the efforts of others. The benefits of collective compensation schemes are said to be that they help in the creation of group norms for effort. It is claimed that individuals' feelings of guilt and shame at letting their colleagues down are the forces which create the group norm. Chapter 5 of this thesis develops a theoretical model of the formation of group norms for effort, under a particular kind of collective compensation system; that of collective sanctions, whereby all group members suffer a sanction if one of their number is found to have not been supplying effort.

A further wage plan designed to motivate workers is the upward-sloping, experience-earnings profile. The idea in this case is to pay individuals less than they are worth when they are young, and more than they are worth when they are older. Young people are motivated to work hard and keep their jobs by the promise of higher earnings later in their career, while the receipt of these high earnings, and the desire to continue receiving them, is what motivates older workers. A possible problem which such an arrangement is that, although the firm can design the experience-earnings profile so that average earnings just equal what the individuals are worth across their working lives, when it comes to the end of their natural working lives, employees are reluctant to relinquish their jobs, because they are earning more than they are worth at this point. Thus, upward-sloping experience-earnings profiles need to be used in conjunction with mandatory retirement, or some other incentive to retire such as pensions, if they are to be successful. Indeed, the existence and use of such motivating schemes has been advanced as a reason for why mandatory retirement occurs (Lazear, 1982).

A final motivation mechanism considered by Lazear (1995) is the use of relative compensation. Such schemes involve employees being rewarded not on the basis of their absolute level of performance, but on their performance relative to others. The discussion of such motivational devices has become known in the literature as tournament theory, and the analogy with sport is easy to recognise. Just as professional sports players, such as tennis players or golfers, receive payment for performing better than others and winning their tournaments rather than for their absolute level of performance, so employees in such schemes are rewarded for out-performing their colleagues, regardless of their actual level of performance. The motivational impact of such schemes should be clear, in that employees put forth effort in an attempt to win the 'tournament' and claim the 'prize.' The greater the difference in rewards between winning and not winning, the more motivated will be the employees. Similarly, the more homogeneous the workers, the more motivated they will be, since they will all consider that they have a good chance of winning. If one worker is of a much higher quality than the others, then the latter will consider that they do not have a chance of winning the reward and so do not even try, while the former will be aware of this lack of competition and so will feel she does not need to work too hard herself.

Promotion to a higher-paying position is usually the prize on offer, when tournament theory is applied to workplaces. Workers lower down in the firm's hierarchy are therefore hypothesised to work hard to prove themselves more worthy than their colleagues of promotions up the corporate ladder. Such theory can be used to explain the very high salaries earned by chief executive officers (CEOs) of large firms. It is not that a CEO is worth her salary in terms of her contribution to the firm, but that the possibility of being promoted to a CEO and earning such salaries provides the motivation for employees lower down in the firm's hierarchy to work hard.

The main drawback of tournament-style motivation schemes is that, by definition, employees are in competition with each other. In work settings where cooperation between workers is important for getting the job done, this can be a serious problem. Individuals are unlikely to help a colleague, if this means that they could be helping that person to win the prize, at the expense of themselves. Some individuals may even go as far as sabotage of others' work, if by doing so they can increase their own chances of winning the prize. Therefore, tournament theory should work best when cooperation is less important, for example when individuals from different regional branches are competing for a well-paid position at the head office of a firm.

The emergence of all of these theories has been important in developing our understanding of worker motivation and the determinants of workers' effort choices. They will be discussed again at various points in this thesis. For example, in Chapter 4, the existence of performance related pay is used as an explanatory variable in regression equations where the dependent variable is the average effort levels of manual workers in various establishments. Also, as mentioned above, the existence of a particular collective compensation scheme is taken as the basis for deriving group norms for effort in Chapter 5. However, most of the thesis is concerned with the efficiency wage hypothesis, and so the remainder of this literature review will concentrate on that theory. The following section reviews the theoretical approaches that have been taken by efficiency wage researchers. Such a review is not original, and can be found, for example in Akerlof and Yellen (1986). The section that follows does represent an original contribution to the topic, in that it extensively describes the attempts that have been made by various authors to test efficiency wage theory empirically.

The basic idea of efficiency wage theories is that the effort or productivity of a firm's workforce depends upon the wage it pays. A difference between this theory of effort choices and those described above is immediately apparent. Whereas in all of the theories outlined above the payment of a reward followed the effort choice, in the case of efficiency wages, the payment precedes it; workers make their effort choices on the basis of the wage that they are already receiving. Psychological theory such as expectancy theory, described in the following chapter, says that payment systems can only have a motivational impact if earnings are explicitly tied to performance, thus leaving no room for efficiency wage-type relationships. It has therefore been an important task for efficiency wage researchers to develop theoretical models for why workers should respond with more effort to higher wages.

Although he did not use the term 'efficiency wage,' the first identifiable efficiency wage model was developed by Leibenstein (1963). Leibenstein's original model set out to explain why, in developing countries, agricultural workers receive a positive wage, even when their marginal product is zero, in the sense that even if some workers were taken away, the smaller labour force could still cultivate the same amount of land and produce an unchanged amount of food. Leibenstein argues that competitive forces do not force the agricultural wage towards zero, because as wages fall the nutrition levels consumed by workers will fall, and as nutrition falls, productivity will also be adversely affected. At subsistence levels, both assumptions seem reasonable, and Leibenstein backs them up with evidence from studies conducted by dieticians. Wages are therefore kept above their market-clearing level, because if wage cuts were made, total output would suffer, thus making the wage cut undesirable in the first place. Thus, this was the first model to establish a link between the wage received by workers and their productivity.

In the 1970s, Leibenstein considered the concept of effort explicitly as a choice variable under the control of workers, although he moved away from the efficiency wage relationship between effort and wages. In a series of articles and books, for example Leibenstein (1975, 1976, 1978 and 1979), he introduced and developed the idea of 'X-inefficiency' which refers to 'the degree to which actual output is less than maximum output (for given inputs)' (Leibenstein [1976]). The less effort that workers supply relative to the level that they are capable of, the more X-inefficient will be the firm.

Leibenstein hypothesised that individual workers will choose at which level to supply effort by maximising a utility function. The personality characteristics of the individual, and the demands of the organisation are included as arguments in this utility function, and hence affect the level of effort chosen. Although he adopts the economist's idea of utility maximisation, Leibenstein's theory is essentially psychological in nature. As for its motivational content, it principally considers 'sticks' rather than 'carrots,' the idea of the demands of the organisation covering pressure put on the workers by their employers. There is therefore no room for efficiency wage-type relationships of workers' effort responding to the wage received in this theory. Nevertheless, Leibenstein's work was very important in that it stressed that effort should not be regarded as either a constant that can be ignored, or as something that can be set at a particular level and hence is totally within the control of management. Leibenstein made it clear that effort will be a discretionary variable that can be set, to a greater or lesser extent depending on the characteristics of the work involved, by the individual worker. Adding this idea to his early work about an efficiency wage-type relationship in under-developed countries led to the formation of modern efficiency wage models relevant to developed countries, to which this review now turns.

2.2 : Theoretical Efficiency Wage Models

One of the earliest works to outline the implications of the existence of efficiency wage relationships between earnings and effort is Solow (1979). From this paper the now well-known Solow condition for efficiency wage payments has been derived. Consider a firm with the production function

$$Q = f\,[e(w)L]$$

Output, Q, is a function of the number of efficiency units hired, defined as the number of workers hired, L, multiplied by the effort, e , input by each. Effort is assumed to be a function of the wage paid by the firm, in accordance with the efficiency wage hypothesis.

The firm's profit function is

$$\pi = P.f\,[e(w)L] - wL$$

Maximizing with respect to the firm's choice variables, w and L, gives two first order conditions:

$$w : PLe'(w)f'\,[e(w)L] - L = 0 \tag{2.1}$$

$$L : Pe(w)f'\,[e(w)L] - w = 0 \tag{2.2}$$

The 'symbol denotes the derivatives of the effort and production functions. This notation will be used throughout, where it is clear with respect to which variables functions are being differentiated.

Rearranging equation 2.2:

$$Pf' = \frac{w}{e(w)}$$

Substituting this expression into equation 2.1:

$$e'(w).\frac{w}{e(w)} = 1$$

This is the Solow condition, which states that the firm should increase the wage it pays until the elasticity of effort with respect to the wage is unity. An increase in the wage beyond this point will bring forth a proportionately smaller increase in effort, and so is not profitable for the firm. Equation 2.2 is the usual condition that says that the firm should employ labour until the point at which the marginal revenue product equals the wage paid.

Efficiency wage theories have been used to explain the existence of involuntary unemployment. If total labour demand, as determined by equation 2.2, is less than total labour supply, then there will be involuntary unemployment, in the sense that individuals without a job are willing to work at the going wage, but cannot find a job. In competitive theories of the labour market this would put downward pressure on the wage level. In efficiency wage theories, however, there is no incentive for firms to reduce their wages, since effort would fall by a proportionally larger amount. Individuals out of work therefore remain involuntarily unemployed.

One immediate criticism of the theory is that an effort-wage elasticity of unity seems very high, and may not be obtainable. Akerlof and Yellen (1986), in the introduction to their book, offer a defence against this objection. They show that the Solow condition is a result of the multiplicative, 'efficiency unit' form of the production function, as written above. Such a form assumes that if a worker varies his or her effort, this will only affect their own labour input, and not that of fellow workers. If this is not the case, which certainly seems possible, then other production functions should be used, which Akerlof and Yellen show can result in a lower necessary effort-wage elasticity. The Solow condition is therefore based upon a simplification as an illustration of the efficiency wage hypothesis, which may not hold exactly in reality.

Solow does not go into much detail about *why* effort should depend on the wage paid by the firm, only briefly mentioning the possible morale effects of higher wages. The task of efficiency wage authors following Solow has been to present more detailed models, offering theoretical microfoundations for an effort-wage relationship. One of four general approaches is usually taken, as outlined below.

a.) The Shirking Model

The classic reference for the shirking model is the paper by Shapiro and Stiglitz (1984). This paper is concerned with the old Marxian problem of how the firm can extract effort from its workforce. As outlined by Bowles (1985), however, Shapiro and Stiglitz do not follow the Marxian method of using class conflict to explain the need for effort extraction. Rather, they simply assume that workers always prefer to shirk, requiring employers to offer them some incentive not to do so. In their model this incentive is initially a higher wage than workers would receive elsewhere. There is therefore a cost to job loss, and if the firm makes it clear that workers caught shirking will be fired, this provides the correct incentives. In the simple model of identical firms, this is the optimal policy for all firms. Since the wage they all pay is above the market-clearing level, there will be unemployment. When all firms are paying the same wage, it is this unemployment that provides the cost of job loss, since a worker caught shirking would be fired and forced to spend time in the unemployment pool, with an associated loss of work income. This in turn provides the incentive for workers not to shirk.

It is assumed in the model that firms and workers cannot simply write contracts specifying the amount of effort required of the worker, because of the difficulties involved in observing the effort input by each worker. Firms therefore use the incentive methods described above.

Each worker has a utility function defined, with suitable normalization, as u=w-e. This clearly shows the assumed disutility of effort. A simplifying assumption is that the worker can make one of only two possible effort choices; either she can input an effort level, e, or she can shirk, in which case her effort is zero.

The expected lifetime utility of a shirker is then

$$rV_E^S = w + (b+q)(V_U - V_E^S) \tag{2.3}$$

b+q is the joint probability that a shirker moves from the employed into the unemployed state, the latter having expected utility, V_U. b is the separation rate of workers from firms for reasons other than shirking, for example voluntary quits. It is assumed to be exogenous. q is the probability of being caught shirking. It is assumed that firms will always fire workers it catches. r is the worker's discount rate.

Similarly, the expected lifetime utility of a non-shirker is

$$rV_E^N = w - e + b(V_U - V_E^N) \tag{2.4}$$

Equations 2.3 and 2.4 can be re-arranged so that only the expected utility from employment terms are on the left-hand sides:

$$V_E^S = \frac{w + (b+q)V_U}{(r+b+q)} \qquad (2.5)$$

$$V_E^N = \frac{w - e + bV_U}{(r+b)} \qquad (2.6)$$

The firm wants to provide sufficient incentive to persuade their employees to work rather than shirk. The workers will compare the expected utility of each state, and choose to expend effort if the utility derived from doing so is greater than the utility they would gain if they shirked; that is, if V_E^N is greater than V_E^S. Using equations 2.5 and 2.6, this implies that to avoid shirking by its workforce, a firm must pay a wage such that:

$$\frac{w - e + bV_U}{(r+b)} \geq \frac{w + (b+q)V_U}{(r+b+q)}$$

This can be re-arranged to give

$$w \geq rV_U + \frac{(r+b+q)e}{q} = w^* \qquad (2.7)$$

This is the critical, no-shirking wage, which is the wage that the firm actually pays, given that further increases in the wage will no have no further effect on effort.

Shapiro and Stiglitz then go on to describe the full market equilibrium, defined as the situation where each identical firm finds it optimal to pay the going wage, taking as given the wages and employment levels of all other firms. Each firm therefore pays the no-shirking wage, and the equilibrium is a no-shirking one.

Now define an asset-type equation, similar to those above, for the expected utility from unemployment:

$$rV_U = \bar{w} + a(V_E - V_U) \qquad (2.8)$$

where $\bar{w}$ is income received whilst unemployed, and a is the probability of acquiring a job, which will yield expected utility V_E. In equilibrium, utility in employment will be at the no-shirking level, since nobody shirks in the equilibrium state. Equations 2.6 and 2.8 can be solved simultaneously for V_E and V_U, substituting V_E for V_E^N in equation 2.6. After substituting the resulting expression for V_U into the no-shirking condition (equation 2.7), the result is the aggregate no-shirking condition.

$$w \geq \bar{w} + e + \frac{e(a+b+r)}{q} \tag{2.9}$$

From equation 2.9 certain predictions can be made. For example, the critical wage which firms must pay to prevent shirking will necessarily be higher, (I) the more effort they require, (ii) the higher the level of unemployment income, (iii) the easier it is to find a job whilst unemployed, (iv) the higher the separation (quit) rate, (v) the higher the discount rate, and (vi) the lower the probability of being caught shirking. Effects (ii) and (iii) reduce the cost of job loss, making workers less concerned about being caught shirking. If a worker intends to leave the firm anyway, or if there is little chance of being caught, then he will be more likely to shirk, explaining points (iv) and (vi). Finally, the higher the worker's discount rate, the more he or she will value the present utility gained from shirking rather than the future gains of continued employment.

Shapiro and Stiglitz take equation 2.9 one step further, by modelling the probability of finding a job whilst unemployed, a. In a steady-state equilibrium, the flows into unemployment must equal the flows out of unemployment. Letting N denote the total size of the labour force, and L be the number who are actually employed, this condition can be written as:

$$bL = a(N - L)$$

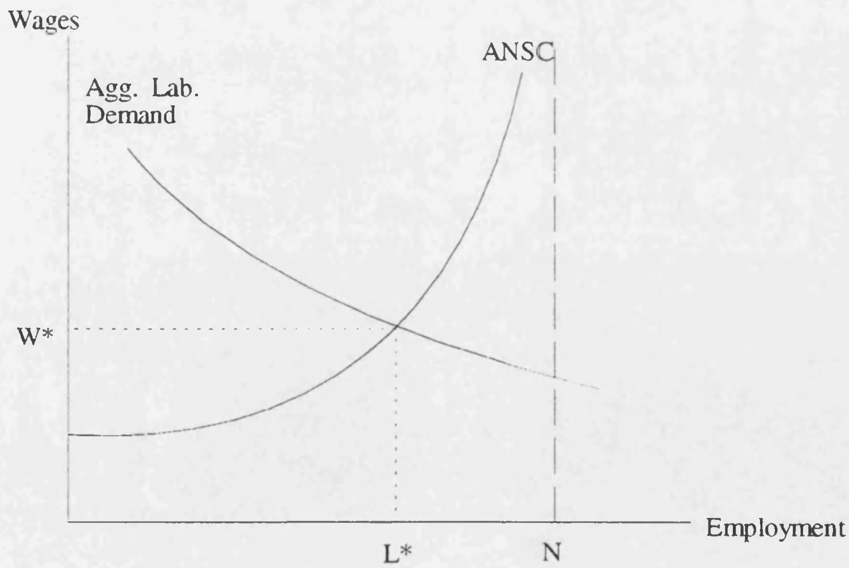
Hence,

$$a = \frac{bL}{(N - L)} \tag{2.10}$$

Equation 2.10 can then be substituted into equation 2.9, the aggregate no-shirking condition, which, after suitable re-arranging becomes

$$w \geq \bar{w} + e + \frac{e}{q}\left(\frac{b}{u} + r\right) \tag{2.11}$$

where u is the unemployment rate, (N-L)/N. Equation 2.11 can be graphed as a relationship between the wage paid by the firm to prevent shirking and the unemployment rate, labelled as the aggregate no-shirking condition (ANSC) below.



The equilibrium wage chosen by all firms is w^*, at which the labour demand curve intersects the aggregate no-shirking condition. It is clear that there will always be some unemployment (N-L*), since the full employment line is an asymptote to the no-shirking condition, implying that an infinite wage would be required to prevent shirking if the economy was at full employment. Some unemployment is therefore always needed, otherwise there would be a zero cost to job loss, and so workers would always shirk.

The remainder of Shapiro and Stiglitz's paper examines welfare implications, and shows that the unemployment rate that results is not efficient. The problem is that firms see the wage they pay, w, as the private cost of hiring one more worker. The social cost of one more worker in employment is only the effort which that worker must input, which is necessarily less than w, otherwise nobody would work. Thus the private cost to the firm is greater than the social cost, and so there is under-employment; the level of unemployment is too high.

b.) The Bonding Critique

A criticism of efficiency wage theory is that certain contracts could be written that would negate the need for efficiency wage payments. This has become known as the 'bonding critique' of the efficiency wage hypothesis. As Katz (1986) points out, the criticism arises from the observation that in efficiency wage theories, the wage is performing a dual role. On the one hand, it is allocating labour in the usual way, but on the other hand, it is also providing incentives to workers. If some other device performed the latter role, this would leave the wage free to act in the standard way, clearing the market for labour.

If there was some other cost to job loss, then it would not be necessary to pay high wages as an incentive device. One possibility is a straightforward fine of a worker caught shirking. Other schemes could involve requiring a worker to pay a bond or a fee upon taking up a job with the firm. A bond would be returnable unless the worker was caught shirking, in which case he or she would forfeit their bond. In the case of the employment fee, this would not be returnable under any circumstances, but the worker still would have an incentive not to shirk, since doing so would result in the sack and therefore the need to pay another entrance fee to regain employment elsewhere. All of these schemes free wages to perform their usual role of clearing the labour market, implying that there need be no involuntary unemployment.

Dickens *et al* (1989) show that such bonding schemes should always be adopted by firms, since they involve no cost to the firm (assuming negligible collection costs), whereas paying the high no-shirking wage clearly increases costs. Casual inspection of the vast majority of employment situations shows that bonding schemes, in whatever form, are not adopted. There must be something missing from the analyses of those who propose such schemes that explains why firms do not, or cannot, use bonds to provide incentives. Dickens *et al* (1989), amongst others, suggest some possibilities. One of the most frequent explanations put forward to explain the lack of bonding is that workers cannot afford to pay the sums involved. This is particularly the case where the probability of being caught shirking is low, so that a large bond is required to ensure that the expected utility derived from shirking is less than the expected utility of not shirking. With imperfect capital markets, workers will also be unable to borrow the money necessary to post an appropriately sized bond. However, as pointed out by Carmichael (1985), firms should still extract whatever workers are able to pay in the form of entrance fees or bonds. These fees would be bid up by unemployed workers, until the expected lifetime utility from working would just equal the expected lifetime utility from not working. That is, the entrance fee would be bid up until almost all the money that workers could get hold of would be going towards paying it, with the resulting poverty and starvation in the current period reducing lifetime utility down to the reservation level. Thus the involuntary unemployment result of efficiency wage theory disappears, in the sense that unemployed workers are just indifferent between accepting employment or not.

Although arguing that involuntary unemployment can be removed by making workers suffer a period of starvation upon commencing a job seems to be merely playing with definitions rather than being an acceptable solution, the main point of Carmichael's criticism still remains - even if workers are liquidity constrained, firms should still extract as much as workers are able to pay in entrance fees, rather than pay efficiency wages. Liquidity constraints do not seem to be the reason why bonding is not generally observed.

Another possible reason for the lack of bonding observed is moral hazard on the part of the firm. If effort cannot be verified by a third party, then firms have an incentive to claim that each worker has been caught shirking, thus keeping the bond or creating a vacancy which must be filled by another fee-paying individual. Workers would then be unwilling to pay bonds, since firms cannot credibly commit not to label them falsely as shirkers.

Carmichael (1985) again suggests a problem with such a defence of efficiency wage theory against the bonding critique. In his scheme the worker still pays a bond, but the firm commits to give it to a third party, such as a charity, in the event of the worker being caught shirking. Thus, Carmichael argues, the incentive for the firm to cheat by falsely accusing a worker of shirking disappears, since it has nothing to gain from doing so. However, the worker only cares about losing his or her bond, and it is of no importance whom the beneficiary is. If the worker thought that the firm would threaten him with passing on his bond unless he accepts some unreasonable demands, then he will still not be willing to pay the bond in the first place. As Dickens *et al* (1989) ask, 'Would the reader be willing to entrust us with $10,000 even if we could not keep it but could only give it to charity? If we were well positioned to ask for a favour, it seems unlikely,' (p.339).

A more fruitful attack against the firm moral hazard defence of efficiency wage theory is the claim that firms care about their reputation and will not want to run the risk of this being damaged by claims that they are wrongly sacking workers in order to appropriate their bonds. If this were the case then workers need not fear the moral hazard problem, and should be willing to post bonds. Efficiency wage supporters could fight back by saying that only large national companies in the public eye will care about their reputation, while the lack of this reputation constraint on small, single location firms means that they will be tempted by the moral hazard problem. Thus bonding should not be possible for such small firms, requiring them to rely on efficiency wage payments to provide incentives to their workforce. This is not what is observed, however, since it seems to be precisely the large national companies who pay higher wages. The moral hazard argument must explain this observation, before it can be taken as a viable reason for the existence of efficiency wages rather than bonding schemes.

A more promising reason for explaining why bonding is not observed in practise, highlighted by Dickens *et al* (1989) but not mentioned much elsewhere, is legal restrictions on contracts. It may be that contracts including bonding clauses are simply not legally possible. This will be the case if the sums forfeited by the accused worker are far in excess of the actual damage suffered by the firm as a result of the shirking. Then the courts will view the arrangement as an attempt by the firm simply to impose a penalty on the worker, rather than gain compensation for damages suffered, and as reported in the legal references quoted by Dickens *et al*, the courts will not enforce any claims which it views as penalties.

Finally, other versions of the efficiency wage hypothesis, described below, may offer the most compelling reasons why bonding is not used. The adverse selection model says that firms pay higher wages in order to attract better quality workers. Unlike the shirking model, bonding cannot substitute for this policy. Charging entrance fees or extracting bonds would most likely have a negative impact on the quality of job applicants, since the more able would find it increasingly worthwhile to go into self-employment, rather than 'buying' jobs from employers. In the adverse selection model, therefore, bonds and high wages have opposite effects on average worker productivity, and so if applicant quality is the reason why firms pay higher, efficiency, wages, then there is no incentive for them to introduce bonding schemes.

Similarly, in Akerlof's (1982) 'gift exchange' model, adopting a policy of appropriating bonds would not have the positive effect on effort that paying higher wages would have. The generous wages are taken to be a 'gift' from the firm, reciprocated through a 'gift' of higher effort by the workers. Charging employment fees or bonds would not be viewed as generous behaviour by the firm, however, and if workers consider this auctioning of jobs to be unfair, it would likely lead to a withholding of effort. A further effect could be that bond-charging signifies a lack of trust by management in their workforce , which would reduce the latter's desire to partake in any gift exchange.

Thus the perceived fairness of bonding may well be the strongest constraint on any firms who want to introduce it. The right to work is seen as a basic privilege of mankind, and charging individuals a fee for access to this right will be seen as unfair by most. The likely result of such perceived unfair treatment will be retaliation, and withholding of effort is the best weapon in the workers' arsenal. For this reason bonding as an incentive effect remains almost exclusively in theoretical models, while firms paying wages above their market-clearing level continues to be a widely observable fact.

c.) The Labour Turnover Model

Another model, closely related to the shirking version, is the labour turnover model. Now the firm is not concerned with workers shirking, but with them quitting, which imposes turnover costs. Such costs can be direct, such as those involved when advertising a vacancy, choosing an applicant, and conducting the necessary training, as well as indirect, for example the cost of the lower productivity of the new recruit. The firm therefore raises the wage it pays, to persuade its workers not to quit. If all firms adopt this strategy, then the unemployment resulting from the higher wages provides the necessary cost of job loss that prevents quits.

Salop (1979) develops a formal model of these ideas. He differentiates between an internal labour market for experienced workers, and an external labour market for new applicants. For morale reasons, Salop claims that the firm is constrained to pay the same wage to all workers. Thus the wage once again has a dual role, attempting to clear two separate markets for two different types of workers. When it fails to succeed in both of these roles, involuntary unemployment again results.

The innovation of the model is to specify a quit rate, q:

$$q = q\left(\frac{W}{Z}\right) \qquad q' < 0 \ , \ q'' > 0$$

where W is the wage paid by the firm, and Z is an average wage in the labour market, adjusted for the probability of obtaining a job. The first derivative of the quit function shows that if the wage received was to fall relative to the outside alternative, each worker would be more likely to quit and search for another job.

In a steady state, the number of new hires by the firm, N, must equal the number of quits.

$$N = q(\frac{W}{Z})E$$

where E is the total number of workers employed by the firm.

The firm chooses the wage, W, the number of new hires, N, and the total employment level, E, to maximize profits, given by

$$\pi = f(E) - WE - T(N) - F$$

subject to the constraints

$$N = q(\frac{W}{Z})E$$

$$N \leq A(\frac{W}{Z})$$

f(E) is a non-increasing returns production function, with labour as the only input. T(N) is the cost of labour turnover, which rises at an increasing rate as the firm hires more new recruits, so that T', T''>0. The first constraint is simply the condition for a stationary state, while the second constraint ensures that there are sufficient applicants, A, written as a function of the firm's relative wage. Let λ and μ be the Lagrange multipliers on the two constraints, respectively.

Differentiating with respect to the three choice variables gives three first order conditions

$$E\colon\ f'(E) - W - \lambda q(\frac{W}{Z}) = 0 \qquad (2.12)$$

$$W\colon\ -E - \frac{\lambda}{Z}q'(\frac{W}{Z})E + \frac{\mu}{Z}A'(\frac{W}{Z}) = 0 \qquad (2.13)$$

$$N\colon\ -T'(N) + \lambda - \mu = 0 \qquad (2.14)$$

At this stage, Salop simply assumes that the firm has sufficient applicants, and so the second constraint does not bind, implying $\mu=0$. He later shows that the resulting equilibrium is possible, and thus demonstrates that involuntary unemployment (i.e. excess applicants) can be an equilibrium position in this model.

Setting $\mu=0$ in equation 2.14 gives the expression $\lambda=T'(N)$. Substituting for λ in equation 2.12:

$$f'(E) = W + q(\frac{W}{Z})T'(N) \qquad (2.15)$$

Equation 2.15 is the usual profit maximizing condition that the marginal product of an additional worker should equal the marginal cost of hiring that worker, taking into account the turnover costs incurred when that worker is hired. A similar substitution for λ in equation 2.13 gives the expression

$$E + \frac{T'(N)}{Z}q'(\frac{W}{Z})E = 0 \qquad (2.16)$$

Equation 2.16 shows clearly the wage-turnover cost trade-off facing the firm. If the firm were to raise its wage by 1 unit, its wage bill would increase by E units, the number of workers it employs. However, there would be a resulting fall in the number of quits, as shown by the derivative of the quit function, with the corresponding fall in turnover costs represented by T'(N). Equation 2.16 says that the firm should increase the wage until these two effects just offset each other.

Equation 2.16 can be written in a different way. Dividing through by E and re-arranging gives

$$T'(N) = \frac{-Z}{q'(W/Z)} \tag{2.17}$$

Substituting equation 2.17 into equation 2.15:

$$f'(E) = W - q(W/Z)\frac{Z}{q'(W/Z)}$$

$$\Rightarrow f'(E) = W(1 + \frac{1}{\epsilon}) \tag{2.18}$$

where ϵ is the quit rate elasticity (defined positive) with respect to the relative wage. The wage setting rule is therefore a variation of the usual monopsony wage formula, here including the quit rate elasticity.

The full equilibrium for the firm can be found by simultaneously solving equations 2.15, 2.18, and the constraint that N=q(W/Z)E, for the three choice variables; W, E and N. If the chosen wage for experienced workers, taking into account their propensity to quit, results in more applicants, A(W/Z), than desired hires, N, then involuntary unemployment will occur. This relies on the assumption that the firm must pay the same wage to new applicants that it pays to experienced workers. Salop goes on to show that such a scenario is possible, a fact that was used when assuming that the applicant constraint was non-binding above. The proof revolves around showing that there is no incentive for new firms to enter the market and hire the excess applicants, because at the margin an extra firm would make negative profits.

The labour turnover model is subject to the same challenge from the bonding critique as the shirking model. In particular, an employment fee could be charged, and raised until the number of applicants equalled the desired number of hires, thus clearing the external market for new applicants and removing the involuntary nature of unemployment. If all firms charged entrance fees, this would also reduce the attractiveness of quitting to find alternative work, and so the turnover problem would also be solved. Moreover, the bonding critique does not suffer the problem of possible firm moral hazard in this case. The firm is unlikely to fire experienced workers to replace them with new fee-paying employees, since the whole point of the model is that the firm is trying to reduce costly labour turnover. To the extent that workers fear that firms will hire them, take their fee and immediately fire them (i.e. the firm is in the 'application industry') then there may still be a reason why workers should resist paying employment fees. However, the legal, social and moral defences against the bonding critique, outlined above, seem to represent the most convincing arguments for why the efficiency wage, rather than the bonding, solution should exist in this case.

d.) The Adverse Selection Model

The adverse selection version of the efficiency wage hypothesis differs from the previous two examples in that it does not argue that higher wages alter the behaviour of the firm's existing workforce, in particular their propensities to shirk or quit. Rather, the link between wages and productivity is explained by arguing that a firm offering higher wages will receive applications from higher quality, and hence more productive, workers.

To make this link between wages and applicant quality, two assumptions are critical, as detailed by Weiss (1980). First, it must be the case that wages received are not proportional to a worker's productivity, otherwise a firm will just pay any worker what he or she is worth, and not worry too much about variations in quality. This assumption can be justified by the cost of regularly collecting accurate productivity data, or by risk aversion on the part of workers making variability in earnings undesirable. It could be argued that firms would then at least base wages on expected productivity at the time of hiring. Even this policy would face objections, however. First, there are the morale effects of differences in earnings, which would be accentuated if one worker turns out to be more productive than was initially expected of him, so that his wages were undeservedly low. Second, unions often push for uniform wage policies. For example, if the median worker's output level is less than the mean output level, then more than half of the workforce will be better off receiving a wage based on the mean output level, rather than based purely on their own work, and hence a democratically run union will vote for a uniform wage based on mean output.

The second critical assumption is that the acceptance wage of workers is an increasing function of their productivity. This can be justified by arguing that an individual's ability and his or her potential earnings from self-employment are positively correlated. At low wage offers, high quality workers will therefore prefer to be self-employed, but as the wage is increased, individuals of increasingly higher quality will be tempted to apply, rather than work for themselves. Alternatively, consider two firms competing for labour. If one firm's workers are made a wage offer by the other, then the first will match the wage offer to all its workers who are productive enough to make such a payment profitable, the rest being allowed to leave. If the second firm then increased its wage offer, the original firm would only find it profitable to match this offer to its most productive workers, and hence the average quality of workers accepting the wage offer of the second firm will increase, as that firm increases its wage offer.

Weiss continues with a simple model to demonstrate the effects of this link between wages and applicant quality. He specifies a function q(w), where q is the expected quality of a worker hired at wage w. By the assumptions of the model, q is increasing in w. The firm's profit function is therefore

$$\pi = pf\,[xq(w)] - wx - F \tag{2.19}$$

where p is the price of the firm's output, f is a concave production function with labour as its only input, x is the number of workers employed, and F is the firm's fixed costs. In keeping with the efficiency wage hypothesis, the production function is specified in terms of efficiency units of labour, rather than simply the number of workers employed. Profit maximising by the firm requires that it chooses the wage which minimises the cost per efficiency unit, w/q(w). Assume that this occurs at some wage $\hat{w}$. If all workers are observationally equivalent (so that the quality differences are due to unobservable characteristics) then the firm will offer this wage to all workers. To show that such a policy maximises profits, consider a firm that pays a wage w^*, not equal to $\hat{w}$, and hires x^* workers. Let $x^0=w^*x^*/\hat{w}$, then

$$f\,[x^0q(\hat{w})] - \hat{w}x^0 - F = f\,[\frac{w^*x^*}{\hat{w}}q(\hat{w})] - w^*x^* - F$$

$$> f\,[x^*q(w^*)] - w^*x^* - F$$

The inequality holds since, by the definition of $\hat{w}$, $\hat{w}/q(\hat{w}) < w^*/q(w^*)$. Hence, employing x^0 workers at wage $\hat{w}$ always yields higher profits than paying any wage w^* and hiring x^*

workers.

The common wage paid by profit maximising firms will therefore be $\hat{w}$, and this can be above the market-clearing wage. Each firm will hire workers until the usual 'wage equals marginal revenue product' condition is satisfied, found by differentiating the profit function, equation 2.19, with respect to x:

$$p\ f'[\hat{x}q(\hat{w})]\ q(\hat{w}) = \hat{w} \tag{2.20}$$

Total labour demand is determined by multiplying the number of workers hired by each firm, x, by the total number of firms, the latter being determined in competition by a zero profit constraint, taking into account the fixed costs. There is no reason why this total labour demand could not be less than total labour supply, and furthermore, those unemployed will not receive work even if they offer to work for less than $\hat{w}$, say $\hat{w}$-Δ. This would place an upper bound on their actual acceptance wage (i.e. their acceptance or reservation wage must be no greater than $\hat{w}$-Δ), and hence an upper bound on their expected productivity, $q(\hat{w}-\Delta)$. Since $(\hat{w}-\Delta)/q(\hat{w}-\Delta)$ is greater than $\hat{w}/q(\hat{w})$, the firm will not lower its wage offer to accommodate such workers. Wages therefore remain above their market-clearing level, and so this model, in keeping with the previous efficiency wage models, offers a theory of involuntary unemployment.

The model also offers an explanation of the empirically observed fact that following, say, a fall in demand, firms lay off workers rather than cut wages. The wage that minimises the cost per efficiency unit will be unchanged at $\hat{w}$. From equation 2.20, if p has fallen, the firm will hire fewer workers to ensure that their marginal revenue product equals the unchanged wage. This follows from the concavity of the production function. The shirking model is unable to make this prediction, since in the diagram above, if the labour demand curve shifts, both the employment level and the wage will change.

The adverse selection model can be extended to the case of heterogeneous labour, where workers differ in their observable as well as their unobservable characteristics. For example, assume that firms can perfectly observe the level of education achieved by an applicant. There will be a different expected quality function, $q_i(w)$, for each level of education, I, and thus each category of worker will have a wage unique to that group which minimises the cost of hiring an 'efficiency unit' from their ranks. For example, imagine that there are three different levels of observed education, each leading to a different q(w) function. Then, barring coincidences, it will be the case that

$$\frac{\hat{w}_1}{q_1(\hat{w}_1)} < \frac{\hat{w}_2}{q_2(\hat{w}_2)} < \frac{\hat{w}_3}{q_3(\hat{w}_3)}$$

Since the minimum cost of an efficiency unit is lowest for group 1 workers, firms will hire only from this group initially, paying the wage $\hat{w}_1$. If the group is not large enough to satisfy total labour demand, then competition amongst firms will bid up their wage, until the cost of hiring an efficiency unit from group 1 is $\hat{w}_2/q_2(\hat{w}_2)$. At this point, firms will also start hiring group 2 workers. Should groups 1 and 2 together provide more than enough workers to satisfy total labour demand, then hiring will stop when labour demand is satiated. All group 1 workers will be hired, but some group 2 workers may not receive a job. As for group 3 workers, hiring stops before they are even considered , and so they will all remain unemployed. The model therefore provides an explanation of varying unemployment rates across observationally distinct groups of workers. For the same reason as described above in the homogeneous case, the unemployment is still of an involuntary nature.

Finally, as pointed out in the section describing the bonding critique, this criticism cannot be directed at the adverse selection model. If firms adopted a policy of bonding rather than paying higher wages, then this would have a detrimental, rather than a positive, effect on the quality of applicants, and so the two policies are not substitutes in this case.

e.) The Gift Exchange Model

The final version of the efficiency wage hypothesis, due to Akerlof (1982), is somewhat different to the previous models, in that it relies more on a sociological explanation for the positive effect of wages on effort.

The starting point for Akerlof's paper is the observations made by Homans (1954) of the 'cash posters' at a utility company in the USA. Homans recorded that, although the company expected its cash posters to record 300 customer payments per hour, all ten workers achieved a better rate than this, with the group average being 353 per hour. Two possible explanations for this behaviour can be ruled out. First, the cash posters did not work harder for fear of being sacked, since it was made clear to them that even if they fell below 300 per hour, their punishment would simply be a mild rebuke. Second, the workers were not working harder than necessary in order to demonstrate their suitability for promotion, since all of the young women involved claimed to have no desire to be promoted anyway.

As Akerlof points out, standard neoclassical theory finds it very difficult to explain such observations. For example, consider the simple problem of a worker maximising a utility function

$$u = u(w,e)$$

subject to the constraint that he or she must work at least at the level of the firm's minimum standard, e_{min}. If u_e, the marginal utility of effort, is negative at e_{min}, then raising effort levels will reduce the worker's total utility, and hence they will choose to work the minimum required of them, and no more. To explain the behaviour of the cash posters, it would be necessary to claim that the marginal utility of effort at e_{min} is positive. If this was the case, however, a profit-maximising firm would increase the minimum work standards, since the extra effort would increase output and therefore raise revenue at no extra cost to the firm, while the workers would be quite willing to supply this extra effort, given that they do not suffer disutility from supplying effort above the level at which they currently work. Since the utility company in question had no intention of raising the work standards, this possible explanation does not seem to be an accurate representation of the situation either.

Akerlof therefore develops a sociological theory, involving group norms and gift exchange, to explain the behaviour of the cash posters. The basic idea is that workers develop sentiment for their co-workers and the institution for which they work. In common with other relationships where sentiment is involved, it is common to demonstrate this sentiment by exchanging gifts. This exchange of gifts takes the form of the firm offering a gift of wages above the workers' reservation wage, and in return the workers offer a gift by collectively setting a group norm for effort which is above the minimum level required by the firm, and to which all workers conform as best they can.

There are certain 'rules' governing gift exchange, in that the gift should be approximately in the region of the value expected by the receiver. In the context of the workplace, this leads to the idea of a fair wage; the wage that workers think would be a fair gift for the firm to offer, and which they are willing to reciprocate with a gift of extra effort. Akerlof specifies a fair wage function algebraically as

$$w_{i,t+1}^{f} = f\,(w_{i,t},\ w_o,\ b_u,\ u,\ e_i,\ e_o)$$

The first influence on what constitutes a fair wage is the level of wages that the individual received in the past, $w_{i,t}$. He or she will not only consider their own past history, but will normally also take into account the wages earned by others. Clearly, all other workers cannot be considered, therefore the individual forms a reference group and uses its members' wages, w_o, to help get an idea of what a fair wage for himself or herself would be. The reference group will usually consist of friends performing similar work, other workers at the individual's firm, and perhaps similar workers at other firms, if the individual can obtain information about their wages. Akerlof presents a lot of evidence in support of this view that individuals consider their position *relative to others* in deciding whether or not they are being fairly treated.

Not all members of the individual's reference group will be employed at any one time. If some are suffering unemployment, then the individual will take this into account when deciding how fairly he or she is being treated. Thus the proportion, u, of the reference group who are currently unemployed is an argument in the fair wage function, as is the level of income received by such people while unemployed, b_u.

The final arguments in the fair wage function take account of the rate of effort which the individual has to expend at work, e_i, and the amount of effort that members of the individual's reference group input, e_o. The higher the effort level required, the higher must the wage be before it will be considered fair.

Akerlof then describes a group norm for effort function:

$$e_n = e_n\{[w(e,\epsilon)],\ e_{\min},\ u_1,\ldots..,\ u_J;\ w_o,\ u,\ b_u\}$$

The group norm for effort is therefore specified to depend on the wages paid by the firm to a particular type of worker, ϵ, working at an effort level e, relative to the income of the reference group members, summarized by w_o, u and b_u. Thus the degree to which the wage schedule, $w(e,\epsilon)$, is considered fair, by the criteria described above, determines the size of the workgroup's reciprocal gift, that is the group norm for effort. To complete the group norm function, Akerlof includes the utility levels of all J workers at the firm, and the work standards, e_{min}. The last term may have a negative effect on the group norm; the harder the firm pushes the workers, the less willing they are to work.

The efficiency wage nature of the model is clear. The firm can choose the level of wages, the work rules and the number of workers it wishes to employ. If it raises the wage it pays, it will increasingly be viewed as 'fair,' which in turn will lead to a higher group norm for effort, as the workers reciprocate in the gift exchange process. A firm may have no incentive to hire an unemployed worker offering to work for less than the firm's current wage, since this may upset the harmony within the firm that makes reciprocal gift exchange possible.

Akerlof proceeds with two examples which assume specific functional forms, to demonstrate the usual efficiency wage result that involuntary unemployment can exist within the framework of the model. The first example assumes that the work rules are given, and that all workers choose to work at the group norm level of effort. Akerlof shows that a wage above the market-clearing level can result. The second example proceeds in a similar way, except that the wage is taken as given, leaving the firm to vary its work rules. Again, involuntary equilibrium unemployment is shown to be possible.

Finally, as with the adverse selection model, this version of the efficiency wage hypothesis is not subject to criticism from the bonding critique. If the firm was to replace its policy of paying wages above the minimum necessary with a policy of charging entrance fees, this would almost certainly be viewed as unfair by the workers, who would therefore withdraw their gift of extra effort. Thus the two policies are not substitutes, in the sense that they have opposite effects on effort.

Therefore, although the various efficiency wage theories have different starting points and model different aspects of firms' behaviour, they all make the prediction that increasing the wage paid can be profitable for a firm, in terms of higher effort or productivity. The next section describes some of the attempts that have been made to test this prediction empirically.

2.3 : Testing the Efficiency Wage Hypothesis

Efficiency wage payments are modelled as being paid by firms to influence the effort choices of their employees, principally because they cannot observe the effort choice perfectly and so cannot write a direct contract specifying a particular effort level required. The econometrician attempting to test the theory faces a similar problem; data which detail the exact effort expended at work by individuals does not exist. The aim of this section is to describe the approaches which have been used in the literature to examine the relevance of efficiency wage theory, while circumventing this data problem.

a.) Industry Wage Differentials

Studies that have described and analysed the existence of significant inter-industry wage differentials probably represent the most indirect evidence in favour of efficiency wage theory, essentially providing evidence 'by process of elimination.' The key studies in this area are those by Katz (1986), Dickens and Katz (1986), Krueger and Summers (1986, 1988) and Groshen (1991). The first four calculate industry wage differentials from various large microeconomic data sets, while the last paper mentioned reviews the available evidence.

To determine the size of inter-industry wage differentials, the first step is to add industry dummy variables to a standard wage equation which already includes all of the usual controls for human capital, region and so on. Katz (1986) includes 1-digit industry dummies, while the remaining papers adopt a broader range, with Krueger and Summers (1988) even taking the analysis to the 3-digit level. Once the coefficients on the industry dummy variables have been estimated, their employment-weighted average is calculated, treating the omitted dummy as having a zero effect on wages, since a constant is usually included. The difference between the actual coefficient on a particular industry's dummy and this average is then the estimated wage differential for that industry, as reported in these studies. The differential therefore estimates the proportional difference in wages between an employee in that industry and the average employee in the sample, after controlling for individual characteristics.

The inter-industry wage differentials estimated by this procedure have been found to be large and significant, even when detailed controls for individual characteristics are included. Typically, they range from approximately 40% above the average to 40% below, implying a standard deviation of about 11-14%. Further analysis, best described in Krueger and Summers (1986), has shown the distribution of estimated differentials to be extremely regular. For example, across time, the correlation coefficient between industry wages in 1984 and a range of other years has been consistently above 0.75 for all years since 1920, and even for years before then, the correlation is always at least 0.6. Similarly, across countries, the correlations between the industry wage structures in various developed countries are typically above 0.7, rising to 0.95 between the UK and the US. Other regularities that have been found are that the inter-industry wage structure is remarkably similar when comparing young with old workers, short tenure with long tenure workers, and blue collar with white collar workers. This last comparison is greatly expanded upon by Katz (1986), who examines a large number of occupations. He finds that the industry wage structure is highly correlated for all pairs of occupations, with most of the correlation coefficients lying between 0.7 and 1. Hence, if one occupational group in an industry is well paid, it is likely that all occupations in that industry will also be well paid.

Another line of investigation usually taken is to examine the correlates of the industry wage differentials, that is, seeing how they vary with industry characteristics. The available evidence is well summarised by both Krueger and Summers (1986) and Dickens and Katz (1986). The latter add to this area of research by performing an extensive analysis of correlates, trying 432 specifications in all. The key results seem to be that industry wages are affected by ability to pay and unionisation rates. Evidence for the former assertion is that more concentrated industries (that is industries containing firms with more market power) tend to be high wage industries, although some studies have found this result to be sensitive to the degree to which labour quality is controlled for (presumably due to firms using some of their market power to attract better quality workers). Also, industry wage differentials are typically found to be positively correlated with profit rates. Further results show that industry wages are negatively related to labour's share of costs, and positively to the industry capital-labour ratio. As Krueger and Summers (1986) point out, these last results are consistent with the Marshallian conditions of factor demand elasticities, whereby the smaller is the share of factor payments going to labour, and the higher is the proportion of capital, the less the firm is affected by wage increases, and so the more able it is to pay higher wages.

As for unionisation rates, these are consistently found to be positively related to industry wage differentials. However, there is a question concerning causality in this relationship. Instinctively, it is usually argued that the higher the unionisation rate, the more powerful are unions, and so the more able are they to obtain higher wages. The causality could equally run in the opposite direction, though, with unions being attracted to industries which already have high wages for other reasons, for example firms with market power.

Armed with this evidence concerning industry wage differentials, each of the studies under consideration then attempts to describe which theories of the labour market are consistent with these findings. This analysis is perhaps best described by Krueger and Summers (1988) and Groshen (1991), although all five papers arrive at broadly the same conclusions. The starting point is usually to dismiss competitive explanations for the industry wage differentials. First, one possibility is that the wage equation results are biased because of unobserved quality differences between workers, so that the high wage industries are simply the ones with the high (unobserved) quality workers. If this was the case, however, then including observed quality differences should substantially reduce the differentials (on the reasonable assumption that observed and unobserved quality will be strongly correlated), which is not what is observed in practice. Further evidence comes from longitudinal studies, which follow workers from one industry to another. The differentials calculated using such data sets are very similar to the cross-section differentials described above, so that if a worker leaves a low (high) wage industry for a high (low) wage industry, he will experience a wage rise (fall) similar to the estimated inter-industry wage differential. This is the case, even when only involuntary displacements from one industry to another are considered, thus not allowing self-selection by workers into appropriate industries for their skills. Assuming such workers' unobserved skills are remaining constant as they move across industries, then competitive theories cannot explain these wage differentials. The final argument against such competitive reasoning is the similarity of the industry wage structure across occupations. It is very unlikely that all skills will be equally valued in all occupations, so that it does not appear that skill differences are the reason for the existence of industry wage differentials.

Another competitive explanation of the industry wage differentials is that they are being caused by the existence of compensating differential payments. Some industries may have less attractive working conditions, in which case they need to offer higher wages in compensation, so that workers in that industry still receive the market rate of utility. A failure to control for working conditions could therefore be the reason for the observed differentials. However, when investigators have included a range of working conditions in their wage equations, the pattern of industry wages barely changes. Indeed, Krueger and Summers (1988) find that the standard deviation of industry wage premiums actually increases following the inclusion of working conditions. Further evidence against the compensating differentials argument is supplied by the fact that high wage industries have significantly lower quit rates than their low wage counterparts. If compensating differential payments were indeed the reason for the high wages in some industries, then such industries should not exhibit fewer quits since they would still only be offering the market rate of utility. Finally, the finding that the industry wage structure is highly correlated across occupations can be used again, since it is unlikely that working conditions will vary in the same way across all occupations. The petroleum industry may be a high risk industry requiring compensating payments to those working on oil rigs, but surely city secretaries are no more at risk working in the petroleum industry than any other industry.

Another possible explanation for the existence of industry wage differentials is that they reflect differences in human capital accumulation. If certain industries have workforces with more accumulated human capital, then they may have to pay higher wages, rather than risk losing such workers. Again, however, the evidence is against such an explanation. As mentioned above, Krueger and Summers (1986) show that the estimated industry wage structure is very similar for both young and old workers, both short tenure and long tenure workers, and both blue collar and white collar workers. Within each of these pairings, the degree of human capital accumulation will be different and yet the industry wage structure is highly correlated, suggesting that differences in human capital accumulation cannot be the reason for the observed industry wage differentials.

A fourth suggestion which has been offered is that the observed differentials are only transitory. For example an increase in labour demand in one industry may lead to higher wages in that industry, which may persist in the short run and not be immediately competed down again, if labour is not perfectly mobile, or if it takes time for firms to re-design their wage structures. However, given that the industry wage structure has been remarkably consistent over the last seventy years or so, this clearly rules out the idea that the differentials are only temporary, and therefore they are not caused by short-term persistence following a shock. Also, if such shocks are hypothesised to be random, then it would be difficult to explain the regular correlations between industry wages and industry characteristics, noted above.

Having ruled out possible competitive explanations for the observed pattern of industry wages, the authors then turn to non-competitive arguments. Katz (1986) and Dickens and Katz (1986) argue that union threat models are relevant, whereby industries whose firms find it more difficult to prevent unionisation pay higher wages in an attempt to pre-empt unionisation. Clearly, the finding that the industry wage structure is correlated with the degree of unionisation is consistent with this reasoning. However, Krueger and Summers (1988) point to evidence which is less supportive of the theory. First, the wage structure has been very similar over a long period of time, which has included periods of large changes in union behaviour. Second, the wage structure is also similar across different areas, so that the wage structure in the northern states of the US (where the threat of unionisation is high) is highly correlated with its counterpart in the southern states, where firms face a much smaller risk of unionisation. Similarly, the wage structure is very similar across countries, whether they have a reputation of being very tough or very sympathetic to unions.

Thus, by a process of elimination, most authors arrive at efficiency wage arguments to explain the existence of industry wage differentials. If monitoring technologies or the costs to the firm of employee-shirking differ across industries, then some industries will be more likely to adopt efficiency wage strategies than others, thus explaining why some industries pay high wages. One problem with this reasoning is the finding that the industry wage structure is correlated across occupations. It is unlikely that an industry would face the same problems monitoring all the different occupations that it employs. In answer to this point, it has been argued that 'fair wage' constraints are also at work, so that if an industry pays a high wage to employees in some occupations, it must pay a high wage to all employees. Thus, most authors argue that the observed industry wage structure is explained by a combination of the shirking and the sociological versions of the efficiency wage hypothesis. The fact that the wage structure is correlated with 'ability to pay' characteristics leads some others to mention rent-sharing theories, but as Krueger and Summers (1988) point out, such ideas are closely related to efficiency wage theory, since firms presumably share rents in order to prevent shirking or to appear to behave 'fairly' towards the workforce.

Clearly, the main problem with this evidence in favour of efficiency wage theory is its indirect nature. The available evidence does not prove efficiency wage payments are made, but only rules out alternative theories. Other tests of the efficiency wage hypothesis attempt to provide more direct evidence.

b.) Testing Implications of the Efficiency Wage Hypothesis

One way around the problem of obtaining effort data that has been adopted in the literature is to test implications of the theory, rather than test the central proposition that effort will rise as wages rise.

For example, Cappelli and Chauvin (1991) argue that since higher wages, and also a higher local unemployment rate, make shirking less attractive, the firm will not be forced to sack as many of its workers for poor performance, as these variables rise. Thus, they expect the dismissal rate across firms to vary negatively with the wage premium that they pay and with the unemployment rate in their local areas.

The data which they use come from the records of a single company, with plants across the US. The fact that all observations are from the same company means that a number of factors are standardised, which otherwise might have been important determinants of worker performance or dismissal rates. For example, all data are for production workers, doing virtually the same job with the same equipment, within a given category of plant. Also, management practices concerning dismissal procedures, as well as the union response to such practices, are centrally determined and therefore consistent across plants. The wage paid is also centrally determined, which means that it will not be influenced by the amount of effort supplied by the workers at any particular plant, and so is not subject to any problems of endogeneity. However there is still variation in the wage premium within the sample, since the average local wage to which the actual wage is compared shows significant variation across the plants.

Other factors that are not constant across plants, but which may influence the dismissal rate, are taken into account. For example, whether the plant is an assembly plant or not, whether the workers are more militant (using dummy variables for areas of the country with a history of militancy), and whether the local unions have important 'voice' effects (by including the management-rated problem solving ability of the local union) are all considered. Thus, although testing implications of the theory can be problematic, given that there could be other interpretations of the implied relationship under investigation, the strength of Cappelli and Chauvin's data set is that it controls for virtually all conceivable factors, making alternative interpretations difficult.

The results show that dismissals are indeed significantly negatively related to the wage premium. The local unemployment rate also has the predicted negative coefficient, but in the case of this variable, the coefficient is insignificant. Efficiency wage theory is therefore supported in that workers are less likely to shirk when they receive a higher wage premium, but the argument that a rise in local unemployment raises the cost of job loss and so makes workers less likely to shirk does not find much support.

Drago (1993) also examines the determinants of dismissals, looking in particular at the state of the labour market, rather than wages. However, he has the opposite prior belief to Cappelli and Chauvin concerning the effects of this variable on dismissals. Cappelli and Chauvin consider the workers' point of view, and argue that a rise in the local unemployment rate will make employees more fearful of losing their job, so that they are less likely to shirk, thus reducing the dismissal rate. Drago, on the other hand, looks at things from the firms' side, and argues that as the unemployment rate rises, dismissal policies will become more effective in providing incentives not to shirk, and so firms will substitute away from paying high wages towards using dismissals more, thus increasing the number of dismissals. Thus one of the problems with testing implications of efficiency wage theory is immediately revealed, in that authors can derive different, and in this case opposite, implications, making interpretation of the results much more difficult.

Drago extends the analysis of Cappelli and Chauvin by allowing for a zero dismissal equilibrium, which can occur if no shirking actually takes place. For example if effort requirements are low, or if firms use alternative incentive schemes when the cost of replacing workers is high, a zero dismissal rate can emerge. Drago therefore adopts the Heckman (1979) sample selection procedure, first estimating the determinants of whether a firm chooses to use a policy of dismissals, and then the determinants of the actual dismissal rate, given that dismissals are used.

To analyse the impact of local labour market conditions, Drago makes use of two variables in his data set (the Australian Workplace Industrial Relations Survey), namely whether there are skilled workers present at the workplace, and the manager's opinion of whether demand for the firm's product has risen in the previous year. Affirmative answers to these questions should indicate that a firm's workers will be in demand, and so these variables proxy the inverse of the relevant unemployment rate. Other variables included by Drago are various controls, for the required effort level at a workplace (capital intensity), the cost of dismissing workers (presence of a union, proportion of union delegates per union member, and whether or not the workplace is operating at full capacity), the use of alternative incentive schemes (performance-related pay and promotion incentives), and for possible mismatches which may affect dismissal rates (by examining the tenureship of dismissed workers).

The results show that having skilled workers and having a rising product demand are both significantly negatively related to whether or not a firm adopts a policy of dismissals, indicating that a firm is more likely to use dismissals when alternative employment conditions for its workers are poor, exactly as predicted by Drago. However, in the equation for the dismissal rate, conditional on dismissals actually being used, both variables have insignificant coefficients, and the coefficient on the product demand variable actually takes the wrong sign. This could be caused by Cappelli and Chauvin's arguments having an offsetting effect, so that as unemployment rises, firms want to use dismissals more, but workers become more fearful of losing their jobs and so are less likely to shirk, so that the observed dismissal rate can actually fall even though the firm has tightened its dismissal policy.

A final issue considered by Drago is the split between primary and secondary sectors, which also appears in several other studies to be described below. He includes a variable measuring the employment level at each workplace, on the assumption that larger workplaces will belong to the primary sector. This is a somewhat arbitrary assumption, since primary sector status is more accurately determined by wages and working conditions, but it has been used by other authors. The results are somewhat confusing, showing that large (primary sector) firms are more likely to use dismissals, but have a lower actual dismissal rate. This issue will be discussed further at the end of this section.

Campbell (1994) also examines dismissals, but again using a different estimation technique. The Employment Opportunity Pilot Project data which he uses are at the firm level, but include data on the most recently hired worker (including whether he or she is still at the firm, and if not, the reason for the departure). Therefore Campbell can look at the individual probability of being dismissed, rather than examining the determinants of dismissal rates. He does this by estimating a hazard model of the form:

$$\lambda(t) = \theta \exp[x'\beta]\lambda_0(t)$$

where θ is a heterogeneity parameter and $\lambda_0(t)$ is the baseline hazard. The latter is estimated non-parametrically, to allow for heterogeneity between workers. The results show that the hazard does indeed decline with employment spell. This could occur for a number of reasons, for example experienced workers are trusted more and so monitored less closely, or older workers become more skilled at avoiding detection, or long tenure workers are the ones with the lowest disutility of effort, their high disutility contempories having been dismissed long ago.

x is a vector of individual and firm characteristics, with estimated coefficients, β. The key variable of interest is again the individual specific unemployment rate, with Campbell including two possibilities, the unemployment rate in the individual's local area, and the individual's industry unemployment rate. Wages are again not examined because of endogeneity problems. The other explanatory variables are controls for personal characteristics, occupation, region (since different states have different dismissal laws), the cost of monitoring (proxied by firm size), the capital-labour ratio of the firm, and the proportion of the firm's workforce that is unionised (again allowing for different dismissal procedures). Campbell therefore has a very rich data set, including both individual level and firm level characteristics that can influence the probability of dismissal.

Campbell forms his hypothesis from the workers' point of view, as in Cappelli and Chauvin (1991), and expects the unemployment rate to have a negative effect on the probability of dismissal, since workers will be more fearful of losing their job the higher is the unemployment rate, and so will be less likely to shirk. This predicted negative coefficient is obtained on both unemployment rates used, significant at least at the 5%

level. The elasticity of fires with respect to the industry unemployment rate is estimated to be in the range -0.45 to -0.76. The elasticity with respect to the local unemployment rate is more erratic (from -0.70 to -6.71), because of the larger standard error on this estimated coefficient, presumably caused by multicollinearity between the local unemployment rate variable and the state dummies.

Moving on from dismissals, Leonard (1987) examines supervision, in particular, examining the prediction of the shirking model that wages and supervision are substitutes when it comes to effort extraction and so should be negatively correlated across firms. He therefore examines particular occupations, notes that wages are not constant across firms, and tests to see whether the various levels of supervision in the different firms can explain this variation in wages. Thinking of wages and supervision as two inputs to the effort extraction, and hence the production, process, it is important to control for the level of activity, so that the analysis is capturing movement along an isoquant. If output is not held constant, then larger scale firms may have both higher supervision and higher wages, biasing the analysis against finding a negative correlation. Leonard therefore includes the number of workers as an explanatory variable, as a proxy for output.

Leonard estimates a wage equation for six different occupations with the ratio of supervisors to employees as a explanatory variable, representing the degree of supervision. In none of these six equations does the supervision variable attract a significant negative coefficient, as predicted by efficiency wage theory. One potential problem with this analysis is that the number of supervisors will be a choice variable for the firm, and so failing to take account of this endogeneity will lead to simultaneous equation bias. If the wages of production workers rise, then the firm may well substitute towards supervisory staff. This reverse causality between the two variables therefore suggests a positive relationship, and so is likely to bias the results against efficiency wage theory.

Leonard attempts to get around this problem by performing a supplementary analysis, regressing the ratio of supervisors to workers against the ratio of their prices (that is the ratio of supervisory wages to production wages). Standard theory would predict a negative relationship, so that as the ratio of prices falls, firms substitute towards supervisory staff. Efficiency wage theory, on the other hand, predicts a trade-off between wages and the level of supervision, so that a rise in the wages of production workers (a fall in the ratio of supervisory to production wages) should be associated with a reduction in the degree of supervision (a fall in the ratio of supervisory to production staff). Efficiency wage theory therefore predicts that the ratio of wages should attract a positive coefficient. However, when Leonard estimates such regressions for his six occupations he finds only one positive coefficient on the ratio of wages, and even that is insignificantly different from zero. He therefore again concludes against efficiency wage theory. However, it is not clear that this new analysis really solves the endogeneity problem. He is now using the ratio of wages as an explanatory variable, and wages are clearly choice variables for the firm. The results are therefore still likely to be affected by simultaneous equation bias.

A paper by Groshen and Krueger (1990) attempts to avoid the endogeneity problems in Leonard's framework, by using a particular data set, the Bureau of Labor Statistics' 1985 Hospital Industry Wage Survey. This data set contains data on the employees of 300 different hospitals, from which Groshen and Krueger concentrate on four particular occupations; radiographers, physical therapists, registered nurses and food service employees. The authors argue that supervisor to staff ratios are set to a large extent by local regulatory authorities, particularly for nurses, so that this variable is largely taken out of the control of the individual hospitals, so that it is no longer a choice variable at the workplace level. Groshen and Krueger do not have data on the actual supervisory levels stipulated by the local authorities, but they do know under which authority's jurisdiction each hospital lies. Their method is therefore to estimate wage equations in a two-stage least squares framework, treating supervisor to staff ratios as endogenous, and omitting indicators of the appropriate local authority, to provide the exclusion restrictions necessary for identifying such wage equations.

The results show that in a straightforward OLS regression of the type estimated by Leonard, the supervisor to staff ratio attracts an insignificant coefficient in the estimated wage equation for all four occupations, but, in the case of nurses, it has a significantly negative coefficient when instrumented by the procedure described. The remaining occupations all continue to supply insignificant results even when the instrumenting procedure is carried out, although, as Groshen and Krueger point out, the local authority regulations are less rigid for these occupations, and indeed the test of the identifying restrictions suggests that they are not valid in these cases. Groshen and Krueger therefore conclude that, when valid instruments can be obtained for the supervisor to staff ratio, as in the case of the nursing profession, this variable is negatively associated with wages, as predicted by efficiency wage theory.

A third paper to study the relationship between wages and supervision is Arai (1994). The central focus in his paper is a point mentioned by Groshen and Krueger (1990), that although compensating differential theory and efficiency wage theory both predict a positive relationship between wages and effort, they can be distinguished by their predictions concerning the relationship between wages and supervision. Efficiency wage theory says that the two are substitutes in the effort extraction process and so suggests a negative relationship between them. Compensating differential theory, on the other hand, predicts a positive relationship between the two variables, assuming close supervision is an unpleasant working condition requiring higher wages in compensation. In making this point, Arai's is one of the very few studies to have explicitly considered compensating differential theory, and to attempt to differentiate between this and efficiency wage theory.

Using individual-level data from the 1981 Swedish Level of Living Surveys, Arai defines a dummy variable which takes the value of one if the individual has an autonomous job, defined as one where the worker has the possibility of deciding the work pace, does not use a punch (time-keeping) card, *and* has flexible working time. Such a job is presumably not closely supervised, so this autonomy variable represents the inverse of the degree of supervision. When Arai includes this variable in a wage equation, controlling for human capital, the use of alternative pay systems, plant characteristics and occupation, it is found to have a positive coefficient in the private sector equation, thus implying a negative relationship between supervision and wages, as predicted by efficiency wage theory, but a negative coefficient in the public sector, in support of compensating differential theory. Arai explains this difference across sectors as being due to a lack of possibilities and incentives to pay efficiency wages in the public sector, because of reasons such as non-observability of output or lack of profit motives. He therefore concludes that efficiency wage theory is relevant to wage determination, but only in the private sector.

The final paper to be discussed in this section is that by Rebitzer and Robinson (1991), which also considers the relationship between wages and supervision, but takes a different approach to the previous papers. The starting point is the same, that supervision and wages are substitutable inputs in the effort extraction process. The authors then argue that monitoring will be more difficult in large firms, so that such firms will substitute higher wages for close supervision, leading to a prediction of the shirking model that wages should increase with firm size.

Rebitzer and Robinson extend the analysis by testing the Bulow and Summers (1986) proposition that only primary sector firms set wages above their market-clearing level and use the threat of dismissal to prevent shirking. Secondary sector firms, so the argument goes, can monitor their workforce perfectly, and so have no need for such incentive schemes. Thus, the predicted relationship between wages and firm size should only be observed in the primary sector. To avoid arbitrary classifications between primary and secondary sectors, the authors use a switching regression technique, which allocates each observation to the sector to which it is best suited, in probabilistic terms, based on the characteristics of the individual. Thus three equations are estimated, the switching equation, and the wage equations for the two sectors. Since some individuals in the 1983 Current Population Survey work for multi-plant firms, the authors include both measures of plant size and firm size (in terms of number of employees) in the wage equations. Other variables in the wage equations include human capital controls and a measure of union density.

The results reveal that wages do indeed increase monotonically with plant size in the primary sector, with all coefficients being significantly greater than zero. Similarly, larger firms seem to be associated with higher wages in the primary sector. However the results are less consistent than those for plant size, since the size of the coefficients does not increase monotonically and also two of the four coefficients on the firm size dummy variables are insignificantly different from zero. These results could be taken as tentative support for the 'monitoring difficulties' interpretation, since such difficulties would be expected to be present more particularly in large plants than in large firms. However, there is still some firm size effect which needs explaining.

As hypothesised, the plant size effects are less clear in the secondary sector. The first dummy variable (25-99 employees) actually has a significantly negative coefficient, indicating that such firms pay lower wages than the smaller size plants in the omitted category (less than 25 employees). After this, however, the coefficients are all positive, and although they do not increase monotonically, the two largest plant size coefficients are significant at the 10% and the 5% level respectively. In general the coefficients are smaller than the respective ones in the primary sector. As for the firm size coefficients in the secondary sector, all are positive but only one is significant, and again they are, in general, smaller than those in the primary sector.

There is therefore evidence that plant and firm size have more effect on wages in the primary than in the secondary sector. This in turn would seem to support the assertion that primary sector firms face monitoring difficulties and so pay higher wages, together with dismissal threats, to provide incentives against shirking. It should be noted, however, that the original argument said that the secondary sector faced no such monitoring difficulties, while the evidence above suggests that plant size has at least some effect on wages in the secondary sector as well.

The seven studies described in this section therefore provide some illuminating evidence, with all but one, Leonard (1987), finding in favour of efficiency wage theory. The main problem with testing implications of the theory is that other theories may lead to the same implications, making it difficult to differentiate between them. For example, the negative correlation found between wages and supervision by Groshen and Krueger (1990), and in the private sector by Arai (1994), may be consistent with efficiency wage theory, as well as ruling out compensating differential theory, but it could equally be explained by an unobserved ability story, and so be consistent with standard human capital models. In particular, assuming that the more able require less supervision and earn higher wages, this will bias the results towards finding a negative correlation, to the extent that ability is not fully controlled for.

Similarly, with Rebitzer and Robinson's (1991) study, other reasons could be, and have been, put forward to explain the positive association between wages and firm size. First, such effects can be explained without departing from competitive theory, for example, if employment in large firms is more regimented, requiring higher wages in compensation. Other explanations lie outside competitive reasoning, for example, if large firms have product market power, and if they share some of these rents with their workforce, then wages would be higher in large firms. Sociological efficiency wage theories could offer another reason for firm size wage effects, if large employers are more worried than their smaller counterparts about the adverse effect of using dismissal threats on worker loyalty, and so are more prone to use gifts of higher wages to promote loyalty. If such arguments are correct, then since the firm size effects are more evident in the primary sector, this would indicate that primary sector firms are ones which share rents or offer gifts of higher wages to their workers, rather than the Bulow and Summers' (1986) idea that they are the ones which use dismissal threats. The former interpretation would seem more consistent with the observation that quit rates are lower in the primary sector.

c.) Productivity Data as a Measure of Effort

The potential problems of interpretation when testing implications of efficiency wage theory have led some authors to attempt more direct tests of the theory. The basic premise of efficiency wage models is that effort rises with the wage paid to workers. The 'shirking' model described above also makes the prediction that an individual's effort should rise with the degree of unemployment in the labour market specific to that individual. A higher unemployment rate implies an increased chance of entering a spell of unemployment, and hence reduced income, following a termination of the current job, so that the costs of dismissal to the individual are increased, giving him or her more incentive to supply effort rather than shirk. A direct test of the efficiency wage hypothesis therefore involves examining whether effort does indeed rise with wages and unemployment.

Of course, the problem is that effort is not directly observable. One solution to this data problem has been to use a measure of productivity as a proxy for effort. Such an approach has been followed by Rebitzer in two papers (Rebitzer [1987,1988]), both of which focus on the 'unemployment effect.' Both papers use essentially the same data set, pooling data across 2- digit industries in the US for the years 1960-1980. Rebitzer uses a simple OLS estimation technique to regress the rate of growth of productivity (output per labour hour) in the industry against the civilian unemployment rate. To allow for other factors which may influence productivity growth, Rebitzer includes a range of other explanatory variables as controls, such as the rates of growth of capital services and labour hours employed, the rate of change of capacity utilization in the industry, the average age of the industry's capital stock, and the fraction of the capital stock that is new investment (all of which are intended to control for the quality of the capital services used), the fraction of the industry's workforce that belongs to a union (to allow for possible union restrictions having an effect on productivity growth), and finally dummy variables for the periods 1966-72 and 1973-80 (allowing for secular trends in productivity growth).

The results show that the unemployment rate, whether included as a level or a change, has a significantly positive effect on productivity growth, as predicted by efficiency wage theory. A supplementary hypothesis examined by Rebitzer is that this unemployment effect should be smaller in industries with long-term employment relations (LTERs). This is confirmed by an interaction term between the unemployment rate and a measure of the prevalence of LTERs (proxied by the standardised average job tenure in each industry) attracting a significantly negative coefficient. The intuition behind this result is that if LTERs are present, whether caused by implicit contracts or the presence of a trade union, then the firm may be unwilling, in the former case, or unable, in the latter case, to respond to higher unemployment with a tightening of its dismissal policy, so that the unemployment threat is reduced, and so the 'unemployment effect' on workers' effort is smaller.

The LTER result adds more evidence to the primary sector/secondary sector debate above. On the reasonable assumption that firms with LTERs, as measured by high job tenure, will be primary sector firms, Rebitzer's study suggests that primary sector firms are less likely to use dismissal threats to elicit work effort. His results are therefore contrary to the predictions of Bulow and Summers' (1986) theoretical model, and also the interpretation of their empirical results put forward by Rebitzer and Robinson (1991), (although the alternative, rent-sharing, interpretation of their results would have been more consistent with Rebitzer's (1987) results).

A third paper to use productivity measures to proxy effort is that by Weisskopf (1987). This study is unique amongst the papers under review here, in that it offers crosscountry comparisons by estimating a series of equations for different countries. Weisskopf is more careful than Rebitzer to explain why he uses the *rate of growth* of productivity as a dependent variable. First he uses standard arguments, although given from a neo-Marxist perspective, as to why a higher *level* of unemployment should increase the level of productivity. The higher unemployment rate puts capitalists in a stronger position relative to the workers, and so they can extract more effort, using dismissal threats, as found in standard shirking models. Thus, with the rate of growth of productivity as a dependent variable, the rate of growth of unemployment should have a positive effect for this reason.

The additional analysis supplied by Weisskopf is that the *level* of unemployment can also affect the *rate of growth* of productivity. The direction of this effect cannot be signed *a priori*, however. A positive relationship could be found, if capitalists are concerned about the strong position of workers when unemployment is low, and so invest more in technology which increases their control over the workforce, rather than in technology which offers efficiency improvements. Additionally, the power given to them by a low unemployment rate may be used by the workers to prevent technical progress. For both of these reasons, which are typical of adversarial industrial relations systems, the level of unemployment would be expected to have a positive effect on technical progress and hence on the rate of growth of productivity. However, in a more harmonious industrial relations setting, low unemployment need not be used by the workers to prevent innovation, and firms need not fear losing control over their workforces. Only when unemployment is high might workers fear the employment implications of the adoption of new technology, and so begin to resist such innovation. Hence, in such settings the level of unemployment may exert a negative influence on the rate of growth of productivity.

To test these theories, Weisskopf uses time series data from 1958 to 1985 on the manufacturing sectors of eight countries. The dependent variable is the annual logarithmic rate of growth of average labour productivity. To examine the short-run, work intensity, effect, the annual change in the unemployment rate is included as an explanatory variable, while the moving average of the level of unemployment in the previous five years is used to capture the longer-term, innovation effects. As in Rebitzer's studies, additional explanatory variables are included, to control for other influences on the rate of growth of productivity. These are the current and lagged values for the capacity utilisation rates, the annual logarithmic rate of change of capital intensity, the change in the average age of the capital stock, and the annual logarithmic rate of change of the relative price of external inputs, averaged over the previous three years.

Looking first at the variable measuring the change in unemployment, measuring the short-run, work intensity effect, the results show this to have a significantly positive coefficient in only three of the eight countries under consideration; Britain, Italy and the United States. In the remaining five countries, France, Sweden, Japan, Canada and West Germany, there is no evidence of an unemployment effect on productivity. This may be explained by better employment relations in those countries, following Rebitzer's arguments above that the unemployment effect is lessened by the presence of LTERs.

As for the longer-term innovation effect, the unemployment level moving average only takes a significantly positive coefficient in the United States, suggesting industrial relations are most adversarial in that country. As hypothesised by Weisskopf, countries at the other end of the industrial relations spectrum may only have innovation blocked when unemployment is high, and such a significant negative effect of the level of unemployment on productivity growth is indeed found particularly in West Germany and Sweden, but also in Canada and France. The remaining three countries, Britain, Italy and Japan, have negative but insignificant coefficients.

By analysing data for eight different countries, Weisskopf therefore shows that there is more to the unemployment-effort relationship than hypothesised in the basic shirking model, and that the unemployment effect can depend on national institutions, in particular the state of industrial relations. The analysis shows that in systems such as in the United States, large reductions in unemployment may be difficult given the associated falls in effort and productivity growth that will result. More harmonious industrial relations systems, such as those found in Germany and Sweden, may be more consistent with low unemployment, because of the related high productivity growth predicted by Weisskopf's results.

The final paper under review to use productivity data is that by Holzer (1990). However, this paper differs from the previous ones in that it uses a subjective productivity rating given to the worker by his or her manager, rather than an objective measure of productivity in terms of physical units of output, as the dependent variable. The advantage of such data, taken from the Employment Opportunity Pilot Project surveys of 1980 and 1982, is that data with a lower level of aggregation can be used, in this case at the level of individual firms, including data on the most recently hired individual. The disadvantage is obviously the subjective nature.

Holzer analyses the effect of wages on a whole range of indicators of firm performance, for example hiring and training costs. For brevity, only the productivity results will be discussed here. Holzer makes allowance for the endogenous nature of the (firm-level) wage variable by estimating the model using two-stage least squares. The crucial assumption is that certain variables in the wage equation can be excluded from the productivity equation, and hence can be used to instrument wages. Holzer tries four different specifications, each with different exclusion restrictions. In the first he excludes the proportion unionised, industry and firm size dummy variables; in the second only proportion unionised and firm size; in the third only proportion unionised; and finally in the fourth only firm size. The justification offered for all of these exclusion restrictions is that individual effort decisions should not be affected by firm or industry characteristics, which should therefore play no part in the productivity equation. The omission of unionisation from the productivity equation can however be questioned, since a number of studies, for example Gregg *et al* (1993), are devoted to examining this very relationship.

The results show that effort, as measured by the productivity score, increases significantly with the wage paid by the firm, in all specifications except that which examines the firm-size wage effect, (that is the fourth specification, when the firm size dummy variables represent the only exclusion restrictions). Thus efficiency wage theory is, in general, supported.

The main question mark hanging over all the studies in this sub-section is that productivity may not be a very good measure of effort. Effort is only a single input to the production process, used with many other inputs, which together determine output, and hence productivity. Therefore, if a variable is found to affect productivity, it is not clear whether this influence is via effort, or one of the other inputs. This point is made clear by the Weisskopf (1987) study described above. While the effect of unemployment on the level of productivity is predicted to work through its influence on workers' effort, Weisskopf's theory explicitly states that unemployment will affect the rate of growth of productivity via capital accumulation effects, which have nothing to do with effort. The problem of interpreting the coefficient on wages in a productivity equation can be even more severe, which is the probable reason why three of the four studies considered above concentrate on the unemployment implication of efficiency wage theory, rather than testing the more direct link between effort and wages. Efficiency wage theory predicts that a rise in the wage will increase effort, which in turn should raise productivity, giving wages a positive coefficient in a productivity equation. However, a higher wage could lead to higher productivity, independent of effort. For example, the trade union literature describes how a firm can react to a higher wage being forced upon it, by substituting capital for labour, which is likely to raise output per head amongst the workers remaining. Similarly, a higher union wage can shock a previously inefficient management into making better use of its resources, again leading to the prediction of a positive relationship between the wage paid and productivity. It seems that any study using productivity data to test efficiency wage theory must be very careful in specifying the relationships involved, which only Weisskopf (1987) made any attempt to do, of the studies considered above.

d.) Production Function Analysis

An approach related to the previous section, in that it considers outputs, is the production function approach. The idea is to estimate a production function and treat effort as a residual, that is the part of output that cannot be 'explained' by variations in other inputs, such as capital and labour. Effort can then be replaced by variables hypothesized to influence it, so that the coefficients on each of these variables will represent the effect of that variable on effort and the effect of effort on output, combined. For example, consider the Cobb-Douglas production function with all variables in log form:

$$y_i = \beta_1 k_i + \beta_2 l_i + \beta_3 e_i + \xi_i$$

where y_i is the log of firm I's output, k_i is the log of the capital stock, l_i is the log of the number of employees, and e_i is the log of the average effort level at firm I. ξ_I is all other factors that affect output, for example hours worked, the skill mix, the price of inputs and technological progress. An error term, v_i, will also form part of ξ_i. Effort, e_i, can then be written as:

$$e_i = \frac{1}{\beta_3}[y_i - \beta_1 k_i - \beta_2 l_i - \textit{other factors} - v_i]$$

Therefore, all variations in output that cannot be attributed to other factors are assumed to be explained by changes in worker effort. The usual approach is then to specify an effort function, using as arguments the variables hypothesized to affect effort. For example, the shirking version of the efficiency wage hypothesis predicts that effort depends on wages and unemployment. In this case, the effort function can be written as:

$$e_i = \alpha_1 w_i + \alpha_2 u_i$$

Substituting this expression for effort in place of e_i in the previous equation, and re-arranging, gives the estimating equation:

$$y_i = \beta_1 k_i + \beta_2 l_i + (\beta_3 \alpha_1) w_i + (\beta_3 \alpha_2) u_i + \xi_i$$

Since effort is assumed to increase output (that is, β_3 is positive), the signs of the coefficients on wages and unemployment will reflect the signs of the effect of these variables on effort (that is, the signs of α_1 and α_2).

Wadhwani and Wall (1991) adopt this method using a panel of company data, and find that the relative wage and the unemployment rate have a significantly positive effect on output, and thus, by implication, effort. Machin and Manning (1992) derive the dynamic prediction that both current and future wages should have a positive effect on effort if efficiency wage theory is correct, and verify this using similar methods and data to Wadhwani and Wall. Using basically the same procedure, Levine (1992) also finds a significant coefficient on a relative wage variable, measured by the manager's opinion of the average wage at his or her firm, relative to that of its three largest rivals. He also finds that this effect is larger in the non-union subsample than in the union subsample, the coefficient being insignificant in the latter case. As Levine points out, this is to be expected under efficiency wage theory, since a union wage may be higher than an efficiency wage optimally chosen by the firm, and thus the marginal effort response to a union wage increase will be lower, assuming diminishing marginal utility of wages and increasing marginal disutility of effort. An alternative interpretation, also consistent with the shirking model, is that discipline and dismissal procedures may be a lot tighter in union settings. Then any high wages which unionised employees receive will not have an incentive effect on effort, to the extent that the workers have faith in the dismissal procedures and so do not fear losing these wages even if they shirk.

These studies represent an improvement, but they are still not directly testing for a positive relationship between wages and actual effort data. There is also a danger that the positive correlation between wages and output may have an alternative explanation. For example, a third variable, that is not controlled for, may be influencing both wages and productivity in the same direction. Possibilities for such a variable include unmeasured worker ability, whereby the most able workers will be the most productive and also earn the highest wages, or transitory productivity shocks, so that firms affected by a positive productivity shock may increase their wages in an attempt to attract more workers to take advantage of the shock. Other potential explanations of a positive wage-productivity correlation revolve around the possibility that higher wages are actually the result, not the cause of higher productivity. This could be the case in a rent-sharing theory, in which the most productive firms are forced, by unions or insider power, to share some of their productivity gains with their workforce.

In fairness to the authors of the papers described above, they do consider such potential alternative explanations of their results. Levine (1992), in particular, considers all of the arguments just presented, and by using new data sets or developing new hypotheses, provides a counter-argument to them all. For example, he uses his union result to dismiss rent-sharing as a theory capable of explaining his results. The reasoning is that unionised workers should be more able than their non-unionised colleagues to obtain a share of any productivity gains, so that a stronger relationship between wages and productivity would be expected in unionised firms, rather than the weaker result that is actually obtained. Efficiency wage theory then emerges as the theory which seems most consistent with the results.

e.) Hours Function Analysis

A final, indirect method that has been used in the literature to examine efficiency wage relationships is to estimate hours functions. Oster (1980) outlines the basic principle. Desired hours, $H^*_{i,t}$, are hypothesised to be a positive function of both desired labour services and 'worker resistance to work,' the latter in effect being the inverse of effort:

$$H_{it}^{*} = g_i(L_{it}^{*} , A_{it})$$

Desired labour services, in turn, are written as a function of the output level, available capital and the state of technology:

$$L_{it}^{*} = f_i(Q_{it} , K_{it} , T_{it})$$

Similarly, worker resistance can be described as an inverse function of the unemployment rate. This is where the efficiency wage relationship of higher unemployment leading to higher effort enters the model.

$$A_{it} = k_i(U_t)$$

Substituting both expressions into the one for desired hours, assuming a log-linear specification, and representing technology by a quadratic time trend gives

$$\log H_{it}^{*} = a_0 + a_1 t + a_2 t^2 + a_3 \log Q_{it} + a_4 \log K_{it} + a_5 U_t$$

Finally, assuming actual hours follow desired hours via a partial adjustment process, an estimating equation is derived as

$$\log H_{it} = \lambda a_0 + \lambda a_1 t + \lambda a_2 t^2 + \lambda a_3 \log Q_{it} + \lambda a_4 \log K_{it} + \lambda a_5 U_t + (1-\lambda)\log H_{it-1} + u_{it}$$

Thus, since worker resistance is expected to increase the number of hours necessary to get a job done, the coefficient on the unemployment rate gives the sign of the effect of unemployment on resistance. When Oster runs this equation separately for twenty different industries for the period 1954-1975, the coefficient on the unemployment rate is indeed negative in every single case, as predicted, achieving significance at the 5% level in thirteen of these cases. Therefore, the efficiency wage relationship of higher effort at times of higher unemployment is confirmed.

Further analysis by Oster reveals that the strength of this 'unemployment effect' is negatively correlated with both the extent of unionisation and the degree of specific skills in an industry. Thus, if workers have some power which makes them more difficult to dismiss, whether via union-backing or the possession of job-specific skills, then the risk of unemployment loses its motivating effect. On the assumption that industries that are more unionised or more highly skilled represent primary sector industries, these results cast further light on the primary sector/secondary sector debate, and suggest that primary sector firms are *less* likely to use dismissal threats to motivate their workers, consistent with the results of Rebitzer (1987).

Green and Weisskopf (1990) perform a very similar analysis to Oster (1980), and arrive at much the same conclusions. Estimating an hours function separately for 143 3-digit industries for the period 1958-1981, they found a positive 'worker discipline' unemployment effect in 87% of their industries, which is significant in 55% of the total number of industries, at the 5% significance level. Using factor analysis to classify the industries according to their characteristics, the authors then examine under what circumstances the worker discipline effect seems to be significantly important. The results of this analysis reveal that the worker discipline effect is significantly negatively correlated with an indicator of belonging to the primary sector, and with such primary sector characteristics as the product market concentration ratio and union coverage. In addition, the unemployment effect is stronger in industries where the quit rate is higher, a typical characteristic of the secondary sector. Once again, therefore, the evidence suggests that primary sector firms and industries are *less* likely to use dismissal threats to motivate their workers, contrary to the theoretical model of Bulow and Summers (1986).

The hours function evidence is therefore consistent with efficiency wage theory, and suggests that the shirking model is particularly relevant in the secondary sector. As with the previous methods considered for testing this theory, the main problem with this evidence is that it is only considering an indirect link between effort and, in this case, unemployment, via hours. It is possible that the negative relationship observed between unemployment and hours could have another explanation, unrelated to effort. For example, it could be that in times of low aggregate demand, firms will cut back on the hours worked by their employees, while at the same time national unemployment is likely to rise. The primary sector/secondary sector evidence described above goes some way towards answering this criticism, since it points to a regular pattern of the strength of the hours-unemployment relationship across industries, which is consistent with the theories proposed, and is more difficult to explain under the alternative interpretation.

f.) Self-Reported Effort Studies

The most direct method of investigating a link between effort and other variables, such as wages and unemployment, is to use data which actually measures effort inputs. Clearly, the main problem is one of measurement. One solution, which will be taken up in the empirical sections of this thesis, is to use self-reported effort levels. To my knowledge, the only other papers which use a similar method to test efficiency wage theory are Drago (1991), Drago and Heywood (1992), Belman *et al* (1992) and Fairris and Alston (1994). The first of these studies uses four different proxies for effort, but the author prefers the estimated equation using a dependent variable derived from a survey question asking whether a worker would willingly work unpaid for twenty minutes after the official finishing time, the idea being that those who are more willing are the ones who put more effort into their jobs. Using a two-stage least squares method, weekly earnings are found to have a positive effect on this dependent variable. Industry dummies provide the identifying restrictions. Drago goes on to attempt to differentiate between competing efficiency wage stories. For example, the fact that workgroup pay dispersion and average workgroup pay are both insignificant in the effort equation is taken as evidence against the gift exchange version due to Akerlof (1982). There is also a problem with the shirking version, however, given that he finds an insignificant unemployment effect on effort.

Drago and Heywood (1992) use as a dependent variable a dummy constructed from a survey question asking whether the respondent exerts a great deal of effort when working. Efficiency wage theory is again supported, since the hourly wage and hours worked per week both have significantly positive coefficients, although it should be noted that no attempt is made to treat the wage as endogenous. Further support for the shirking version of the hypothesis is provided by the fact that the degree of supervision (+) and the re-employment probability (-) both significantly affect effort. The lay-off probability and spouse's income also take their predicted signs, but their coefficients are insignificant.

Fairris and Alston (1994) use a very similar dependent variable, a dummy to indicate whether the respondent agrees with the statement 'My job requires that I work very hard.' A potential problem is that the word 'requires' implies that these are not the freely-made, individual effort choices predicted by the theory. The paper explicitly considers the possibility of compensating differentials leading to a reverse causality of effort influencing wages, by estimating effort and wage equations simultaneously. Excluding human capital variables, working conditions, region and industry dummies from the effort equation, to provide the identifying restrictions, wages are found to have a positive effect on effort which is weakly significant. The unemployment rate also has a positive effect, but this is statistically insignificant. In the wage equation, the coefficient on effort is small and very insignificant. The authors therefore conclude that the causation is running from wages to effort, in support of efficiency wage theory.

Belman *et al* (1992) do not actually use self-reported effort data, but 'dobbing' data, whereby the respondents in their survey are asked to comment on the effort levels of their colleagues rather than of themselves. This gets around one problem of using self-reported effort data, which is that individuals typically overate their own performance. Another innovation of the paper is that it uses nine different measures of effort, some of which reflect individualistic behaviour, for example 'taking days off when not sick,' and some of which reflect typically group behaviour or low effort group norms, for example 'sticking to rules to slow work down.' Two ordered probit equations are run for each effort measure, one for a sample who report they work in cohesive work groups, and the other for a non-cohesive sample, on the assumption that group cohesion can be an important determinant of behaviour. This analysis therefore produces eighteen estimated equations, which makes interpretation somewhat difficult, given that no variable is significant in all, or in fact even in half, of the equations. One variable which does attract some sort of pattern is the hourly wage rate, which consistently reduces individualistic shirking, however measured, in the non-cohesive sample. Again, however, no attempt is made to treat wages as endogenously determined. The only other variable to achieve such consistency is an indicator of union density, which is positively and significantly associated with all measures of group shirking, in the cohesive sample. Other results include expected unemployment duration reducing some individualistic measures of shirking, more efficient monitoring technologies reducing some individual and group shirking, and higher actual monitoring levels increasing two shirking measures. The wage, unemployment and monitoring technology results all support the shirking version of the efficiency wage hypothesis, while the monitoring levels result can be explained using a 'gift exchange' story, whereby more intensive monitoring represents a reduction of gifts on offer from the firm, which will be reciprocated by lower effort levels. However, as evidence against this version theory, the degree of wage dispersion within the workgroup does not attract the significantly positive coefficient predicted of it by gift exchange theory in a single estimated equation. The authors therefore seem to decide that the shirking version is the more relevant version of efficiency wage theory. Finally, they conclude by mentioning some of the problems of using 'dobbing' data, for example that individuals in more cohesive work groups are less likely to report low effort levels amongst their colleagues than those in non-cohesive groups, so that it is not clear whether 'dobbing' data does have any advantage over self-reported effort data. Whatever the problems of using self-reported data, I believe such methods are the most appropriate for analysing the determinants of workers' effort, rather than the indirect methods described above, and the rest of this thesis will report the results of using such data.

g.) Non-Econometric Tests

Finally, before ending this review of efficiency wage tests, two more papers add an interesting new angle to the debate, in that they do not use econometric evidence. The first of these, by Raff and Summers (1987), involves a case study of Henry Ford's decision to pay five dollars per day to his employees in 1914, which represented a substantial increase in their pay. The authors point out that potential competitive explanations of

Ford's behaviour are unsuccessful. It does not appear that Ford had to pay such high wages because of a failure to attract workers in sufficiently large numbers, since contemporary accounts describe that there were already long queues for Ford jobs before the pay increase, especially as the US economy slipped into recession. Neither does it seem that Ford's aim was to improve the quality of his workforce, since the recent change to mass-production techniques had lowered rather than raised the skill requirements of the job.

Raff and Summers therefore argue that Ford's five dollars per day was an efficiency wage payment, to improve productivity at his workplaces. Certainly there appeared to be a need for such payments, since a daily absenteeism rate of 10% and an annual turnover rate of 370% in 1913 suggests that there were serious morale problems amongst the workforce. This interpretation seems to be supported by the results of Ford's experiment. A large increase in wages should lead to an increase in prices and/or a fall in profits. Data analysed by Raff and Summers show that neither of these results occurred. Profits continued to rise throughout 1914 and 1915, suggesting that the wage increase must have paid for itself in terms of increased productivity, as the efficiency wage literature predicts. Estimates based on output and employment data, described in the paper, imply a 40-70% increase in productivity. A dramatic fall in the turnover rate to 16% by 1915, which Raff and Summers consider to be 'visible manifestations of less easily quantified changes in workers' behavior' (p.S81), suggests how this increase in productivity came about. Of course, this case study evidence cannot prove the existence of a link between the wage increase and the change in workers' behaviour towards a more productive attitude, but the story is clearly consistent with the existence of efficiency wage payments, while more competitive theories have trouble explaining such events.

A second non-econometric approach is adopted by Drago (1990), who attempts to get around the problem of effort data being difficult to obtain by setting up a natural experiment in which the hypotheses can be tested under controlled conditions. The method is that the experiment subjects are told to pick an integer from 1 to 100 inclusive. This number, e, represents their effort choice (of course, the experiment is not cast in the language of effort choices). There is a cost to picking this number, given by $e^2/2000$, which represents the disutility of effort. As long as this 'effort' choice is not deemed to be too low, the participant receives an actual monetary payment, representing a fixed wage for performing a job, net of the cost of choosing the number. If the choice is too low, the individual receives no payment, but still has to pay the cost of making that choice. A card is drawn to determine the level at which the 'effort' choice is deemed to be too low, representing the point at which an employee is fired for shirking. Finally the probability of being monitored is modelled by giving each participant either one or three cards with an integer between 1 and 10 written on them. If they can match a randomly drawn card, a participant receives their payment whatever their 'effort' choice; they have not been monitored. If they cannot match the randomly drawn card then they have been observed, and only receive their payment if their 'effort' choice is high enough, as described above. Thus, the probability of 'being monitored' is m, which can take the value of either 0.7 or 0.9, and the probability of 'job loss' is m[1-(e/100)].

A base experiment gave each participant three 'monitoring' cards, so that the probability of detection was 0.7. The pay was $4.11 and there were twelve repetitions of the experiment. The next experiment was a high monitoring experiment, raising the probability of detection to 0.9 by giving each participant only one 'monitoring' card, and holding the payment at the same level. There were again twelve repetitions, with theory predicting that the average 'effort' choice across participants should be higher than in experiment 1, because of the greater chance of 'job loss.' This was the case for women in eight of the twelve repetitions, but for men in only two rounds. The third experiment was a high wage experiment, paying $5.28, while returning the monitoring probability to 0.7. Again, women seemed to behave in the manner predicted by efficiency wage theory, their average 'effort' choice being higher in this experiment than in experiment 1 in eleven of the twelve repetitions, while men raised their average 'effort' choices in only three rounds.

Thus it would appear that women act in a way consistent with efficiency wage theory, but this is not true of men. Drago cannot really explain this difference, simply hypothesising that maybe women are more intelligent or diligent and so work out the optimal responses in the experiment more successfully, or that women are more responsive to efficiency wage considerations. A problem with the experiment, acknowledged by Drago, is that when participants are 'fired' they receive no income but still pay the cost associated with the 'effort' choice, and so receive negative utility. In the real world, of course, utility is reduced following job loss, but is not negative. It may be that individuals behave differently when a potential outcome involves negative utility, for example avoiding the negative outcome at all costs rather than calculating the optimal response, and this may be affecting the results.

Such an experimental approach has been used extensively in psychological research, but is rarely used in economic work. The main problem seems to be that it can be argued that the decisions which individuals make in such controlled experimental conditions will not reflect the decisions which they would take in real life work conditions, where alternative incentives and a whole range of other influences may affect their behaviour. Even so, such evidence is an interesting addition to the more usual econometric tests that use 'real world' data.

2.4: Summary

The above discussion has revealed a large literature in the area of efficiency wages, but one that has yet to establish a definitive theoretical model of such phenomena, or conclusive proof of their existence.

The interest in efficiency wage models has been natural. Their ability to offer a potential explanation as to why wages can remain above their competitive, market-clearing level, and hence of involuntary unemployment, has meant that they can contribute an important element to macroeconomic debates of the problems facing modern industrial economies. Such implications follow, once the efficiency wage relationship between earnings and effort is accepted. The task of the efficiency wage literature is to persuade its readers of the legitimacy of this relationship.

Four theoretical arguments as to why effort should depend on wages have been detailed in the literature, as described above. I find the assumptions and ideas involved in all of these models plausible, and it may be that each consideration can have a particular relevance to a particular situation, or that all considerations are jointly important in all situations. *The* efficiency wage model may never emerge. Perhaps the least well developed model, in terms of explaining individual's behaviour within the economist's traditional choice-theoretic, utility-maximisation framework, is the 'Gift Exchange' approach, due to Akerlof. The theoretical chapter that follows will therefore focus on this approach, in particular offering a potential way to model the formation of, and the adherence to, group norms for effort.

The empirical evidence for efficiency wage theory has followed a range of different procedures, each with their own strengths and weaknesses. Despite these different approaches, the uniformity of results is impressive. With the exception of Leonard (1987), all papers reviewed above find evidence in favour of efficiency wage theory, to a greater or lesser degree. In particular, the studies which examine interpretations of efficiency wage theories all find their predictions supported by available data, except for Leonard (1987). With respect to papers that directly examine the relationship between effort on the one hand and earnings and unemployment on the other, using self-reported effort data or some proxy for effort, a significantly positive relationship between earnings and effort is consistently obtained. The relationship between effort and unemployment is less robust. Although the predicted positive relationship is usually obtained, the relevant coefficient is often insignificantly different from zero. It may be that the individuals are more affected by personal variables such as the wage they receive, rather than 'outside' variables such as unemployment rates. A more personalised view of job security, as measured by individual feelings, may have more relevance in explaining individual behaviour. This issue will be considered in the empirical chapters that follow.

An issue in the empirical literature has been whether efficiency wage theory, and in particular the shirking model, is more relevant in the primary or the secondary sector of the economy. The evidence is mixed, but tends to lean towards secondary sector establishments being more likely to use dismissal threats as an incentive to stimulate effort. Only Rebitzer and Robinson (1991) argue definitely in favour of the contrary, and their results rely on the classification of primary sector and secondary sector firms being respectively large and small. Green and Weisskopf (1990), Oster (1980) and Rebitzer (1987, 1988) all find the efficiency wage relationship between unemployment and effort is weaker in primary sector firms. Given the small amount of evidence on this issue, however, the empirical analysis that follows this review will consider this matter further.

In particular, the chapter based on the Workplace Industrial Relations Survey will examine how efficiency wage relationships vary according to union strength. This is the first time an empirical study has explicitly considered the relationship between unions and effort as its main theme.

Perhaps the most serious problem with all of the evidence reviewed is that alternative interpretations can be placed on all of the results. The papers that test relationships implied by the theory have the flaw that theories or ideas unrelated to efficiency wage theory can lead to the same implied relationships between variables. Similarly, the studies which adopt the more direct approach of using measures of effort, whether self-reported effort data or productivity proxies, face a problem in that the central positive relationship between earnings and effort in efficiency wage theory is also predicted by compensating differential theory, albeit with the reverse causality between the two variables. Causality is very difficult to prove, however, and so usually the evidence cannot discriminate between the two theories. Only three of the empirical studies reviewed above, Arai (1994), Fairris and Alston (1994) and Machin and Manning (1992), have explicitly considered compensating differential theory as an alternative explanation for their observed results. All three argue that their results are more consistent with efficiency wage theory, although only for the private sector in Arai (1994). Given this small number of studies to have considered reverse causality and compensating differentials, important parts of the empirical chapters to be presented below will be devoted to such considerations. This will involve new tests, not previously used by the three papers mentioned that have attempted to differentiate between the theories. It is hoped that the results presented will add to the available evidence on efficiency wage theory, by being more rigorous, and less susceptible to alternative interpretations.

References

Akerlof, G.A. (1982) 'Labor Contracts as Partial Gift Exchange,' *Quarterly Journal of Economics*, 97, 543-569.

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

Arai, M. (1994) 'Compensating Wage Differentials versus Efficiency Wages: an Empirical Study of Job Autonomy and Wages,' *Industrial Relations*, 33, 249-262.

Belman, D., Drago, R. And Wooden, M. (1992) 'Workgroups, Efficiency Wages and Work Effort,' *Journal of Post Keynesian Economics*, 14, 497-521.

Bowles, S. (1985) 'The Production Process in a Competitive Economy: Walrasian, Neo-Hobbesian, and Marxian Models,' *American Economic Review*, 75, 16-36.

Bulow, J.I. and Summers, L.H. (1986) 'A theory of Dual Labor Markets with Application to Industrial Policy, Discrimination, and Keynesian Unemployment,' *Journal of Labor Economics*, 4, 376-414.

Campbell, C.M. (1994) 'The Determinants of Dismissals. Tests of the Shirking Model with Individual Data,' *Economic Letters*, 46, 89-95.

Cappelli, P. and Chauvin, K. (1991) 'An Interplant Test of the Efficiency Wage Hypothesis,' *Quarterly Journal of Economics*, 105, 769-787.

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

Dickens, W.T. and Katz, L.F. (1986) 'Inter-Industry Wage Differences and Industry Characteristics,' in Lang, K. And Leonard J.S. (eds) *Unemployment and the Structure of Labor Markets*, London, Basil Blackwell.

Dickens, W.T., Katz, L.F., Lang, K. and Summers, L.H. (1989) 'Employee Crime and the Monitoring Puzzle,' *Journal of Labor Economics*, 7, 331-347.

Drago, R. (1990) 'Efficiency Wages: an Experimental Analysis,' *Australian Economic Papers*, 29, 68-84.

Drago, R. (1991) 'Incentives, Pay and Performance: a Study of Australian Employees,' *Applied Economics*, 23, 1433-1446.

Drago, R. (1993) 'Do Efficiency Wages Explain Dismissals?' *Applied Economics*, 25, 1301-1308.

Drago, R. and Heywood, J.S. (1992) 'Is Worker Behaviour Consistent with Efficiency Wages?' *Scottish Journal of Political Economy*, 39, 141-153.

Fairris, D. and Alston, L.J. (1994) 'Wages and the Intensity of Labor Effort: Efficiency Wages Versus Compensating Payments,' *Southern Economic Journal*, 61, 149-160.

Green, F. and Weisskopf, T.E. (1990) ' The Worker Discipline Effect: a Disaggregative Analysis,' *Review of Economics and Statistics*, 72, 241-249.

Gregg, P., Machin, S. and Metcalf, D. (1993) 'Signals and Cycles? Productivity Growth and Changes in Union Status in British Companies, 1984-9,' *Economic Journal*, 103, 894-907.

Groshen, E.L. (1991) 'Five Reasons Why Wages Vary Among Employees,' *Industrial Relations*, 30, 350-381.

Groshen, E.L. and Krueger, A.B. (1990) 'The Structure of Supervision and Pay in Hospitals,' *Industrial and Labor Relations Review*, 43, S134-S146.

Heckman, J.J. (1979) 'Sample Selection Bias as a Specification Error,' *Econometrica*, 47, 153-162.

Holzer, H.J. (1990) 'Wages, Employer Costs, and Employee Performance in the Firm,' *Industrial and Labor Relations Review*, 43, S147-S146.

Homans, G.C. (1954) 'The Cash Posters,' *American Sociological Review*, 19, 724-733.

Katz, L.F. (1986) 'Efficiency Wage Theories: a Partial Evaluation,' in Fischer, S. (ed) *NBER Macroeconomics Annual 1986*, Cambridge, MA, MIT Press.

Krueger, A.B. and Summers, L.H. (1986) 'Reflections on the Inter-Industry Wage Structure,' in Lang, K. And Leonard J.S. (eds) *Unemployment and the Structure of Labor Markets*, London, Basil Blackwell.

Krueger, A.B. and Summers, L.H. (1988) 'Efficiency Wages and the Inter-Industry Wage Structure,' *Econometrica*, 56, 259-293.

Lazear, E.P. (1982) 'Severance Pay, Pensions, and Efficient Mobility.' NBER Working Paper 854.

Lazear, E.P. (1995) *Personnel Economics*, Cambridge, Mass., MIT Press.

Leibenstein, H. (1963) 'The Theory of Underemployment in Densely Populated Backward Areas,' in *Economic Backwardness and Economic Growth*, chapter 6, New York, John Wiley and Sons.

Leibenstein, H. (1975) 'Aspects of the X-Efficiency Theory of the Firm,' *Bell Journal of Economics*, 6, 580-606.

Leibenstein, H. (1976) *Beyond Economic Man*, Cambridge, MA, Harvard University Press.

Leibenstein, H. (1978) *General X-Efficiency Theory and Economic Development*, New York, Oxford University Press.

Leibenstein, H. (1979) 'A Branch of Economics is Missing: Micro-Micro Theory,' *Journal of Economic Literature*, 17, 477-502.

Leonard, J.S. (1987) 'Carrots and Sticks: Pay, Supervision, and Turnover,' *Journal of Labor Economics*, 5, S136-S152.

Levine, D.I. (1992) 'Can Wage Increases Pay for Themselves? Tests With a Production Function,' *Economic Journal*, 102,1102-1115.

Machin, S. and Manning, A. (1992) 'Testing Dynamic Models of Worker Effort,' *Journal of Labor Economics*, 10, 288-305.

Oster, G. (1980) 'Labour Relations and Demand Relations: a Case Study of the "Unemployment Effect",' *Cambridge Journal of Economics*, 4, 337-348.

Raff, D.M.G. and Summers, L.H. (1987) 'Did Henry Ford Pay Efficiency Wages?' *Journal of Labor Economics*, 5, S57-S86.

Rebitzer, J.B. (1987) 'Unemployment, Long-Term Employment Relations, and Productivity Growth,' *Review of Economics and Statistics*, 69, 627-635.

Rebitzer, J.B. (1988) 'Unemployment, Labor Relations, and Unit Labor Costs,' *American Economic Review*, 78, 389-394.

Rebitzer, J.B. and Robinson, M.D. (1991) 'Employer Size and Dual Labor Markets,' *Review of Economics and Statistics*, 73, 710-715.

Roy, D. (1952) 'Quota Restriction and Goldbricking in a Machine Shop,' *American Journal of Sociology*, 57, 427-442.

Salop, S.C. (1979) 'A Model of the Natural Rate of Unemployment,' *American Economic Review*, 69, 117-125.

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

Solow, R.M. (1979) 'Another Possible Source of Wage Stickiness,' *Journal of Macroeconomics*, 1, 79-82.

Wadhwani, S.B. and Wall, M. (1991) 'A Direct Test of the Efficiency Wage Model Using UK Micro-Data,' *Oxford Economic Papers*, 43, 529-548.

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

Weisskopf, T.E. (1987) 'The Effect of Unemployment on Labour Productivity: an International Comparative Analysis,' *International Review of Applied Economics*, 1, 127-151.

Chapter 3

How Hard do People Work?

Effort Choices and Individual Characteristics

3.1 : Introduction

The introduction to this thesis provided a motivation for studying effort choices by asking a list of questions to which few answers have been given in the economics literature. The aim of this chapter, continuing into the next, is to gather empirical evidence in an attempt to answer some of those questions.

The evidence to be presented in this chapter is based on data from the 1989 British Social Attitudes Survey. This is a survey of individuals, and hence the questions dealt with are the ones which relate to the influences of individual characteristics on individual effort choices.[1] Using a particular question from the survey, an effort variable is derived which is regressed against a range of such individual characteristics, such as sex, age, ethnic origin, education, union membership and domestic responsibilities, in addition to characteristics of their place of work, such as firm size and working conditions. Hence the empirical results will suggest which individuals supply the greatest effort. This individual-level data can also be used to examine theories of effort determination that have been developed in the economics literature. Principal amongst these is efficiency wage theory, which argues that individuals' effort will respond to the wage that they receive. By including wages as one of explanatory variables in the estimated effort equations, this proposition can be analysed and tested.

The paper is organised as follows. Section 2 describes the data, which is then

1 This is to be contrasted with the evidence in the following chapter, which is principally concerned with the influence of the characteristics of individuals' places of work on their effort choices.

analysed in subsequent sections. The third section looks at how effort varies across individuals, while the important test of the efficiency wage hypothesis is conducted in the fourth section by introducing a wage variable into the estimated effort equation. This is followed by a section which includes local unemployment as a further determinant of individual effort choices. The concluding section is followed by an appendix which estimates similar effort equations for a range of other countries, using data from the International Social Survey Programme.

3.2 : Data

This paper presents estimated effort equations, using a variable from the 1989 British Social Attitudes Survey (BSAS) as a dependent effort variable. Few attempts have been made to use self-reported effort in empirical work. One example is Bielby and Bielby (1988). This paper is an answer to one by Becker (1985), which claimed that the reason women are paid less than men is because women specialize in household labour, and thus put less effort into their daytime jobs. Bielby and Bielby, using data from the 1973 and 1977 Quality of Employment Surveys, find that household factors do indeed reduce effort. For example, being responsible for household duties or childcare duties reduces effort put in at work, while having a non-working spouse increases job effort. However, they also find that, controlling for these factors, women input more effort than men, and so a woman with such household duties will still put in the same amount of effort as a man with no such responsibilities.

Other papers have used self-reported effort to test the efficiency wage hypothesis, for example Drago (1991), Drago and Heywood (1992), and Fairris and Alston (1994)[2]. Briefly describing the data used by each to measure effort, Drago (1991) uses four different proxies for effort, but prefers the estimated equation using a dependent variable derived from a survey question asking whether a worker would willingly work unpaid for twenty minutes after the official finishing time. Drago and Heywood (1992) use a dummy

[2] These three studies are described in detail in the literature review.

variable constructed from a survey question asking whether the respondent exerts a great deal of effort when working. Fairris and Alston (1994) use a very similar dependent variable, again defining a simple dummy variable based on whether or not respondents agree with the statement, 'My job requires that I work very hard.'

The use of different dependent variables in these studies reveals clearly the main problem with empirical analyses of effort; that an unambiguous measure of effort does not exist. Ideally, there would be a scale for effort, understood and accepted by everyone, so that an observer, or the individual herself, could accurately measure the amount of effort exerted. For example, consider the question of how many hours individuals work. There is an accepted scale for measuring such a concept. Either an observer or the individual herself know that once sixty minutes have passed, they can report that one hour has been worked. In the same way, with a perfect effort measure, once a certain amount of effort had been expended, everyone would know that one unit of effort could be recorded. Of course, such a measure does not, and never will, exist, because it cannot be described exactly how much effort is involved in 'one unit.' For this reason, the studies mentioned above use individual respondents' own subjective opinions as to the effort that they exert at work. Because there is no accepted question to ask in order to obtain such subjective responses, different authors have used different measures.

This paper follows such previous work in using a self-reported subjective effort measure. The variable used in this study comes from a question in the BSAS data set which asks 'How often do you come home from work exhausted?' Answers were coded 5=always, 4=often, 3=sometimes, 2=hardly ever, and 1=never. There was a reasonable spread of answers across the five possible response categories, as shown below.

Response	Percentage of Respondents
Always	6.5
Often	37.4
Sometimes	47.0
Hardly ever	7.7
Never	1.5

When comparing this data with the 'ideal' effort measure described above, two problems are immediately obvious. First, it is not an objective measure, based on a scale understood and accepted by everyone, but a self-reported subjective measure. Second, the question does not ask about effort *per se*, but exhaustion.

Considering the first problem, support for the use of self-reported effort data is provided by Hogan and Fleishman (1979). First, they report on laboratory studies, for example Borg (1962), in which laboratory tasks are performed, such as riding a bicycle machine or running on a treadmill. The advantage of using such tasks is that the equipment used can measure the work or effort input of the subject. After the experiment, the subjects were asked to estimate how much effort they had been inputting during the experiment. Borg finds a high correlation of 0.85 between perceived effort and actual effort, as measured by the equipment.

The problem with this and similar studies is that they only refer to laboratory tasks, and may not be suitable for generalizations to real world settings. Hogan and Fleishman therefore drew up a list of 30 well defined occupational tasks, for example, 'lift and carry objects weighing 85-100 pounds,' 'work heavy levers,' 'stock shelves with groceries,' and 'drive a truck.' Each of these tasks was given a physiological evaluation by work study experts. A group of 26 personnel specialists were then asked to rate the amount of physical work involved in each of these tasks, on a seven point scale ranging from very, very light to very, very hard. There was a high degree of agreement on the rating of each task, with the standard deviation for each task lying in the range 0.3 to 1.3. The authors then examine the correlation between actual metabolic cost and the average rating given by the respondents, and find a correlation coefficient of 0.81, which is highly significant. This accuracy in rating can also be extended to recreational tasks, where the authors find a correlation coefficient of 0.83.

It may be argued that personnel specialists will be more accurate in identifying the exertion necessary to complete tasks than general workers would be. Hogan and Fleishman therefore repeat the experiment for a sample of 54 general workers. The standard deviation of male ratings ranged from 0.8 to 1.6 for individual tasks, with the female range being from 0.8 to 1.8. There was therefore again a reasonably high consensus among respondents about the effort rating for each task. For male respondents, the correlation coefficient between perceived and actual effort was 0.80, with the equivalent female statistic being 0.75. Both of these correlations are highly significant.

The effort variable used in this study is not based upon predictions of effort for different tasks, but upon self-reports of workers' own effort. However, a further experiment by Hogan and Fleishman involved 20 subjects actually performing the task of moving boxes of varying weights, and then being asked to rate the effort involved. Again there were only small differences in the ratings given to any particular box, which supports the reliability of such data. When the average rating was compared with a combined weight and amount-of-movement index, the correlation was found to be 0.77. It therefore appears that workers can place a reasonably accurate rating on the amount of effort they input to a task.

All of these experiments relate to the physical effort involved in completing a task. Meijman *et al* (1986) examine mental effort, by asking subjects to perform a range of tasks including visual search, navigation and bus driving. The mental effort required for these tasks was measured by certain physiological indicators. When these measurements were compared with the average rating by subjects of the effort involved in each task, a strong positive correlation was found.

It therefore appears that self-reported effort levels may be a reasonably accurate indicator of the actual effort individuals put into their work. This was also the conclusion reached by Guest (1990), after surveying the literature on effort measurement. Guest claimed that self-reported effort levels appear to be a more accurate indicator of actual effort than more indirect measures based on productivity, such as the percentage utilization of labour (PUL) index.

If it seems that self-reported subjective effort levels are reasonably accurate indicators of individuals' actual effort levels, what would the ideal question to elicit such opinions look like? It might look something like, 'in a typical hour's work, how much effort do you exert; that is, how hard or intensely do you work.' A scale of possible answers should be offered, otherwise respondents would have difficulty in quantifying their answers. Such a question attempts to define effort, so that the respondents understand that they are expected to describe how hard they work. Asking about effort in a typical hour also makes it clear that the question is not concerned about the time dimension. Individuals who work long hours may report that they work hard, but while the reason for working long hours is an interesting question in itself, the focus of the present analysis is how hard individuals choose to work within a given time period.

It can be seen that the question actually used falls some way short of this ideal. In particular, the question asks about exhaustion at the end of the working day, rather than the actual effort exerted at a particular point during the day. There are likely to be many reasons other than effort exerted for why respondents should be exhausted when they return home from work. Given that an ideal effort measure is therefore not available, a second best solution is to control for such other reasons for exhaustion, by including indicators of them among the explanatory variables. An ideal data set would therefore include, for example, measures of the respondents' health, lifestyle, sleeping patterns, travel-to-work time, hours worked, type of work and working conditions. Again, the available data falls short of the ideal. While the respondents' opinions as to their working conditions can be included, as well as hours worked, personal characteristics such as health, lifestyle, and probably many others which influence exhaustion levels, are unobserved.

The problem can be expressed as a measurement error problem. As an indicator of effort, the exhaustion variable is measured with error. The effort equation can be written as

$$e = \beta x + v$$

where x is a vector of explanatory variables and v is a disturbance term. The effort data available come in the form of the exhaustion variable, h, which can be decomposed as

$$h = e + w$$

where w represents all other factors which may influence exhaustion. Substituting into the effort equation gives the estimating equation

$$h = \beta x + u$$

where u=v+w is a composite error term. Hence the β coefficients will still give unbiased estimates of the effects of the x variables on effort, as long as the x variables are distributed independently of u. In effect, this requires that the other reasons for exhaustion, w, do not vary systematically with any explanatory variable. Whether this assumption holds will be commented upon at various points in the results sections. It should also be noted that the addition to the error term will increase the standard errors of the estimated coefficients, reducing the chances of obtaining statistically significant coefficients on the explanatory variables. Any significant effects that are found may therefore be regarded as particularly strong.

Having discussed the dependent variable, what are the independent variables used in this study?[3] To see whether effort varies across individuals, variables were included reflecting the sex, age and ethnic origin of respondents, as well as their highest educational achievement, union status, socio-economic group, industry and region of work. Two 'household' variables were also included, reflecting the analysis of Bielby and Bielby (1988). In particular, dummy variables indicating respondents who were primarily responsible for childcare, and who were partnered by a non-working spouse, were used. A third variable indicating respondents primarily responsible for household duties was initially also used, but was never found to have a significant impact on effort, and so was excluded from the analysis. Dummy variables for firm size were included, as a proxy for the monitoring capabilities of the respondents' firms. It is hypothesised that monitoring

[3] A full list of variables used, together with their mean values, is given in Table 1.

is more difficult in larger firms, which may therefore receive lower effort levels if employees believe that there is less chance of their shirking being detected. Next, weekly hours of work were included, since length of time spent working may clearly influence reported exhaustion upon returning home. Finally, indicators of the respondents' type of work and working conditions were included.[4] There are two rationales for including these variables. First, they could affect the efficiency wage relationships to be investigated later in the chapter. For example, efficiency wage theory says that employees are less likely to shirk, the greater is the cost of job loss. This cost of job loss will be less, in utility terms, when the job involves bad working conditions. Hence, the working conditions included may be expected to attract negative coefficients. The second reason for including these variables is that possible reasons for exhaustion, other than effort, should be controlled for, as far as possible. As described above, if these reasons for exhaustion are not controlled for, and they are correlated with other explanatory variables, then the coefficients on those explanatory variables will be biased. It is likely that working conditions will affect reported exhaustion; for example, it is easy to understand why individuals who perform hard physical work or stressful work might report being exhausted when they return home form work. It is also possible to argue that such variables could be correlated with other explanatory variables of interest. For example, if compensating differential payments for poor working conditions are paid, an issue that is discussed later, then working conditions and earnings may be correlated. Likewise, if individuals with household responsibilities take jobs with particular characteristics, then

4

The variables used are indicators of whether the respondents perform hard physical work, stressful work, boring work, dangerous work, work in unhealthy conditions, or work in unpleasant conditions. The actual questions ask how often the respondent works in these conditions, with the answers coded on the same 1-5 scale as the exhaustion variable.

the coefficients on the household variables may be biased by the omission of the working conditions.

Simple correlations between some of these variables and the effort variable were first examined, by looking at mean effort levels (on the 1 to 5 scale) for different categories. The results are displayed in Table 2. Average effort was higher among females than males (3.38 compared to 3.37), among respondents of ethnic origin, rather than white or European respondents (3.77-3.36), among union members relative to non-union members (3.45-3.33) and among those who work longer hours. Respondents with no responsibilities for childcare had a higher average effort level (3.40 compared to 3.16), as did those with a non-working spouse (3.49-3.35). With respect to socio-economic group, higher average effort was reported by more senior workers for those with non-manual occupations, while the opposite was true for manual workers. The education variable provided mixed results, the two groups with the highest reported average effort being those with higher education and those with no qualifications at all. The firm size categories suggest that effort is on average higher among those who work in larger firms. Finally, average reported effort was highest in the North-West with respect to region, and in energy and water with respect to industry.

The next section begins to analyse more formally how effort varies across individuals.

3.3 : How Effort Varies Across Individuals? - Estimated Effort Equations

This section begins to answer some of the questions posed in the introduction. The 715 respondents who answered the exhaustion question were used as the unit of observation for OLS regressions, with the effort variable as the dependent variable, in an attempt to see how effort varies across individuals. The sample size was reduced to 553, once individuals with missing responses for any of the explanatory variables were removed. The results are given in Table 3. Controls for the occupation, industry and region of the respondent were also included in this and all regressions, but were never significant, unless mentioned in the text, and are not reported.

Looking first at the female variable, the Bielby and Bielby (1988) result that females input significantly higher effort, *ceteris paribus*, is replicated here. The studies by Drago (1991) and Drago and Heywood (1992) also found a positive effect, but in both cases their coefficient was insignificant. On the basis of the coefficient estimated here, being female is associated with an effort level that is about one-quarter of a standard deviation higher on the effort scale, relative to males. However, females are more likely to have childcare responsibilities, and the significant negative coefficient on this variable shows that such responsibilities are associated with lower effort at work. Although the childcare coefficient is not identical in absolute size to the female coefficient, an F-test (F=0.06 [1, 513]) cannot reject the null hypothesis that the two are equal but opposite in size, suggesting that a female with childcare responsibilities still puts as much effort into her job as does a male with no such responsibilities.

The other 'household' variable, having a non-working spouse, also has a significant coefficient, at the 10% significance level. The estimated coefficient suggests that a worker whose partner stays at home will increase his or her job effort by just under one-quarter of a standard deviation on the effort scale.

The results so far replicate the findings of Bielby and Bielby (1988). They are also consistent with a 'disutility of effort' story. Having to perform chores outside of work, such as childcare, makes effort at work more costly, and so reduces the amount of effort the individual chooses to input. Being married to someone who stays at home to perform these chores makes effort at work less costly, and so increases the chosen level of effort.

These results also seem to support the use of the exhaustion variable as a proxy for effort. Possible reasons for reporting exhaustion, such as poor sleeping patterns, could be expected to be positively correlated with the childcare variable, and hence a positive bias might have been expected on this variable. The fact that a significantly negative coefficient is obtained suggests that the effects of any such bias are not too strong. Put another way, if all the dependent variable is measuring is pure exhaustion, it should be positively correlated with the childcare variable, since childcare responsibilities on top of normal work would be expected to lead to greater exhaustion. Similarly a negative coefficient would be expected on the non-working spouse variable, since such individuals will have fewer domestic chores, and so should be less exhausted after work. The opposite is found in both cases, though, lending support to the interpretation that the dependent variable is measuring effort choices, rather than simply pure exhaustion.

Looking at the other variables in Table 3, age, defined as a normal continuous variable, has a significantly negative effect on effort. One possible explanation for this result is that younger workers have more to gain from pay rises and promotions, in terms of expected lifetime utility, than older workers, and so the gains from increasing their effort now are larger. This result also provides support for the use of the exhaustion variable to represent effort. It might be expected that poor health, as a reason for exhaustion, would be generally positively associated with age, and hence a positive bias would be expected on this variable due to the omission of controls for health. Put differently, it might be expected that older workers would be more exhausted after a day's work. Since the negative coefficient shows older workers report less exhaustion, this suggests that they are actually putting less effort into their work. Bielby and Bielby also found a negative effect of age on effort, although their study also presented evidence of a quadratic relationship between the two variables.[5]

The number of hours worked per week attracts a positive, significantly different from zero, coefficient, as was found by Drago and Heywood (1992). A one standard deviation (12.9 hours) increase above the mean in the number of hours worked will increase effort by roughly one-eighth of a standard deviation on the effort scale. Continuing the disutility of effort story, it could be argued that individuals with responsibilities outside of work accept only reduced-hours or part-time jobs. Since such individuals also choose lower effort levels, this will lead to a positive correlation between hours and effort. This is the conclusion Bielby and Bielby reached to explain their negative significant coefficient on a part-time variable. Of course, in the case of the hours variable, the exhaustion interpretation of the dependent variable cannot be ruled out, since a positive correlation between hours of work and exhaustion would naturally be expected.

5

An age-squared term was tried in the estimated effort equations, but attracted a highly insignificant coefficient, while also driving the coefficient on the linear term to insignificance. It would therefore seem that the relationship between effort and age in this data set is linear.

The three education variables all attract negative coefficients, although only the A- or O-level coefficient is significant, at the 10% level. Although the relationship appears to be monotonic, the omitted category represents individuals with no qualifications at all. The estimated coefficients suggest that people with no qualifications input the most effort into their jobs, but it is not the individuals at the other end of the scale, with higher education, who input the least, but those whose highest qualifications are CSEs, who input an effort level just under one-third of a standard deviation lower than individuals with no qualifications at all. The other papers using self-reported effort offer conflicting results concerning the effect of education on effort. Bielby and Bielby obtain a significantly positive coefficient on a years of schooling variable, but this becomes insignificant when earnings are also included, suggesting that the more educated only input more effort because of their higher earnings. Drago and Heywood find a significantly positive coefficient on a similarly defined education variable, but their result continues to hold after earnings are controlled for. The negative education effect found in this paper is replicated in the Drago paper, where a weakly significant negative coefficient is obtained on a dummy variable indicating those individuals who attended high school.

The remaining coefficients reported in Table 3 are all statistically insignificant. The ethnic minority variable attracts a large, positive coefficient. However, it is not well-defined, and fails a significance test at the 10% level. The likely cause is the small number (less than 3%) of respondents who report being from an ethnic minority. The only other paper to include such a variable was that by Drago and Heywood, which found that ethnic minority individuals are associated with significantly lower effort, contrary to the result obtained here.

The variable indicating respondents who are union members attracts a positive, but statistically insignificant coefficient.[6] Again, the other studies offer conflicting results for this variable. Drago and Heywood obtain the same result as the current analysis, with their union member variable attracting a positive but insignificant coefficient. The paper by Drago, however, obtains a negative and statistically significant coefficient on a similarly defined variable, although the estimation does differ in that it treats union status as endogenous.

As for the firm size dummy variables, all attract positive coefficients, none of which are statistically significant. The papers by Drago, Drago and Heywood, and Fairris and Alston all obtain the negative firm size effect predicted by efficiency wage theory, if monitoring is more difficult in large firms. The first of these results is strongly significant, the second weakly significant, and the last statistically insignificant. The two papers obtaining a significant result use the actual number of employees as an continuous indicator of firm size, rather than firm size dummy variables.

Finally, the coefficients on all of the control variables were statistically insignificant, except for those on the dummy variables indicating junior non-manual and skilled manual employees, both of which were significantly negative. As suggested by the cross-tabulations in Table 2, senior workers therefore seem to exert more effort among non-manual employees, while the reverse is true among manual employees.

Table 4 adds the six working condition variables to the specification of Table 3.

6

A variable indicating individuals who work at an establishment with at least one recognised union was used as an alternative to the union member variable reported here. The result was very similar, the variable attracting a positive but insignificant coefficient. Given that this chapter is concerned with individual characteristics, the remaining tables will include only the union member variable. The following chapter examines in more detail the relationship between effort and unions, and in particular the degree of union power, using establishment level data.

All six attract statistically significant coefficients, and it is clear that they are an important sources of variation in the dependent variable. Naturally, an F-test strongly rejects the hypothesis that the coefficients are jointly insignificant (F=29.46 [6, 507]). All of the tables that follow will therefore also include these variables. The previous section outlined two reasons for why working condition variables should be included. First, bad working conditions may reduce the utility associated with a job, and so make individuals more likely to shirk. Second, they should be included because they may offer reasons for exhaustion that are correlated with the other explanatory variables, leading to biases in the estimated coefficients if they are omitted. When the coefficients on the working condition variables are examined, the fact that some attract significantly positive coefficients, and others significantly negative ones, may suggest that both of these factors are at work, in different proportions for the different variables. Two of the bad working conditions, dangerous work and unpleasant conditions, have negative coefficients, so it would seem that it is these two conditions in particular that reduce the utility of a job sufficiently to leave individuals indifferent to losing their jobs and so deciding to shirk. The remaining working condition variables all attract positive coefficients, with the most likely explanation being that they are picking up the effect that certain jobs will lead to higher reported exhaustion. This explanation seems particularly plausible with respect to the variables indicating that the respondent performs hard physical work or stressful work. In fact, it is these two conditions which attract the largest coefficients, a one standard deviation increase in these variables raising reported exhaustion by just less than one-third of its own standard deviation, in the case of hard physical work, and approaching one-half a standard deviation in the case of stressful work.

Considering now the other coefficients in Table 4, it can be seen that some have been affected by the introduction of the working condition variables. The gender and age effects remain statistically significant, but others do not. Small reductions in the absolute value of the education coefficients mean that none achieve significance, although the 'A- or O-level' and 'CSE' coefficients are close to significance at the 10% level. Hours worked is another variable which loses its significant effect once working conditions are controlled for. However, the largest change in estimated coefficients occurs with respect to the household variables. The coefficients on the variables indicating respondents who are responsible for childcare or have a non-working spouse are both greatly reduced in absolute size, and are now statistically insignificant. While the reasons for these results can only be speculated on, a possible explanation suggested by the previous discussion is that bad working conditions tend to increase reported exhaustion. If, for example, individuals primarily responsible for childcare are less committed to the labour market, and only accept jobs with good working conditions, then the significant negative coefficient on this variable in Table 3, that was attributed to such individuals choosing to supply less effort, may have in fact been caused by this choice of jobs with good, less exhausting, working conditions. Rank-order correlations reveal that respondents with childcare responsibilities are indeed significantly less likely to have jobs with any of the bad working conditions, with the exception of a boring job. Similarly, those with a non-working spouse are significantly more likely to have jobs with some of the bad working conditions, particularly a stressful job, which makes them more exhausted, which may have led to an upward bias in the earlier coefficient. Thus the emphasis of the 'household' results changes slightly, in that it is now being argued that individuals with domestic responsibilities are more likely to choose less exhausting jobs, rather than that they put less effort into their jobs. Of course, the two effects will be related, and it is likely that both have some relevance.

All of the regressions discussed so far have used OLS analysis. However, the dependent variable is grouped into five discrete categories, with the order of these categories representing higher levels of effort. The numbers assigned to the categories, 1 to 5, are purely ordinal in nature, the numbers themselves having little meaning. OLS treats the five numbers as coming from a continuous cardinal scale, however, and for example treats the difference between effort=4 and effort=3, and effort=2 and effort=1 as equal, or treats effort=4 as representing twice the amount of effort as effort=2. Applying OLS when the analysis involves an ordered, limited dependent variable can therefore lead to biases in the estimated coefficients.

A more appropriate estimation technique is ordered probit analysis.[7] This technique assumes there exists some unobserved continuous scale for the dependent variable, effort, e*.

$$e^{*} = \beta x + u$$

where x is a vector of all the right-hand side variables, and u is a standard normally distributed error term. The data that are actually observed only put effort into five categories. The category chosen will depend on certain cut-off points on the continuous scale. Therefore observed effort e=1 if $e^*<c_1$, e=2 if $c_1<e^*<c_2$, e=3 if $c_2<e^*<c_3$, e=4 if $c_3<e^*<c_4$, and e=5 if $c_4<e^*$, where the c's are the cut-off points. Then, the probability that an individual chooses the first effort category, e=1, can be given as:

[7] This technique is described in detail in, for example, Maddala (1983).

$$Pr(e=1) = Pr(e^* < c_1)$$

$$= Pr(\beta x + u < c_1)$$

$$= Pr(u < c_1 - \beta x)$$

$$= \Phi(c_1 - \beta x)$$

Φ is the cumulative normal distribution.

Similarly,

$$Pr(e=2) = \Phi(c_2 - \beta x) - \Phi(c_1 - \beta x)$$

$$Pr(e=3) = \Phi(c_3 - \beta x) - \Phi(c_2 - \beta x)$$

$$Pr(e=4) = \Phi(c_4 - \beta x) - \Phi(c_3 - \beta x)$$

$$Pr(e=5) = 1 - \Phi(c_4 - \beta x)$$

Taking the log of each probability and summing, with a suitable indicator to show which effort category each observation falls into, gives the log likelihood function. This can then be maximized, using a suitable optimization technique, with respect to the parameters of interest and the cut-off points. The ordered probit coefficients estimated in this way are displayed in Table 5. The pseudo R^2 statistic reported is defined as:

$$pseudo\ R^2 = 1 - \frac{L_1}{L_0}$$

where L_0 is the value of the log likelihood function with only a constant in the model, and L_1 is its value with all of the independent variables included. Therefore, if the variables offer no explanation at all, then L_0 will equal L_1 and the value of the pseudo R^2 will be zero. At the other end of the scale, perfect prediction will yield a pseudo R^2 of 1, the same as when the usual R^2 statistic is used.

The estimated ordered probit coefficients are clearly different from the OLS coefficients. This is because the interpretation that must be put on each coefficient is different. With ordered probits, it is usual to talk about the marginal effects of the regressors on the probability of a particular value of the dependent variable being chosen. For example, the marginal effect of regressor x_i on the probability that effort takes its lowest value, 1, can be derived from the estimated coefficient using the formula:

$$\frac{\partial\ Pr(e=1)}{\partial x_i} = \frac{\partial\ \Phi(c_1-\beta x)}{\partial x_i}$$

$$= -\phi(c_1-\beta x)\beta_i$$

For the x values in the probability term, it is usual to use the mean values of each regressor.

Therefore, the coefficients cannot be interpreted in the same way as when OLS is used, but significance of the variables still works in the same way. Comparing Table 5 with Table 4 reveals that very little changes when ordered probit analysis is used. The significance of all coefficients is the same in both tables, with the exception of that on the CSE variable, which was just insignificant in the OLS equation, and is just significant in the ordered probit equation. The importance of all coefficients in terms of their relative size is also unchanged from the OLS equation. Thus the variables with the largest effects in the OLS equation, such as 'hard physical work,' 'stressful work' and 'female,' remain the most important variables in the ordered probit specification.

The fact that ordered probit analysis does not qualitatively alter the results is important in the next section, where a least squares technique is used to test the efficiency wage hypothesis.

3.4 : Testing the Efficiency Wage Hypothesis

There have been a limited number of attempts made to empirically test efficiency wage theory. Presumably because of the difficulty in measuring effort itself, most have attempted to test implications of the theory, for example that larger firms are more likely to pay high, efficiency wages (Rebitzer and Robinson [1991]), that disciplinary dismissals will be lower when higher wages are paid (Cappelli and Chauvin [1991]), or that firms which pay high wages should employ fewer supervisors (Leonard [1987]). Such tests do not provide conclusive proof in favour of efficiency wage theory, however, since alternative interpretations can be given to such predictions. A better approach is therefore to test the theory directly by obtaining effort data. Some authors have used productivity measures as a proxy for effort, for example Rebitzer (1987) and Holzer (1990). Others estimate a production function, treating effort as the residual (see Wadhwani and Wall [1991], Machin and Manning [1992], and Levine [1992]). The problem with both of these approaches is that any positive correlation found between wages and output or productivity may have an alternative explanation, for example unobserved worker quality or rent-sharing, despite the authors' attempts to check for this. The most direct approach is to regress actual effort data on wages, as performed by Drago (1991), Drago and Heywood (1992), and Fairris and Alston (1994). With the exception of the paper by Leonard (1987), most studies have, in general, supported the efficiency wage hypothesis.[8] The aim of this section is to add more evidence to this still small literature, using the final approach described, that is including wages in the effort equation that was estimated

[8] All of these studies are described in much greater detail in the literature review.

above.

The prediction that the set level of wages should positively affect effort is a distinctive feature of efficiency wage theory, and does not follow from other, psychological theories of individual behaviour. A popular example of the latter is expectancy theory. This theory states that the force on an individual to perform an act is the product of the subjective expectancy (probability) that the act will lead to a certain outcome and the utility associated with that outcome. If there are several possible outcomes, then these products are summed over all possible outcomes. Therefore:

$$F_i = \sum_j P_{ij} U_j$$

where F_i is the force to perform act i, U_j is the utility associated with outcome j, and P_{ij} is the individual's subjective expectancy that act i will lead to outcome j. In terms of the work situation, F_i is the force or motivation to perform effort level i. The various outcomes, j, may be interpreted as possible rewards. The individual will choose the effort level at which she is most motivated to work. Therefore, the greater the utility associated with the various rewards, and the greater the subjective expectancy that increased effort will lead to the gaining of these rewards, the greater will be the amount of effort input.[9] In this, its simplest form, the theory is therefore the same as expected utility theory in economics. Such a theory would interpret F_i as the expected utility associated with effort level i, with the individual choosing the effort level which maximizes this expected utility.

The important point is that expectancy theory predicts that wages will only motivate additional effort to the extent that they are paid dependent on this effort.

[9] For a more complete exposition of this theory, see Lawler (1971).

Therefore a basic guaranteed wage can have no motivational impact, since it does not vary in response to effort (the expectancy that effort will lead to higher wages is zero). Indeed, a supporter of expectancy theory, Vroom (1964), wrote;

'One possibility is that the strength of a worker's motivation to perform effectively is directly related to the amount of his wages. The more wages he receives the higher his motivation to do an effective job. There is little evidence in support of such a relationship. To this writer's knowledge there is no reliable data indicating that increases in wages increase levels of performance, or that decreases in wages decrease levels of performance' (page 252).

This statement is obviously directly opposed to efficiency wage theory, which takes as its cornerstone that effort responds to the level of wages. The results that follow, by including wages in the estimated effort equation, will therefore provide information as to the relevance of efficiency wage theory, as opposed to psychological theories of individual behaviour such as expectancy theory.

a.) Wages in the Effort Equation

The wage data used were taken from a question that asks:

'Which of the letters on this card represents your <u>own</u> gross or total <u>earnings</u>, <u>before</u> deduction of income tax and national insurance?' (emphasis in original).

Each letter corresponds to a band for total annual earnings. To form the variable actually used, the mid-points of these bands were calculated, and then logs were taken. Checks were undertaken, to ensure that the results obtained were not sensitive to the value given to the open-ended highest category.

A potential problem with the earnings variable is that it measures total earnings, and is therefore likely to comprise bonuses and performance related pay, as well as basic wages, for some of the individuals in the sample. In answer to a question asking whether any part of their earnings was made up by piecework payments, performance related pay or commission, 7.4% of the total sample said that this was the case for them. The use of such payment mechanisms is clearly a motivating, effort-inducing policy by the firm. If a positive correlation is observed between effort and earnings, this may then be due to the use of such mechanisms, and to attribute the correlation to efficiency wage effects may be erroneous. An attempt was made to remove such effects, by estimating the following equations using only the subsample of individuals who claimed not to receive piece-rate payments. The estimated coefficients did not change substantially, and all of the conclusions to be presented remained unaltered.

However, this point does lead to a more general one that, even if formal piecework or performance related pay schemes are not used, an individual's earnings may still be affected by her performance. In economic language, effort and earnings are likely to be simultaneously determined, so that the earnings variable is endogenous. In an attempt to allow for this endogeneity, a two-stage least squares method was used. This involves estimating a two-equation model, with effort and log earnings being the dependent variables in the respective equations. They will each also be explanatory variables in the other's equation. Estimating each equation individually by OLS would give inconsistent parameter estimates, because in each case the endogenous right-hand-side variable would be correlated with the disturbance term. It is therefore necessary to instrument these variables, with the instruments being provided by the exogenous variables in the model. In practise, this involves regressing, say, earnings against all of the exogenous variables in the model, then using the predicted values from this equation to instrument earnings in the estimated effort equation. In order to identify the effort equation, however, it is necessary that there are some exogenous variables in the earnings equation but not in the effort equation, and vice versa to identify the earnings equation. Ideally such identifying restrictions would be supplied by economic theory, but that is difficult in this case. In the absence of such theoretically-provided exclusion restrictions, resort was made to *ad hoc* assumptions. Thus, the one-digit industry dummies were excluded from the effort equation. The justification for doing so is that at the one-digit level, industry classifications are very broad. Within each of these nine categories, the complete range of occupations, jobs, firms, monitoring capabilities and so on are likely to be found. It is difficult to think of characteristics particular to such broad industry categories that will influence workers' effort choices. There does not seem to be any reason why two individuals should exert differing effort levels, purely because one works in, say, the energy industry and the other in manufacturing, controlling for their personal characteristics. Thus the one-digit industry dummies are not expected to explain any of the variation in the dependent variable, once the other explanatory variables are included, as was indeed observed in the above analysis. On the other hand, it is well established that an industry wage structure does exist, so that wages differ systematically by industry. It can therefore be argued that the industry dummies need not be included in the effort equation, but should be included in the earnings equation, and thus they can provide the exclusion restrictions required to identify the effort equation.

Another exclusion restriction is provided by the age-squared term, and thus the effort equation is also identified on the basis of functional form. As described above, effort appears to decline linearly with age, but no evidence of a quadratic effect can be found, both age and age-squared attracting highly insignificant coefficients when the latter is added to the effort equation. It can also be noted that a linear age effect for effort can be predicted by theory, as argued above. The older a worker gets, the lower the expected lifetime utility from remaining in a job becomes, and so the smaller her incentive not to shirk, implying the declining age-effort profile observed above. However, it is difficult to derive the prediction that effort should vary non-linearly with age from any theory of effort determination. On the other hand, it is well documented that earnings do not vary linearly with age, and that the inclusion of an age-squared term in earnings equations typically leads to a well-defined inverted-U shaped age-earnings profile. A similar result is found in the BSAS data, with both an age and an age-squared term attracting highly significant coefficients in estimated earnings equations. Thus the age-squared variable provides another exclusion restriction with which to identify the effort equation.

These restrictions are accepted using an F-test (F=0.513 [9, 472]). The F-test used was that suggested by Basmann (1960), based on comparing the restricted residual sum of squares when the exclusion restrictions are implemented, with the unrestricted residual sum of squares when all exogenous variables in the model are included. A high value for the computed F-statistic leads to a rejection of the null hypothesis that the exclusion restrictions are valid. As will always be done when such a test is reported, a check was also undertaken by computing the test statistic developed by Hausman (1983). This test involves regressing the residuals from the original estimation, which adopted the exclusion restrictions, on all of the exogenous variables in the model. The test statistic is $T.R^2$, where T is the number of observations. This is distributed as χ^2 with $K^* - M$ degrees of freedom, where K^* is the number of exogenous variables excluded from the equation in question, and M is the number of endogenous variables that are included. A high value of the test statistic again leads to a rejection of the null hypothesis that the exclusion restrictions are valid. Throughout this study, whenever a reported Basmann test accepts some exclusion restrictions, the Hausman test also accepts them.

The results of the first stage estimation, where the earnings variable is regressed against all of the exogenous variables in the model, are presented in Table 6. Standard wage equation results are obtained. All socio-economic groups, from professionals to skilled manuals, earn significantly more than unskilled manual workers, who form the omitted category. Only semi-skilled workers fail to obtain a significant differential above their unskilled colleagues. With respect to education, those individuals with a degree, A-levels or O-levels earn significantly more than individuals without any qualifications. The acquisition of CSEs, however, has no significant effect on earnings. Female, older and ethnic minority workers earn less than male, younger and white European workers, respectively. The last effect is, unusually, insignificant. The coefficient on the union member variable is also statistically insignificant, although this may be due to the fact that no allowance is made for the strength of the union at the individuals' workplaces. Earnings also increase with firm size, as is normally found. Hours have a significantly positive effect, as expected when the dependent variable is total annual earnings. Being primarily responsible for the care of children is not a variable often found in wage equations, but Table 6 shows that it has a significantly negative effect on earnings. The reason for this effect is presumably that such individuals accept less demanding, and hence lower paid, jobs, because of their family responsibilities, and often become trapped in low-paying, secondary sector labour markets. Finally, of the control variables, earnings are significantly lower in the 'distribution' sector (relative to 'finance'), while respondents living in Greater London or the South East report significantly higher earnings (relative to Wales).

The effort equation estimated in the second stage was the same as the preferred equation estimated in the previous section, except, of course, for the inclusion of the log of annual earnings and the exclusion of the industry dummies. The results are reported in column 1 of Table 7.

Looking first at the instrumented earnings variable, it can be seen that earnings have the positive effect on effort predicted by efficiency wage theory, but the estimated coefficient is insignificantly different from zero. Table 7 also shows that omitted variable bias is present in the estimated coefficients when earnings are excluded from the equation, since some coefficients change significantly after the inclusion of earnings. The effect is to increase the coefficients on variables inversely correlated with earnings, and vice versa for variables positively correlated with earnings. For example, female workers have lower earnings. The estimated coefficient in Table 4 should therefore have been downwardly biased. Examining Table 7, column 1 shows that this was indeed the case, the coefficient having almost doubled. Intuitively, the first regression showed females input more effort. However, those with higher earnings also input higher effort, and so controlling for wages (that is allowing for the fact that females earn less and still put in more effort) makes the 'female effect' on effort even stronger. The size of the coefficient now suggests that being female is associated with an effort level over half a standard deviation higher, relative to males.

Following a similar reasoning, individuals from an ethnic background often earn less, and so the estimated coefficient on this variable would also have been downwardly biased in Table 4. This again turns out to be the case, the 'ethnic' coefficient in Table 7 being larger in size compared to that in Table 4, although it remains statistically insignificant. The increase is smaller than the increase in the size of the female coefficient, because the ethnic minority variable is less strongly correlated with earnings than the female variable.

The pure effect of education on effort was previously biased upwards, since the fact that the more educated earn more and so input more effort for this reason was not controlled for. The inclusion of earnings does indeed reduce the estimated coefficients on the education variables, making them more negative. The coefficient on the CSE variable is now significant at the 10% level.

Perhaps the most important effect is on the hours variable. Since hours and total earnings are obviously positively correlated, the previous 'hours' coefficient was upwardly biased. Column 1 shows that, once earnings are controlled for, hours worked has a tiny, and very insignificant, effect on effort. The same result occurs if a part-time dummy variable is used rather than hours worked; its estimated coefficient is insignificantly different from zero. Therefore, it would seem that hours worked only influence effort via an effect on total earnings, having no independent effect of their own. This fact can be used as further evidence in support of the interpretation of the dependent variable as effort. If all that the dependent variable is measuring is pure exhaustion, then working longer hours would be expected to have an independent positive effect. This is not found, however, once earnings are controlled for.

The shirking version of the efficiency wage hypothesis argues that higher wages stimulate effort because of the higher opportunity cost of being caught shirking and sacked. Higher wages therefore stimulate effort through a higher cost of job loss. As argued by Schor and Bowles (1987), total earnings are a better measure of the cost of job loss, rather than an hourly wage rate measure. Only if the latter is the only measure available, need hours also be included. This argument is supported by the Drago (1991) paper described above. When he includes the hourly wage rate and hours worked as two independent variables, both attract positive, significant coefficients (the former at only a 10% significance level), reflecting their impact on the cost of job loss. However, when he includes weekly earnings he obtains a more statistically significant coefficient, significant at the 5% level, on this variable, while the hours variable loses its significantly positive effect, since it is no longer required to measure cost of job loss. In fact, the hours coefficient becomes significantly negative, which Drago interprets as long hours being a bad working condition, thus reducing the attractiveness of a job and hence the desire to input effort to avoid being sacked.

This discussion suggests that hours worked should be removed from the effort equation, since the BSAS wage data come in the form of annual earnings, and hours worked were found to have no independent effect on effort, after controlling for earnings. The alternative is to include the hours variable and to construct a crude measure of the hourly wage rate, dividing the mid-point of the individual's income range by his weekly hours multiplied by 52, thus assuming employment all year. When this was done the hours coefficient was indeed significantly positive, reflecting the influence of hours on the cost of job loss, while the hourly wage rate coefficient was positive but just failed to achieve significance. This may have been caused by the crudeness of the calculations involved, and because of this it was decided to adopt the former approach, leaving earnings in the annual form that they take in the original data set, and omitting the hours variable from the effort equation.

Column 2 of Table 7 reports the estimation results when this is done. The key result is that earnings now have a significantly positive effect on effort, once the unnecessary and highly correlated variable 'hours' is removed. The estimated coefficient suggests that if log earnings were to increase by one standard deviation (=0.321, which corresponds to £7210) above their mean value, effort would rise by one-third of a standard deviation on the effort scale.

An objection is that the hours variable should not be removed from the model altogether. Even if it has no effect on effort, and therefore need not be in the effort equation, it is still an important determinant of annual earnings, as was observed in Table 6, and so should be included in the two equation simultaneous model. Column 3 of Table 7 allows for this by maintaining hours as one of the exogenous variables in the model, which is therefore used to instrument earnings, since it is not in the effort equation. The results in column 3 show that this has virtually no effect on any of the estimated coefficients, so that there is no question of any of the estimated coefficients picking up an 'hours effect' on effort. This re-enforces the proposition that hours have no part to play in explaining effort, if total earnings are included in the analysis. The coefficient on log earnings is now smaller than the coefficient in column 2, but it remains statistically significant, at the 10% level. Including hours worked among the instruments does not affect the outcome of the F-test for validity of the exclusion restrictions; the null hypothesis that the excluded variables should not be in the effort equation still cannot be rejected (F=0.474 [10, 472]).

It could be argued that such a potentially endogenous variable as hours worked should not be interpreted as an exogenous variable used to identify the effort equation. In answer to this criticism, the results of column 2 can be pointed to, where hours worked was not used as an instrument. However, because hours worked are an important determinant of total earnings, this variable will be kept in the earnings equation in all further analysis, and hence will continue to act as an instrument for the earnings variable in the effort equation. None of the results are sensitive to its exclusion from the model, however, as suggested by column 2.

The evidence presented so far therefore suggests that effort and earnings are significantly positively correlated. This is an important finding in itself, but it is tempting to go further and conclude in favour of efficiency wage theory, as did Drago (1991) and Drago and Heywood (1992). However causality has not been proved, given the *ad hoc* manner in which the exclusion restrictions were chosen, casting doubt on how well identified the effort equation really is. More conclusive results regarding the causality issue could be derived if data sets could be obtained that contain a definite *exogenous* change in wages, for example a change in minimum wage laws, as well as an indicator of effort.

It is possible to go further here, however. For example, the coefficients on the variables other than the earnings variable can be re-interpreted in the light of efficiency wage theory. The shirking version of the theory argues that it is the cost of job loss that motivates individuals to input effort. The expected cost of losing a job will be higher, the greater the difficulty that person anticipates in finding another job. Thus the cost of job loss will be greater for those groups of individuals who are relatively disadvantaged in the labour market, such as women and ethnic minorities, and such groups should therefore input more effort, to avoid being sacked. Table 7 shows that this is indeed the case. Similarly, the more educated an individual, the easier it should be for her to obtain another job in the event of losing her current job. This implies a lower cost of job loss, and so a lower incentive not to shirk. The negative coefficients on the education variables show that educated workers do input less effort. Such results are therefore also consistent with efficiency wage theory.

b.) Compensating Differentials: Job Satisfaction Equations

The positive correlation found between earnings and effort has so far been interpreted as evidence in favour of efficiency wage theory. It could be argued, however, that the causality is running in the opposite direction, that is from effort to earnings. Such causality could be derived from a theory, first advanced by Adam Smith over 200 years ago, which said that 'the whole of the advantages and disadvantages of different employments of labour and stock must, in the same neighbourhood, be either perfectly equal or continually tending to equality' (Smith [1976], Book 1, Chapter X). This idea usually goes by the name of compensating wage differentials today. It argues that all workers receive exactly the same utility level, with higher wages compensating for any negative aspects of the job. Having to input greater effort could be one such negative aspect requiring compensation, and this could be the reason for the positive correlation between earnings and effort. In such a story, higher earnings are the result, rather than the cause, of higher effort.

Only three empirical efficiency wage studies of which I am aware have explicitly considered the possibility of compensating differential payments.[10] Machin and Manning (1992) derive the theoretical prediction that effort should be positively related to both current and future wages in the shirking model, but only current wages in a compensating differentials model, future wages playing no part. This they test and verify in a production function framework. Arai (1994) considers monitoring in an attempt to distinguish between the two theories. He argues that efficiency wage theory predicts a negative

[10] The three papers are described in more detail in the literature review.

relationship between monitoring and wages, the two being substitutes in the effort extraction process, while compensating differential theory predicts a positive relationship between the two variables, on the assumption that workers dislike being monitored and require higher wages in compensation. Arai estimates earnings equations for the private and the public sectors, including a respondent-reported measure of job autonomy (as the inverse of monitoring) amongst the usual controls in wage equations. The results show that autonomy has a positive effect on earnings in the private sector, in support of efficiency wage theory, while the reverse is true in the public sector. Arai argues that the difference across sectors is due to a lack of possibilities and incentives for public sector employers to pay efficiency wages, due to, for example, output being difficult to quantify and the lack of a profit motive in certain public sector institutions.

One problem with Arai's study is that, if certain aspects of worker quality are unobserved, then this will bias the results towards finding a positive relationship between autonomy and earnings, since unobserved skills will likely increase both an individual's earnings and her level of job autonomy. Thus the study will be biased in favour of efficiency wage theory. Arai argues that the fact that the results differ in sign between the two sectors is defence against this criticism, since it is difficult to imagine why, if unobserved worker quality is biasing the results, it should have different effects in the two sectors. This is not totally convincing, however. Given his earlier arguments that employment relations will differ between the two sectors, it could be that the bias is just stronger in the private sector, if worker quality is more valued there.

Fairris and Alston (1994) argue that any empirical efficiency wage analysis must allow for compensating differential payments, since jobs may differ in their effort requirements. Their study adopts a similar strategy, and comes to the same conclusions, as the current study, estimating a wage equation and an effort equation simultaneously by two-stage least squares, and finding earnings to have a significant effect on effort, with no evidence for the reverse causality. As such, however, their study is open to the same criticism as this one, that the exclusion restrictions adopted to identify the equations are chosen arbitrarily, with no reference to any theory, although, as here, they are not rejected by the Hausman test. The variables omitted from their effort equation in order to identify it are all human capital variables, all working condition variables, regional dummies and industry dummies. Their exclusion restrictions are therefore more extensive than the ones used in the current study, and so are even more difficult to justify.

Given this caveat concerning the two-stage least squares results, this paper attempts to provide further evidence on the causality between effort and wages, to help decide which theory best explains the positive correlation between the two variables. This new test uses a question from the BSAS data set which asks the respondent to say how satisfied they are in their job, on a 1 to 7 scale from 'completely dissatisfied' to 'completely satisfied.' This question should give an idea of the utility the respondent derives from her job.

The idea behind the analysis is to estimate job satisfaction equations, including working conditions among the explanatory variables. The one of interest is effort (as measured by the exhaustion question), with the other working condition variables used earlier also included. If the compensating differentials story is correct, then any variation in these variables will immediately be compensated by changes in earnings, leaving utility (satisfaction) unchanged. When earnings are included in the satisfaction equation, they are being controlled for. The coefficients on the working condition variables then represent the effects of these conditions on satisfaction, holding earnings constant. Since all of the working conditions included are expected to yield disutility, they should have negative coefficients in the satisfaction equation, once earnings are held constant. Formally, the partial derivative of the utility function with respect to any working condition, say effort, should be negative ($\partial u/\partial e<0$). However, when earnings are not included in the satisfaction equation, the analysis is allowing them to vary as effort, or any other working condition, varies. Compensating differential theory then predicts that, in this case, earnings should adjust to remove the dissatisfaction associated with the adverse working conditions, so that the estimated coefficients on all working conditions should be zero. Formally, the total differential of the utility function with respect to any of the working conditions, for example effort, should be zero ($du/de=0$), since wages also should vary to keep utility constant. Intuitively, satisfaction should be the same across all individuals, after controlling for all other influencing factors, so that variations in working conditions appear to have no effect on satisfaction, hence the predicted zero coefficients on such variables.

As mentioned, the estimated equations also control for other variables which are expected to influence job satisfaction. Variables included, as suggested by the job satisfaction literature, include gender, age, race, education, occupation, hours worked, union membership, firm size, industry and region. The results are presented in Table 8.

Column 1 displays the regression equation with work conditions and log earnings included. For a given level of earnings, four of the seven working conditions, effort, stressful, boring and unpleasant work, have the anticipated significant, negative effect on job satisfaction. Another, that measuring the extent to which the respondent's job involves hard physical work, is surprisingly positive and significant. Perhaps this reflects the satisfaction derived from doing an 'honest day's work.' The largest coefficient is on the 'boring' variable, suggesting that such work is the most disliked by workers. Only dangerous work and unhealthy work appear to have no effect on satisfaction. Important to this analysis, effort has a significantly negative impact on job satisfaction.

Column 1 also shows that earnings appear to have a negative, though insignificant, effect on job satisfaction. A positive coefficient was expected on earnings, on the assumption that the money received for doing the work will be positively valued, and so increase the satisfaction derived from the job. A similar result was obtained in a previous econometric study of self-reported job satisfaction carried out by Cappelli and Sherer (1988). Meng (1990) obtained the more expected finding that wages significantly increase job satisfaction.

The results of the estimation when earnings are excluded from the analysis are displayed in column 2. Column 1 showed that five of the working conditions significantly affect satisfaction (four negatively and one positively) and hence should receive compensating wage differentials, if the theory is correct. Allowing earnings to vary as working conditions vary should then drive the coefficients on the latter to zero. This is not what is observed, however. The five coefficients that were significant in column 1 remain significant when earnings are excluded from the equation. Indeed, the estimated coefficients barely change at all. This seems to imply that additional wages are not paid to compensate for these working conditions, since they continue to significantly reduce satisfaction (or increase it in the case of 'hard physical work') even when the analysis does not hold earnings constant, and hence allows for the possibility of compensating payments. Of particular interest to this study is that effort still has a significantly negative coefficient, suggesting that the positive correlation found between earnings and effort above is not caused by a compensating differentials story of higher effort leading to higher earnings.

While this does not prove that the efficiency wage interpretation of the previous sub-section's results is correct, it would seem to cast doubt on the 'reverse causality' interpretation.

It is possible to argue that the other variables included in the regressions to control for their possible effect on job satisfaction are in fact proxying earnings, since they are largely the same variables that are usually found in wage equations. If this is true, then it could be argued that earnings are, in effect, still present in the specification of column 2, and this is why the coefficients on the working conditions are not driven to zero. To examine this possibility, the regressions of Table 8 were repeated, including only working conditions and earnings, as appropriate. All other variables were excluded. The results were unchanged. There were no cases among the seven working conditions, including effort, where the estimated coefficient was significantly negative in the first specification with earnings included, but then insignificant in the second specification, when earnings were excluded. This suggests that the control variables in the satisfaction equations estimated above did not proxy earnings, when the latter were removed in the second specification.

Turning to the control variables and comparing the results to those found in previous empirical investigations of job satisfaction,[11] there is general agreement about the influences of these variables. Such replication of previous results lends support for the appropriateness of the satisfaction variable used. All of the papers mentioned use reported job satisfaction, similar to the dependent variable in the current study, except the Clark and Oswald work, which makes use of a General Health Questionnaire in the British

---

11

See, for example, Cappelli and Sherer (1988), Idson (1990) (both for the US), Meng (1990) (Canadian), Clark (1992) and Clark and Oswald (1994) (both using British data).

Household Panel Study to create an index of 'mental well-being.'

The strongest results in the above analysis are on the occupation dummies, which are aggregated to make their effect clearer to observe. The positive, significant coefficients increase monotonically in size as the occupation becomes more senior. To the extent that senior jobs involve greater participation in decision-making, more job autonomy, and a wider variety of tasks, these results agree with those of Cappelli and Sherer, who obtain highly significant, positive coefficients on such variables in a job satisfaction equation.

As a theoretical exercise to illustrate the effect of occupation on job satisfaction, consider an individual who is promoted from a junior non-manual to a senior non-manual position. The estimated coefficients in column 1 suggest that she would be willing to increase effort by almost one point on the effort scale, other things equal, and leave utility (satisfaction) constant. Similarly, a promotion from a skilled manual to a senior non-manual position would allow the individual to increase her effort by over 1.5 points and still keep satisfaction constant.

The only other variable with a significant coefficient in Table 8 is the number of hours worked per week, which, rather surprisingly, has a positive effect on job satisfaction. A positive coefficient on hours is also found by Meng, although in his case it is insignificant. Clark obtains the more expected negative coefficient on hours; a result that is significantly different from zero. A possible reason for the counter-intuitive result found in the present study is that hours of work are being endogenously determined, and that the causality is running from job satisfaction to choice of hours. The positive correlation between the two variables could than be explained by arguing that when individuals are more satisfied in their job, they are more willing to work extra hours.

Among the control variables, agriculture had a significantly positive coefficient. Apparently, the agricultural industry is the most satisfying to work in, although the small number of respondents in this industry may make such a result unstable.

Considering now the insignificant coefficients, of the firm size variables, two have very small, and thus statistically insignificant, positive coefficients, while the two representing 25-99 and 100-499 employees have larger negative coefficients, the latter approaching significance at the 10% level. Such dissatisfaction with working in large firms has been observed in previous studies. Clark (1992) obtains a significantly negative coefficient on his larger firm size dummy, while this relationship is the main focus of study for Idson (1990). The single establishment size variable in his study attracts a significantly negative coefficient. Further analysis reveals that the cause of such dissatisfaction is the lack of flexibility and increased regimentation often found in larger firms. This reasoning had been put forward previously by theorists, and Idson empirically demonstrates the result by showing that the inclusion of variables reflecting flexibility and regimentation drives the firm size coefficient to insignificance, in a job satisfaction equation.

The positive relationship between females and job satisfaction replicates the results of all other studies mentioned, except that of Clark and Oswald, who obtain a positive coefficient on their 'male' variable, possibly as a result of their alternative measure of job satisfaction. The gender dummy variable is significant in every one of these other studies.

Education having a negative effect (here predominantly through the A- or O-level variable) is also a common finding in job satisfaction equations. Cappelli and Sherer do not include any variables relating to education in their analysis, but all of the other studies under consideration find the level of education, however defined, to be inversely related to job satisfaction. The education coefficients in the studies by Meng and Idson are insignificantly different from zero, as in the present work, but Clark, and Clark and Oswald find education to have strong, significantly negative effects on job satisfaction. An attempt was made to obtain similar significant results here by aggregating the education dummies to the same level as used in these British studies (that is, 'higher education' and 'all other qualifications'). The two estimated coefficient were negative, but still insignificant, however. The explanation most often advanced for the negative education result is that the more highly educated have greater aspirations, and are thus relatively unsatisfied with their work when such aspirations are not fulfilled.

The dummy variable indicating union members also attracts an insignificant, negative coefficient. This relationship is the main interest of the paper by Meng, who also obtains a negative but insignificant coefficient on his union member variable in an equation examining overall job satisfaction, similar to the one reported here. Further investigation in his paper reveals that although union workers are satisfied with their pay and security, they are significantly less satisfied than non-union workers about the interest, freedom, personal influence and surrounding conditions associated with their jobs.

With respect to age, there is little agreement about its effect on job satisfaction. Age-squared is included here because it makes the age coefficients, and their t-statistics, larger than if age is entered on its own, although the estimated coefficients remain statistically insignificant. The two other studies using British data find a similar U-shaped relationship between job satisfaction and age, their coefficients being statistically significant. On the basis of the estimated coefficients here, job satisfaction reaches a low at 41 years of age, slightly higher than Clark and Oswald's estimates which are in the mid-thirties. Other studies have not replicated this U-shaped relationship, however. Cappelli and Sherer obtain a negative, linear and insignificant effect of age on job satisfaction, while Meng finds a significantly positive linear relationship.

The job satisfaction equations estimated above are therefore broadly in line with previous work in this area, thus supporting their authenticity. Every variable has a coefficient, the sign of which is replicated by at least some of the studies under consideration. Results for which there does not seem to be a consensus across studies involve possibly endogenous variables such as earnings and hours of work. Such problems are to be expected, given that OLS analysis is an incorrect procedure to use when endogenous variables are included on the right-hand side of the equation. In addition, the relationship between age and job satisfaction seems to vary across studies.

One criticism of the above analysis is that the job satisfaction variable, like the effort variable, is grouped into ordered categories, in this case seven. As explained above, ordered probit is a more appropriate estimation method for such dependent variables. Table 9 therefore repeats the analysis of Table 8, using ordered probit rather than OLS techniques. As was found when estimating effort equations, this makes little difference to the qualitative results. The effort coefficient is now significant at the 5% level in column 1, while the coefficient on the dummy variable representing unhealthy conditions is now significantly negative, at the 10% level. Interestingly, dangerous work is now the only working condition which does not attract a significant coefficient in the job satisfaction equations, and it is this variable that has been analysed more than any other in wage equation testing of compensating differential theory[12]. As for the compensating differential evidence in Table 9, the previous conclusions are clearly unchanged by the use of ordered probit analysis. The six working conditions, including effort, which

[12] Such tests are described in detail in the following sub-section.

significantly affect job satisfaction when earnings are held constant in column 1, still have their significant effects in column 2 when earnings are allowed to vary, suggesting that they do not attract compensating differentials.

As mentioned above, another possible problem with the analysis using job satisfaction as a dependent variable involves endogenous right hand side variables. To allow for this problem, a three equation simultaneous model was estimated, with the dependent variables being effort, log earnings and job satisfaction respectively. Each equation includes the other two variables as endogenous right-hand side variables. As described earlier, such variables will be correlated with the disturbance term of each equation, and so must be instrumented using all of the exogenous variables in the model.

Table 10 reports the resulting job satisfaction equations. To identify these equations, variables had to be excluded. It was decided to exclude the ten regional dummy variables, on the assumption that there is no reason why workers in one region should be fundamentally more satisfied with their jobs than workers in another region.[13] However, the usual caveat that there is no solid theoretical reason why this should be so applies. The null hypothesis that the exclusion restrictions are valid cannot be rejected on the basis of F-tests.

Columns 1 and 3 show, now that earnings are properly treated as endogenous, they obtain the more logical, positive effect on job satisfaction.

In column 2, five of the seven working conditions now have statistically

13

A potential argument against the use of regional dummies as instruments is that individuals' views as to what they expect from their work may differ across regions, for example individuals in more affluent areas may expect more from their job, while those in less well-off regions may be grateful just to have a job. Thus, job satisfaction could vary with region. While acknowledging this point, the analysis will proceed using regional dummies as instruments, because of a lack of alternatives.

insignificant coefficients, as compensating differentials theory would predict, when earnings are excluded from the satisfaction equation, and therefore allowed to vary as working conditions vary. In particular, effort now has an insignificant impact on job satisfaction. However, if the reason why these working conditions do not reduce satisfaction is that the workers are receiving increased earnings in compensation, then they should have significant, negative coefficients at a given level of earnings, that is when earnings are controlled for. Column 1 contains the results of the estimation when earnings are included in the equation. All five of the coefficients in question remain insignificant, suggesting that, even at a given level of earnings, variations in such work conditions do not significantly affect job satisfaction. The implication is that the results of column 2 are revealing not that compensating differentials are paid for the five working conditions, including the amount of effort involved, but that extra payments are not required for such conditions in the first place. The only work which seems to significantly reduce satisfaction is boring or unpleasant work, but even then there is no evidence that differentials are paid in compensation, since they also have a negative effect on satisfaction when earnings are not controlled for, and allowed to vary.

Although not directly relevant to this study, it is interesting to note that in columns 1 and 3 of Table 10, that is the specifications that contain earnings correctly treated as endogenous, the occupation variables become insignificant. This implies that when earnings are properly controlled for, higher position jobs are intrinsically no more satisfying than any other job. When earnings are removed, in column 2, these variables have a significant effect on job satisfaction again. Perhaps it is really only the money associated with them that make more senior positions more satisfying.

As for the other variables in the job satisfaction equations, the qualitative results remain the same, although some of the actual coefficients vary from the OLS analysis. In particular, the dummy variable representing the highest educational level now has a consistently negative coefficient, while the negative coefficient on the 'A- or O-level' variable is consistently larger in absolute value than before, and achieves significance in column 1. The firm size results are similarly stronger, with all firm size dummy variables now being negative, and generally larger in absolute value than previously. The '100-499 workers' size, in particular, now has a negative coefficient which is significant at the 5% level in column 1, and the 10% level in the other two columns. The two-stage least squares approach therefore strengthens the education and firm size results, which are replications of previous findings, as described above.

Summarizing the job satisfaction evidence, the OLS estimates imply that effort reduces satisfaction, but that respondents who report high effort do not receive wage differentials in compensation. The two-stage least squares evidence suggests that compensating differentials are not required for jobs with higher effort levels, since such work does not seem to reduce satisfaction. The latter approach is more appropriate in that it treats effort levels and earnings as endogenous, although there is still the problem of the arbitrarily chosen exclusion restrictions.

c.) Compensating Differentials: Earnings Equations

As a final test of whether compensating differentials are the cause of the observed positive wage-effort correlation, earnings equations were estimated which included the working conditions as independent variables.[14] If compensating differentials are being

---

14

Note that this equation is the same as that reported in Table 6, with the exception that the effort variable is included in Table 11. Because this variable attracts such a tiny coefficient, its omission did not affect the paid to workers, then their earnings should be significantly affected by their working conditions. This is the procedure which has most often been adopted when testing for evidence of compensating wage differentials.

As described in Smith (1979), it is quite acceptable to use OLS estimation, as part of a straightforward single equation procedure, when examining such relationships. Equilibrium wage-working condition combinations occur where a certain type of worker's indifference curve, in wage-working condition space, is tangential to a certain type of firm's isoprofit curve, defined over the same variables. In particular, this isoprofit curve will reflect zero, or normal, profits, if competitive conditions are assumed to exist; an assumption which compensating wage differential theory makes. Differences in preferences and technologies across different types of workers and firms, respectively, imply that a locus of such tangencies will exist. The actual wage-working condition pairs observed in the data will lie on this locus. The analysis is therefore not attempting to identify the separate curves which produce these tangencies, which would require simultaneous equation methods, but merely to identify the locus of tangencies, which can be performed within a single equation framework.

Table 11 displays the results. The estimated coefficients on the working condition variables are all very small, with the coefficient on effort being the smallest of all. Significant effects are only obtained twice; stressful work receiving extra payments, while hard physical work is associated with significantly lower pay. The latter case may be the result of employers reducing wages to compensate for the higher satisfaction found in such jobs, but it is more likely to be reflecting lower skilled, and hence less well paid, jobs

---

earlier results, and indeed the coefficients appear identical at the level of accuracy used for the reporting.

involving hard physical work. To test whether all of the working conditions together had a significant impact on earnings, an F-statistic was calculated, with the result that the null hypothesis of no significance could only be rejected at the 10% level (F=1.79 [7, 471]).

Exactly the same results are found in OLS wage equations estimated by Duncan and Holmlund (1983). Using Swedish data, they report that of their work condition variables, only those indicating hard physical work or stressful work significantly influence wages, the former having a negative effect. The authors point out that such results can be biased, however, if important personal characteristics remain unobserved. For example, if earnings potential is not fully controlled for, and if the respondents 'spend' some of this potential by choosing jobs with less adverse working conditions, then the coefficients on the bad working condition variables will be biased downwards. The reason for this is that respondents with low (unobserved) earnings potential will only obtain jobs with poor conditions, and will also have low wages, causing the negative bias on the bad working condition coefficients. To get around this problem, Duncan and Holmlund exploit the panel element of their data, estimating wage change equations, thus controlling for any individual characteristics that do not vary over time. The working condition coefficients estimated in this way are indeed larger, with the result that the negative hard physical work coefficient becomes insignificant, stressful work continues to have a significantly positive effect on earnings, while dangerous work acquires a positive and significant coefficient.

Brown (1980) reviews other studies which have included working conditions in wage equations. The significantly negative coefficient on a variable representing jobs requiring physical strength, as obtained above, has also been found by other researchers in his review, for example Lucas (1977) and Bluestone (1974). The finding that stressful work receives a positive compensating differential replicates the result of Quinn (1975). Other significant results that have been found include a general measure of bad working conditions, and an indicator of repetitive work, both having a positive impact on wages, as reported by Lucas (1977). A number of studies have also examined whether workers are financially compensated for above average risk of injury or death, generally finding significantly positive relationships. As reported by Smith (1979), however, the results are stronger and more consistently significant when it is a risk of death that is the issue, rather than merely injury. Meng (1991), using Canadian data, takes this argument further and provides evidence that it is only the risk of death due to work-related disease that attracts a significant compensating differential, as opposed to the risk of an accidental death at work. Meng presumes, without any particular justification, that workers have more knowledge about disease risks, or that they are better able to react in their wage demands to work-related deaths due to disease.

Despite the significant results described, a large number of studies have also turned up insignificant, and sometimes 'wrong sign' results. Brown (1980) concludes that the available empirical evidence provides 'an uncomfortable number of exceptions' (p.118). He puts forward the same argument as Duncan and Holmlund (1983), that the insignificant results may be caused by unobserved worker characteristics downwardly biasing the estimated coefficients, but when he performs a similar analysis to theirs using panel data, controlling for individual specific effects, he still can find no working conditions with a significantly positive effect on wages. The only significant coefficient is that on 'repetitive work' but this has the 'wrong' (negative) sign.

As far as evidence using British data is concerned, the only relevant studies of which I am aware are those by Marin and Psacharopoulos (1982) and McNabb (1989).

The former examines the relationship between wages and risk of death, matching data on death rates by occupation from the Office of Population Censuses and Survey's 1978 'Occupational Mortality Decennial Supplement,' into the 1975 General Household Survey. The results are supportive of compensating differential theory, although when the sample is split into professional/managerial, intermediate non-manual, and manual subsamples, only in the last case does the risk of death variable attract a positive coefficient which is significant at the 5% level. This is probably due to very small probabilities of death in the other categories. McNabb (1989) also uses the 1975 General Household Survey, which includes self-reported assessments of a number of job characteristics. Of these, a variable representing poor working conditions is associated with significantly higher annual earnings. Inconvenient hours of work also have a positive and significant effect on annual earnings, but this may be caused by low-wage workers taking on extra hours to boost their earnings. This latter interpretation is supported by the fact that the same variable has a negative impact on hourly wages. Finally, job insecurity appears to have no effect on wages, annual or hourly. On dividing the data set into the same subsamples as used by Marin and Psacharopoulos, McNabb replicates the result that any evidence there is in favour of compensating differential theory, is found only in the manual workers subsample.

The British results therefore follow the results of the other papers reviewed above. Risk of death appears to attract a positive compensating differential, but other working conditions yield fluctuating patterns of coefficients.

The results obtained above with the British Social Attitudes data, using job satisfaction and log earnings as dependent variables, cast doubt on compensating differentials being a significant phenomenon in the labour market. This is in keeping with other evidence on this subject, which at best can be described as inconsistent. While this does not completely prove the efficiency wage interpretation of the positive effort-earnings correlation, it would appear to rule out one of the most serious objections to such an interpretation.

If the efficiency wage interpretation is correct, as it seems, then this also casts doubt on the ability of expectancy theory to explain worker motivation in all circumstances. While the effect of conditional rewards on effort can still be formulated in terms of expectancy theory, the theory would seem to be incomplete, in that it cannot explain how a fixed level of wages can influence motivation to supply effort.

Finally, it is interesting to note here that, although evaluating theories in terms of how well they explain why modern 'human resource' managers pay high wages, rather than why there should be a positive effort-wage correlation, Ulman (1992) concludes that efficiency wage theory (and in particular the sociological or 'gift exchange' version of the hypothesis) is the theory most consistent with available evidence, after considering six possible theories, including the idea of compensating wage differentials.

3.5 : The Effect of Unemployment on Effort

a.) Unemployment in the Effort Equation

The previous section contained references to attempts to test efficiency wage models, full descriptions of which can be found in the literature review in Chapter 2. A number of them look for a positive relationship between unemployment and effort, however effort is measured in their studies. The motivation for doing so is that the shirking version of the efficiency wage hypothesis predicts that as unemployment rises, the cost of job loss will be greater, since it will require a longer time spent searching, with the consequent loss of income, before another job is found. Workers should then input more effort, in an attempt to avoid job loss as a result of being caught shirking. Although most of the studies reviewed find a positive unemployment effect on effort, as many find this to be statistically insignificant as find it to be significant.

The aim of this section is to provide further evidence on the effort-unemployment relationship, using the same effort variable as above. Regional unemployment rates for 1989 quarter 2 (the rates at the time of BSAS) were obtained from Department of Employment data, tabulated in 'Economic Trends'. These data were then matched into the BSAS data set, using the region of each respondent. However, when this variable was entered into the preferred specification of Table 4, it was found to attract a highly insignificant coefficient. As an experiment, unemployment data by region was then collected for quarter 2 of 1988, that is one year before the BSAS survey, and the annual change in the unemployment rate, rather than its level, was entered into the effort equation. The results are provided in Table 12, estimated by OLS, and Table 13, estimated by ordered probit.

It can be seen that the change in regional unemployment variable performs quite well in the effort equation, although it does not quite achieve significance.[15] The estimated coefficient in the OLS equation suggests that a one percentage point increase in the rate of change of unemployment will raise effort by roughly one-sixth of a standard deviation on the effort scale. It should be noted, however, that efficiency wage theory makes a prediction concerning the level of, rather than the change in, unemployment. To explain why the change in unemployment should appear more relevant to effort choices than the level, it is necessary to rely on rather abstract theorising. Perhaps individuals become used to the level of unemployment in their region, so that, even when this is high, they are not too worried about dismissal, maybe considering that they will go to the front of the queue for new jobs. It is only when unemployment is actually rising in their region that they become fearful of job loss, and so begin to exert effort. However, it would be difficult to model such ideas theoretically, and given that the change in unemployment coefficient is statistically insignificant anyway, maybe not too much attention should be paid to it, unless this effect is found to be important in other studies.

There is an econometric problem associated with this regression equation, which has been examined by Brent Moulton. In a series of articles (for example Moulton [1986], [1990]) he has pointed out that merging aggregate data with micro observations by, for example, industry, occupation or region, can lead to a downward bias in the estimated standard errors of the aggregate variables. This will lead to an upwards bias in the t-ratios, with the result that the investigator may infer a significant relationship where in fact

15

Other experiments were performed with unemployment, for example using a simple dummy variable indicating abode in a high unemployment region, rather than including the rates of unemployment. This variable performed very similarly to the change in unemployment variable, attracting an insignificant, positive coefficient, and so did not add anything more to the results.

there is none. Matching regional unemployment rates into the individual BSAS data will have this effect. Intuitively, the problem is that the estimation package assumes that each respondent has their own unique unemployment rate relevant to them, while in fact there is less information than this, there being only eleven different unemployment rates corresponding to the eleven regions. Moulton shows more formally that if the error terms are correlated within each region, which he argues is quite likely, then a computer package estimating by OLS will assume that there is no correlation between error terms, and so will apply an incorrect formula for calculating the covariance matrix of the OLS estimator. In fact, Moulton shows that the resulting bias in the estimated standard errors will be downward.

Given that the unemployment variable is insignificant anyway, and that the standard errors on the other variables are virtually unchanged relative to Table 4 (the equivalent specification without the unemployment variable included), it does not seem that the Moulton problem is very relevant in this instance. However, a way to formally test whether the problem exists is to estimate a random effects model of the form:

$$e_{ir} = \beta x_{ir} + v_r + u_{ir}$$

In regressing the effort of individual i in region r, e_{ir}, against the vector of explanatory variables, x_{ir}, the random effects model allows for a regional-specific component of the disturbance term, v_r, in addition to the usual error term. Not allowing for this regional component to the disturbance term can lead to an under-estimation of the standard errors. A test of how important this problem is therefore comprises testing whether v_r=0. The Breusch and Pagan Lagrange multiplier test for random effects tests whether the variance of the region-specific component of the disturbance term is zero, in effect testing whether

v_r=0 for all regions. Running a random effects regression on the specification of Table 12 and performing this test returns a test statistic of χ^2 (1) = 0.82. Thus, the null hypothesis of no random effects cannot be rejected, and the estimating package[16] returns the same coefficients in the random effects model that are obtained in the OLS analysis. The remaining equations in this section are therefore estimated without further reference to the Moulton problem.

Section 4 showed that earnings are an important influence on effort choices. To the extent that the analysis conducted so far in this section does not consider earnings, it may be considered to be deficient. Table 14 therefore examines the unemployment effect in estimated effort equations with log earnings included. As in Section 4, the two-stage least squares estimating technique is used, because of the endogenous nature of earnings. All exclusion restrictions used to identify the equation are the same as those used previously, and the F-test continues to fail to reject the null hypothesis that they are valid restrictions (F=0.487 [10, 481]).

The usual effects of including earnings on the effort equation coefficients are obtained. However, the coefficient on the unemployment variable is virtually unchanged from Table 12, and remains statistically insignificant. It would therefore seem that 'outside' variables, such as the unemployment rate in the individual's region, have a smaller impact on effort decisions than personal characteristics. It could be noted that the theoretically relevant variable is the individual's opinion as to her own particular chances of finding another job, following dismissal. It may be that the unemployment rate in her region does not accurately measure her opinion as to her own personal re-employment

---

[16] All equations in this thesis were estimated using Stata.

probability. This point was investigated using a variable which asked respondents how long they thought it would take them to find another job, if they lost their current one and started looking for a new one immediately. The effect on effort remained small, however, and the coefficient on this variable, while positive, was statistically insignificant, whether earnings were included or not.

b.) Micro- and Macroeconomic Job Security

The BSAS data set allows further analysis of the relationship between job security and effort. The shirking model predicts that effort rises when unemployment rises because workers fear having to spend longer searching for employment, and so want to avoid losing their current job in the first place. However, in times of unemployment, some workers may feel their jobs are safe, so that they hardly increase effort, while others may feel a lot less secure, and so increase their effort input more. This is introducing a slightly different job security concept, as emphasized by Lindbeck and Snower (1988). They distinguish microeconomic job security, as described by an individual's perceived probability of retaining her current job, from macroeconomic job security, which is characterized by her perceived probability of obtaining a new job after being sacked. It is the latter probability that is affected by the unemployment rate.

Three new variables from the BSAS data set were used to measure microeconomic job security. The first indicated a respondent who expected employment at her firm to fall in the coming year, on the assumption that if a fall is expected, that worker will feel that her job is more at risk. The second variable was more direct and asked whether the respondent actually expected to leave the firm themselves in the coming year. The final extra variable was a dummy to indicate a respondent who had suffered a spell of unemployment in the five years prior to the survey. The justification for including this variable is that if an individual has recently experienced unemployment, she might feel that her current job is more at risk, while an individual who has not experienced being out of work recently may think that she has built up an immunity to unemployment.

These variables have already been examined in the context of earnings equations by Blanchflower (1991). Expecting to leave the firm, and suffering unemployment in the previous five years, are both associated with significantly lower earnings. The third, indicating a respondent who expects employment to fall at his or her firm, is insignificant, but the reverse, expecting employment to rise, is associated with significantly higher earnings. Blanchflower interprets these results by arguing that when workers fear unemployment more than usual, they willingly forgo a portion of their wages in the hope of staying in a job. Will they likewise also increase their work effort, in an attempt to avoid unemployment?

Log earnings continued to be included in the estimated effort equation, and so a two-stage least squares approach was again used, treating earnings as endogenous. The results are displayed in column 1 of Table 15.

Although the three additional variables all have the hypothesized positive effect on effort, none attract statistically significant coefficients, although the coefficient on the variable indicating individuals with recent experience of unemployment is close to significance.

Microeconomic job insecurity can be further investigated by making use of the questions in the BSAS data set that ask respondents who expect to leave their firm, the reason for this expectation. There are seven possible responses, as listed among the variables in Table 15. If it is a fear of job loss that is making workers increase effort, then the variables indicating individuals who expect to be *forced* to leave the firm should be the ones with significant coefficients. As can be seen, however, none of the responses attract statistically significant coefficients, regardless of whether they represent voluntary or forced separations. At least in the BSAS data set, no evidence can therefore be found that individuals who feel more insecure in their current jobs exert more effort.

3.6 : Summary

Empirical investigation into the determinants of effort is quite rare in applied economic work, primarily because of the limited data available. This chapter uses the answers given by approximately 700 respondents to the question 'How often do you come home from work exhausted?' as an indicator of effort. The data source is the British Social Attitudes Survey, and the following results are found.

First, effort does not appear to be a constant, inherent in everyone, but is variable across individuals. Therefore, it is important to learn which individuals input more effort, and also how effort can be influenced. The empirical results of Section 3 suggest that females, the young, and the less well-educated put more effort into their jobs. Also, lower effort is associated with individuals who have responsibilities outside of work, for example childcare, while higher effort occurs if the individual has a non-working spouse to take care of these responsibilities. However, these 'household' effects lose their statistical significance once working conditions are controlled for. It therefore appears that individuals with responsibilities at home deliberately choose jobs that are less demanding or exhausting.

Second, there seems to be a positive relationship between effort and earnings. Efficiency wage theory would predict that the causation runs from earnings to effort, but the analysis of Section 4 does not prove causality, given that the restrictions used to identify the estimated effort equations were chosen arbitrarily, and were not derived from economic theory. Other explanations can be offered for this positive correlation, for example unobserved labour quality or compensating differential stories.

Further tests could find no evidence for the payment of compensating differentials.

Job satisfaction equations were estimated by OLS, including working conditions and log earnings among the explanatory variables. Standard job satisfaction results were obtained, for example satisfaction is higher amongst women, the less well educated, those in more senior positions, and those who work in small firms. Of interest to this study, controlling for earnings, most of the adverse working conditions, including effort, significantly reduce satisfaction. When earnings are then removed from the estimated equation, the analysis is no longer holding them constant, but they do not adjust to remove this dissatisfaction, since the coefficients in question retain their significant effect on job satisfaction. There is therefore no evidence that earnings adjust to equalize utility (satisfaction) across jobs with different working conditions, and in particular for the purposes of this study, that earnings adjust to compensate for different effort levels.

The results changed when a two-stage least squares approach was adopted, treating effort and earnings as endogenous. Then, most of the working conditions, including effort, attract insignificant coefficients, whether earnings are included or not. These results suggest that compensating differentials are not necessary, given that such working conditions do not affect satisfaction in the first place. The conclusion is unchanged, however, that there is no evidence for the 'reverse causality' argument that higher wages are paid in response to high effort levels.

In another test, earnings equations containing working conditions were estimated, thus following the standard method used to test for compensating differentials. Only one condition, stressful work, has a significantly positive effect on earnings, while the insignificant coefficient on the effort variable is the smallest amongst the working conditions. Once again, therefore, no evidence could be found in favour of the hypothesis that earnings respond to effort. However, the ruling out of this 'reverse causality' argument does not prove the validity of the efficiency wage interpretation of causality. In order to obtain stronger evidence in favour of efficiency wage theory, new data must be found, incorporating perhaps a definite exogenous change in wages or maybe a panel element.

Finally, Section 5 found no evidence of a significant unemployment effect on effort. Although the change in regional unemployment seems to have more relevance than the level, neither attracts a statistically significant coefficient. A number of previous studies have also failed to find a significant relationship between unemployment and effort. It may be that local unemployment rates do not accurately measure individuals' opinions as to their own personal re-employment probability, and it is the latter that affects their personal cost of job loss and hence their effort decisions. A question asking individuals how long they expected to be unemployed if they lost their current job was, however, also found to be unrelated to effort.

A final experiment involved examining microeconomic, as well as macroeconomic, job insecurity. Indicators for individuals who expected employment to fall at their firms, who expected to leave their firms in the next year, or who had experienced unemployment in the previous five years, were all included in the effort equation. All were positively associated with effort, although none of these relationships were statistically significant.

In conclusion, progress has been made in answering some of the questions posed in the introduction. It has been shown that work effort does vary across individuals, and hence which characteristics are more likely to be associated with higher effort, other things equal. As for the relevance of efficiency wage theory, the evidence presented above may still not be taken as conclusive, because of the difficulties involved in allowing for the endogeneity of earnings. However, it can be said that the results presented complement the existing empirical literature on efficiency wage theory, in that they arrive at similar conclusions, and also add to it, in that the new job satisfaction test of the reverse causality argument casts doubt on this being a relevant alternative to the efficiency wage interpretation.

Tables

Table 3.1 : Variable Means

Variable	Mean
how often come home from work exhausted	3.374
annual earnings (taking mid-points)	10510.6
female worker	0.463
age of respondent	39.564
member of ethnic minority	0.024
degree or higher education	0.282
A- or O-level	0.349
CSE	0.080
responsible for childcare	0.112
have a non-working spouse	0.157
hours worked per week	38.21
union member	0.382
10-24 employees	0.166
25-99 employees	0.275
100-499 employees	0.219
500+ employees	0.192
hard physical work	2.421
stressful work	3.110
boring work	2.175
dangerous work	1.870
unhealthy conditions	1.915
unpleasant conditions	1.797
SEG=professional	0.067
SEG=employer	0.136
SEG=intermediate non-manual	0.154
SEG=junior non-manual	0.253
SEG=skilled manual	0.184
SEG=semi-skilled manual	0.156

Table 3.2 : Average Effort by Variable Categories

Variable	Category	Mean Effort
sex	male	3.37
	female	3.38
race	ethnic minority	3.77
	white / European	3.36
education	degree or higher education	3.45
	A- or O-level	3.33
	CSE	3.16
	no qualifications	3.41
childcare responsibilities	yes	3.16
	no	3.40
non-working spouse	yes	3.49
	no	3.35
hours of work	10-15 hours	3.09
	16-23 hours	3.14
	24-29 hours	3.36
	30+ hours	3.42
union member	yes	3.45
	no	3.33
socio-economic group	professional	3.35
	employer	3.59
	intermediate non-manual	3.46
	junior non-manual	3.21
	skilled manual	3.32
	semi-skilled manual	3.41
	unskilled manual	3.50

Table 3.2 (continued)

Variable	Category	Mean Effort
size of firm	<10 employees	3.23
	10-24 employees	3.29
	25-99 employees	3.42
	100-499 employees	3.46
	500+ employees	3.40
industry	agriculture	3.00
	energy and water	3.67
	metals and minerals	3.26
	metal goods engineering	3.32
	other manufacturing	3.65
	construction	3.44
	distribution	3.30
	transport/communication	3.29
	financial services	3.28
	other services	3.40
region	Scotland	3.35
	Northern	3.22
	North-West	3.49
	Yorkshire and Humberside	3.42
	West Midlands	3.22
	East Midlands	3.42
	East Anglia	3.21
	South West	3.32
	South East	3.40
	Greater London	3.46
	Wales	3.46

Table 3.3 : OLS Equation, Dependent Variable=Effort

Variable	Coefficient
female	0.203 (0.088)**
age	-0.009 (0.003)**
ethnic minority	0.338 (0.217)
degree or higher education	-0.105 (0.110)
A- or O-level	-0.158 (0.095)*
CSE	-0.226 (0.140)
childcare duties	-0.234 (0.113)**
non-working spouse	0.174 (0.096)*
hours	0.007 (0.003)**
union member	0.091 (0.076)
10-24 employees	0.039 (0.118)
25-99 employees	0.113 (0.110)
100-499 employees	0.121 (0.117)
500+ employees	0.049 (0.125)
constant	3.367 (0.600)**
controls	occupation (6)
	industry (9)
	region (10)
no. of observations	553
R^2	0.121
RSS	286.2

Note: standard errors in parentheses. **=significant at 5% or better. *=significant at 10%

Table 3.4 : OLS Equation Including Working Conditions, Dependent Variable=Effort

Variable	Coefficient
female	0.257 (0.077)**
age	-0.005 (0.003)*
ethnic minority	0.241 (0.190)
degree or higher education	-0.024 (0.096)
A- or O-level	-0.123 (0.083)
CSE	-0.191 (0.122)
childcare duties	-0.121 (0.099)
non-working spouse	0.109 (0.084)
hours	0.003 (0.003)
union member	0.013 (0.066)
10-24 employees	0.030 (0.102)
25-99 employees	0.125 (0.095)
100-499 employees	0.082 (0.102)
500+ employees	0.015 (0.109)
hard physical work	0.194 (0.030)**
stressful work	0.326 (0.035)**
boring work	0.063 (0.033)*
dangerous work	-0.063 (0.036)*
unhealthy conditions	0.112 (0.038)**
unpleasant conditions	-0.100 (0.040)**
constant	1.620 (0.550)**
controls	occupation (6)
	industry (9)
	region (10)
no. of observations	553
R^2	0.348
RSS	212.2

Note: standard errors in parentheses. **=significant at 5% or better. *=significant at 10%

Table 3.5 : Ordered Probit Preferred Equation, Dependent Variable=Effort

Variable	Coefficient
female	0.483 (0.137)**
age	-0.009 (0.005)*
ethnic minority	0.435 (0.336)
degree or higher education	-0.056 (0.170)
A- or O-level	-0.221 (0.146)
CSE	-0.364 (0.216)*
childcare duties	-0.230 (0.175)
non-working spouse	0.184 (0.148)
hours	0.006 (0.005)
union member	0.020 (0.117)
10-24 employees	0.061 (0.181)
25-99 employees	0.228 (0.168)
100-499 employees	0.135 (0.181)
500+ employees	0.011 (0.193)
hard physical work	0.353 (0.055)**
stressful work	0.598 (0.066)**
boring work	0.118 (0.059)**
dangerous work	-0.119 (0.064)*
unhealthy conditions	0.208 (0.067)**
unpleasant conditions	-0.181 (0.070)**
controls	occupation (6)
	industry (9)
	region (10)
no. of observations	553
pseudo R^2	0.188
log likelihood	-511.3

Note: standard errors in parentheses. **=significant at 5% or better. *=significant at 10%

Table 3.6 : Predicting Equation to Obtain Instrumented Earnings, Dependent Variable=Log Earnings

Variable	Coefficient
female	-0.176 (0.023)**
age	0.020 (0.005)**
age^2/100	-0.022 (0.006)**
ethnic minority	-0.038 (0.054)
degree or higher education	0.147 (0.029)**
A- or O-level	0.118 (0.024)**
CSE	-0.001 (0.036)
childcare duties	-0.087 (0.030)**
non-working spouse	0.015 (0.025)
hours	0.007 (0.001)**
union member	0.014 (0.020)
10-24 employees	0.045 (0.030)
25-99 employees	0.106 (0.028)**
100-499 employees	0.103 (0.030)**
500+ employees	0.147 (0.032)**
hard physical work	-0.025 (0.009)**
stressful work	0.024 (0.010)**
boring work	0.001 (0.010)
dangerous work	0.001 (0.011)
unhealthy conditions	-0.001 (0.011)
unpleasant conditions	0.010 (0.012)
professional	0.189 (0.058)**
employer	0.245 (0.052)**
intermediate non-manual	0.205 (0.048)**
junior non-manual	0.163 (0.045)**
skilled manual	0.125 (0.045)**
semi-skilled manual	0.056 (0.044)
constant	3.079 (0.182)**
controls	industry (9)
	region (10)
no. of observations	519
R^2	0.707
RSS	15.9

Note: standard errors in parentheses. **=significant at 5% or better. *=significant at 10%

Table 3.7 : Two Stage Least Squares, Dependent Variable=Effort

Variable	Column 1	Column 2	Column 3
log earnings$_{iv}$	0.706 (0.651)	0.844 (0.480)*	0.594 (0.312)*
female	0.430 (0.137)**	0.474 (0.138)**	0.413 (0.105)**
age	-0.007 (0.003)**	-0.007 (0.003)**	-0.006 (0.003)**
ethnic minority	0.261 (0.194)	0.273 (0.196)	0.257 (0.192)
degree or higher ed.	-0.125 (0.144)	-0.149 (0.128)	-0.107 (0.111)
A- or O-level	-0.173 (0.118)	-0.191 (0.106)*	-0.159 (0.093)*
CSE	-0.208 (0.126)*	-0.205 (0.127)	-0.208 (0.125)*
childcare duties	-0.113 (0.108)	-0.094 (0.112)	-0.118 (0.105)
non-working spouse	0.078 (0.087)	0.075 (0.088)	0.080 (0.086)
hours	-0.001 (0.006)	-	-
union member	0.008 (0.069)	0.009 (0.068)	0.011 (0.067)
10-24 employees	0.013 (0.111)	0.005 (0.109)	0.019 (0.106)
25-99 employees	0.072 (0.124)	0.050 (0.119)	0.084 (0.107)
100-499 employees	0.034 (0.131)	0.009 (0.128)	0.046 (0.114)
500+ employees	-0.084 (0.153)	-0.112 (0.142)	-0.067 (0.124)
hard physical work	0.214 (0.036)**	0.219 (0.034)**	0.211 (0.032)**
stressful work	0.302 (0.040)**	0.296 (0.040)**	0.304 (0.037)**
boring work	0.076 (0.035)**	0.075 (0.035)**	0.077 (0.034)**
dangerous work	-0.068 (0.037)*	-0.070 (0.037)*	-0.068 (0.037)*
unhealthy conditions	0.095 (0.039)**	0.096 (0.040)**	0.095 (0.039)**
unpleasant conditions	-0.103 (0.043)**	-0.107 (0.043)**	-0.102 (0.042)**
constant	-0.395 (2.074)	-0.897 (1.655)	-0.052 (1.099)
controls	occupation (6) region (10)	occupation (6) region (10)	occupation (6) region (10)
no. of observations	519	519	519
R^2	0.312	0.300	0.320
RSS	206.2	209.9	203.9
instruments	industry (9) age^2	industry (9) age^2	industry (9) age^2 hours
F test of exclusion restrictions [d.f.]	0.513 [9, 472]	0.456 [9, 473]	0.474 [10, 472]

Note: standard errors in parentheses. **=significant at 5% or better. *=significant at 10%

Table 3.8 : Are Earnings Compensating for Higher Effort?
OLS Job Satisfaction Equations, Dependent Variable=How Satisfied with Job

Variable	Column 1	Column 2	Column 3
effort	-0.126 (0.068)*	-0.126 (0.068)**	-
hard physical work	0.114 (0.047)**	0.117 (0.047)**	-
stressful work	-0.203 (0.058)**	-0.205 (0.057)**	-
boring work	-0.485 (0.051)**	-0.485 (0.051)**	-
dangerous work	0.037 (0.055)	0.036 (0.055)	-
unhealthy conditions	-0.082 (0.058)	-0.082 (0.058)	-
unpleasant conditions	-0.162 (0.063)**	-0.163 (0.062)**	-
log earnings	-0.107 (0.237)	-	-0.301 (0.273)
female	0.061 (0.126)	0.080 (0.118)	0.109 (0.142)
age	-0.013 (0.025)	-0.015 (0.025)	-0.011 (0.029)
$(age)^2 / 100$	0.016 (0.031)	0.018 (0.031)	0.026 (0.036)
ethnic minority	-0.239 (0.282)	-0.234 (0.281)	-0.393 (0.324)
degree or higher ed.	0.009 (0.146)	-0.008 (0.141)	0.033 (0.170)
A- or O-level	-0.181 (0.128)	-0.194 (0.125)	-0.092 (0.148)
CSE	-0.085 (0.187)	-0.085 (0.187)	0.160 (0.216)
childcare duties	0.091 (0.158)	0.101 (0.156)	0.053 (0.182)
non-working spouse	-0.114 (0.129)	-0.115 (0.129)	-0.107 (0.149)
hours	0.009 (0.005)*	0.008 (0.004)*	0.006 (0.005)
union member	-0.071 (0.102)	-0.073 (0.102)	-0.122 (0.118)
10-24 workers	0.048 (0.155)	0.043 (0.154)	0.068 (0.180)
25-99 workers	-0.108 (0.148)	-0.119 (0.145)	-0.122 (0.171)
100-499 workers	-0.243 (0.158)	-0.254 (0.156)	-0.250 (0.183)
500+ workers	0.026 (0.167)	0.010 (0.163)	-0.130 (0.194)
senior non-manual	0.641 (0.261)**	0.618 (0.256)**	0.706 (0.292)**
junior non-manual	0.537 (0.231)**	0.518 (0.226)**	0.431 (0.260)*
semi-/skilled manual	0.429 (0.217)**	0.420 (0.216)*	0.374 (0.251)
constant	9.311 (1.189)**	8.986 (0.944)**	7.166 (1.344)**
controls	industry (9) region (10)	industry (9) region (10)	industry (9) region (10)
no. of observations	517	517	517
R^2	0.355	0.354	0.109
RSS	427.3	427.5	590.2

Note: standard errors in parentheses. **=significant at 5% or better. *=significant at 10%

Table 3.9 : Are Earnings Compensating for Higher Effort?
Ordered Probit Equations, Dep. Variable=How Satisfied with Job

Variable	Column 1	Column 2	Column 3
effort	-0.164 (0.077)**	-0.164 (0.077)**	-
hard physical work	0.147 (0.054)**	0.151 (0.054)**	-
stressful work	-0.243 (0.066)**	-0.247 (0.066)**	-
boring work	-0.571 (0.062)**	-0.571 (0.062)**	-
dangerous work	0.056 (0.062)	0.055 (0.062)	-
unhealthy conditions	-0.118 (0.067)*	-0.118 (0.067)*	-
unpleasant conditions	-0.162 (0.072)**	-0.163 (0.072)**	-
log earnings	-0.152 (0.269)	-	-0.325 (0.262)
female	0.056 (0.142)	0.083 (0.134)	0.087 (0.135)
age	-0.015 (0.029)	-0.018 (0.029)	-0.008 (0.028)
$(age)^2$ / 100	0.019 (0.036)	0.023 (0.036)	0.024 (0.035)
ethnic minority	-0.224 (0.313)	-0.217 (0.313)	-0.331 (0.304)
degree or higher ed.	0.046 (0.166)	0.022 (0.161)	0.065 (0.162)
A- or O-level	-0.175 (0.146)	-0.193 (0.142)	-0.053 (0.141)
CSE	-0.065 (0.214)	-0.065 (0.214)	0.210 (0.207)
childcare duties	0.097 (0.180)	0.110 (0.178)	0.034 (0.174)
non-working spouse	-0.081 (0.146)	-0.083 (0.146)	-0.054 (0.142)
hours	0.010 (0.005)*	0.009 (0.005)*	0.006 (0.005)
union member	-0.103 (0.116)	-0.105 (0.115)	-0.133 (0.112)
10-24 workers	0.013 (0.177)	0.005 (0.177)	0.012 (0.173)
25-99 workers	-0.130 (0.169)	-0.147 (0.166)	-0.139 (0.164)
100-499 workers	-0.330 (0.181)*	-0.346 (0.179)*	-0.301 (0.175)*
500+ workers	0.011 (0.190)	-0.012 (0.186)	-0.159 (0.185)
senior non-manual	0.821 (0.299)**	0.788 (0.293)**	0.723 (0.280)**
junior non-manual	0.698 (0.265)**	0.670 (0.260)**	0.452 (0.250)*
semi-/skilled manual	0.556 (0.249)**	0.542 (0.247)**	0.404 (0.240)*
controls	industry (9) region (10)	industry (9) region (10)	industry (9) region (10)
no. of observations	517	517	517
pseudo R^2	0.151	0.151	0.041
log likelihood	-631.9	-632.1	-713.4

Note: standard errors in parentheses. **=significant at 5% or better. *=significant at 10%

Table 3.10 : Are Earnings Compensating for Higher Effort?
2SLS Job Satisfaction Equations, Dependent Variable=How Satisfied with Job

Variable	Column 1	Column 2	Column 3
$effort_{iv}$	-0.196 (0.534)	-0.085 (0.510)	-
hard physical job	0.151 (0.122)	0.104 (0.111)	-
stressful job	-0.208 (0.179)	-0.216 (0.174)	-
boring job	-0.478 (0.068)**	-0.488 (0.066)**	-
dangerous job	0.028 (0.067)	0.036 (0.065)	-
unhealthy conditions	-0.078 (0.079)	-0.094 (0.076)	-
unpleasant conditions	-0.173 (0.085)**	-0.149 (0.080)*	-
log $earnings_{iv}$	1.069 (1.038)	-	0.728 (1.139)
female	0.286 (0.298)	0.056 (0.193)	0.279 (0.244)
age	-0.035 (0.032)	-0.015 (0.025)	-0.026 (0.036)
$(age)^2$ / 100	0.039 (0.038)	0.018 (0.031)	0.042 (0.043)
ethnic minority	-0.133 (0.320)	-0.180 (0.308)	-0.306 (0.323)
degree or higher ed.	-0.183 (0.218)	-0.013 (0.138)	-0.156 (0.248)
A- or O-level	-0.334 (0.190)*	-0.197 (0.132)	-0.218 (0.198)
CSE	-0.069 (0.211)	-0.052 (0.205)	0.201 (0.216)
childcare duties	0.151 (0.184)	0.092 (0.170)	0.084 (0.195)
non-working spouse	-0.127 (0.141)	-0.120 (0.138)	-0.142 (0.153)
hours	-0.001 (0.009)	0.007 (0.005)	-0.002 (0.010)
union member	-0.074 (0.101)	-0.063 (0.098)	-0.131 (0.116)
10-24 workers	-0.035 (0.162)	0.010 (0.153)	-0.018 (0.186)
25-99 workers	-0.256 (0.192)	-0.147 (0.156)	-0.266 (0.213)
100-499 workers	-0.382 (0.194)**	-0.278 (0.161)*	-0.379 (0.217)*
500+ workers	-0.192 (0.231)	-0.025 (0.160)	-0.352 (0.267)
senior non-manual	0.404 (0.361)	0.667 (0.249)**	0.425 (0.432)
junior non-manual	0.308 (0.338)	0.536 (0.249)**	0.186 (0.359)
semi-/skilled manual	0.324 (0.281)	0.464 (0.239)*	0.254 (0.280)
constant	5.568 (3.314)	8.780 (1.097)**	3.839 (3.708)
controls	industry (9)	industry (9)	industry (9)
no. of observations	517	517	517
R^2	0.311	0.344	0.068
RSS	456.2	434.3	616.9
instruments	region (10)	region (10)	region (10)
F test of exclusion restrictions [d.f]	0.722 [8, 473]	0.798 [9, 473]	0.804 [9, 479]

Note: standard errors in parentheses. **=significant at 5% or better. *=significant at 10%

Table 3.11 : Compensating Differentials - Including Working Conditions in the Wage Equation

Variable	Coefficient
effort	$1.60E^{-4}$ (0.013)
hard physical work	-0.025 (0.009)**
stressful work	0.024 (0.011)**
boring work	0.001 (0.010)
dangerous work	0.001 (0.011)
unhealthy conditions	-0.001 (0.011)
unpleasant conditions	0.010 (0.012)
female	-0.176 (0.023)**
age	0.020 (0.005)**
$(age)^2$ / 100	-0.022 (0.006)**
ethnic minority	-0.038 (0.054)
degree or higher ed.	0.147 (0.029)**
A- or O-level	0.118 (0.024)**
CSE	-0.001 (0.036)
10-24 workers	0.045 (0.030)
25-99 workers	0.106 (0.028)**
100-499 workers	0.103 (0.030)**
500+ workers	0.147 (0.032)**
professional	0.189 (0.058)**
employer	0.245 (0.052)**
intermediate non-manual	0.205 (0.048)**
junior non-manual	0.163 (0.045)**
skilled manual	0.125 (0.045)**
semi-skilled manual	0.056 (0.044)
childcare duties	-0.087 (0.030)**
non-working spouse	0.015 (0.025)
hours	0.007 (0.001)**
union member	0.014 (0.020)
constant	3.079 (0.183)**
controls	industry (9)
	region (10)
no. of observations	519
R^2	0.707
RSS	15.99

Note: standard errors in parentheses. **=significant at 5% or better. *=significant at 10%

Table 3.12 : The Effects of Local Unemployment, OLS, Dependent Variable=Effort

Variable	Coefficient
female	0.262 (0.077)**
age	-0.005 (0.003)*
ethnic minority	0.261 (0.188)
degree or higher education	-0.013 (0.096)
A- or O-level	-0.118 (0.082)
CSE	-0.179 (0.120)
childcare duties	-0.128 (0.098)
non-working spouse	0.116 (0.083)
hours	0.003 (0.003)
union member	0.012 (0.065)
10-24 employees	0.021 (0.101)
25-99 employees	0.119 (0.094)
100-499 employees	0.088 (0.101)
500+ employees	0.017 (0.107)
hard physical work	0.195 (0.030)**
stressful work	0.328 (0.035)**
boring work	0.071 (0.033)**
dangerous work	-0.060 (0.036)*
unhealthy conditions	0.112 (0.037)**
unpleasant conditions	-0.104 (0.039)**
Δ (regional unemployment rate)	0.129 (0.089)
constant	1.783 (0.557)**
controls	occupation (6)
	industry (9)
no. of observations	553
R^2	0.339
RSS	215.3

Note: standard errors in parentheses. **=significant at 5% or better. *=significant at 10%

Table 3.13 : The Effects of Local Unemployment, Ordered Probit, Dependent Variable=Effort

Variable	Coefficient
female	0.488 (0.136)**
age	-0.009 (0.005)*
ethnic minority	0.470 (0.331)
degree or higher education	-0.037 (0.168)
A- or O-level	-0.211 (0.145)
CSE	-0.341 (0.212)
childcare duties	-0.239 (0.173)
non-working spouse	0.196 (0.147)
hours	0.005 (0.005)
union member	0.021 (0.114)
10-24 employees	0.048 (0.179)
25-99 employees	0.219 (0.166)
100-499 employees	0.152 (0.179)
500+ employees	0.024 (0.188)
hard physical work	0.352 (0.055)**
stressful work	0.593 (0.065)**
boring work	0.131 (0.058)**
dangerous work	-0.113 (0.063)*
unhealthy conditions	0.204 (0.066)**
unpleasant conditions	-0.186 (0.069)**
Δ (regional unemployment rate)	0.239 (0.157)
controls	occupation (6)
	industry (9)
no. of observations	553
pseudo R^2	0.181
log likelihood	-515.6

Note: standard errors in parentheses. **=significant at 5% or better. *=significant at 10%

Table 3.14 : Earnings and Unemployment, 2SLS, Dependent Variable=Effort

Variable	Coefficient
log earnings$_{iv}$	0.521 (0.309)*
female	0.400 (0.104)**
age	-0.006 (0.003)**
ethnic minority	0.258 (0.188)
degree or higher education	-0.081 (0.108)
A- or O-level	-0.139 (0.091)
CSE	-0.192 (0.123)
childcare duties	-0.123 (0.103)
non-working spouse	0.086 (0.086)
union member	0.011 (0.065)
10-24 employees	0.016 (0.104)
25-99 employees	0.085 (0.105)
100-499 employees	0.059 (0.113)
500+ employees	-0.053 (0.121)
hard physical work	0.211 (0.032)**
stressful work	0.306 (0.037)**
boring work	0.084 (0.034)**
dangerous work	-0.064 (0.036)*
unhealthy conditions	0.097 (0.039)**
unpleasant conditions	-0.108 (0.042)**
Δ (regional unemployment rate)	0.102 (0.093)
constant	0.362 (1.141)
controls	occupation (6)
no. of observations	519
R^2	0.314
RSS	205.7
instruments	industry (9)
	age^2
	hours
F test of exclusion restrictions [d.f.]	0.487 [10, 481]

Note: standard errors in parentheses. **=significant at 5% or better. *=significant at 10%

Table 3.15 : Investigating Microeconomic Job Security, Dependent Variable=Effort

Variable	Column 1	Column 2
log earnings$_{iv}$	0.637 (0.326)**	0.606 (0.326)*
female	0.464 (0.111)**	0.464 (0.111)**
age	-0.005 (0.003)*	-0.004 (0.003)
ethnic minority	0.231 (0.188)	0.211 (0.190)
degree or higher education	-0.126 (0.111)	-0.116 (0.112)
A- or O-level	-0.175 (0.093)*	-0.171 (0.094)*
CSE	-0.179 (0.124)	-0.173 (0.125)
childcare duties	-0.140 (0.105)	-0.147 (0.106)
non-working spouse	0.090 (0.087)	0.094 (0.088)
union member	0.018 (0.067)	0.025 (0.068)
10-24 employees	0.020 (0.105)	0.008 (0.107)
25-99 employees	0.075 (0.107)	0.069 (0.108)
100-499 employees	0.026 (0.116)	0.020 (0.117)
500+ employees	-0.057 (0.123)	-0.061 (0.124)
hard physical work	0.213 (0.032)**	0.212 (0.032)**
stressful work	0.295 (0.038)**	0.292 (0.038)**
boring work	0.089 (0.035)**	0.090 (0.035)**
dangerous work	-0.059 (0.037)	-0.056 (0.037)
unhealthy conditions	0.099 (0.039)**	0.101 (0.040)**
unpleasant conditions	-0.120 (0.043)**	-0.122 (0.043)**
Δ (regional unemp. rate)	0.115 (0.095)	0.111 (0.095)
expect emp. to fall at firm	0.072 (0.081)	0.089 (0.085)
expect to leave firm	0.074 (0.078)	-
expect firm to close	-	-0.280 (0.275)
expect redundancy	-	0.005 (0.176)
expect to retire	-	-0.220 (0.304)
expect contract to expire	-	-0.127 (0.341)
expect work at other firm	-	0.150 (0.103)
expect self-employment	-	0.090 (0.199)
expect to look after home	-	0.088 (0.208)
unemployed in last 5 years	0.136 (0.084)	0.124 (0.085)
constant	-0.154 (1.221)	-0.095 (1.216)
controls	occupation (6)	occupation (6)
no. of observations	507	507
R^2	0.317	0.324
RSS	199.1	197.1
instruments	industry (9) age^2 hours	industry (9) age^2 hours
F test of exclusion restrictions [d.f.]	0.329 [10, 466]	0.318 [10, 460]

Note: standard errors in parentheses. **=significant at 5% or better. *=significant at 10%

Appendix: Effort Determinants in Other Countries

The section of the 1989 British Social Attitudes Survey entitled 'Work Orientations,' which contained the effort variable used in the preceding analysis, formed the subject of the 1989 International Social Survey Programme (ISSP), whereby a group of ten countries agreed to include a set of standard questions on this topic in one of their various national surveys. Thus exactly the same effort question was asked in each of these countries, allowing estimation of effort equations for each. This appendix describes the results.

The equation estimated for each country is based on the specification in Table 14 above. Thus the estimations are by two-stage least squares, treating the log of earnings as endogenous, and including the annual change in local unemployment rates as an additional regressor. Israel was immediately ruled out as a subject of analysis because no data was collected on earnings, the prime variable of interest, but equations similar to Table 14 are estimated for Austria, Germany (West), Hungary, Ireland, Italy, the Netherlands, Norway and the United States. Two variables omitted in every case are the ethnic minority and responsibility for childcare variables, as these questions were not included in the ISSP survey. The latter is replaced by an indicator of respondents with prime responsibility for household duties (which was originally included in the initial analysis of the British BSAS data, but omitted from the reported results due to a lack of any significant effects). The indicators of educational attainment obviously differ according to each country's national education system, but in each case the dummy variables are arranged in ascending order, so that education levels with a higher number represent a higher level of attainment. The omitted category is always the lowest education level. For each country, two equations were run with age entering first linearly and then quadratically. Only the results for the specification which best captured the age effect are reported in Table A1. Occupation and firm size controls are included where available. Finally, with respect to the instruments used, the intention was again to copy the specification of Table 14 for the British analysis, but not all of these variables were available for each country. Clearly, age-squared can only be used as an instrument when the age effect on effort appears to be linear and so age-squared is not required in the effort equation. One-digit industry dummies and the hours variable are also used as instruments where available. For every country, F-tests fail to reject the null hypothesis that the instruments are valid.

The results are presented in Table A1. Looking first for the existence of efficiency wage effects, it can be seen that the instrumented earnings variable achieves a significantly positive coefficient in five of the eight countries. Amongst these, the magnitude of the effect is by far the largest in Italy, followed by Austria then Germany. The Italian coefficient suggests that a one standard deviation increase in earnings will increase effort by almost one standard deviation on the effort scale. In Austria, Germany and the United States, a one standard deviation increase in earnings leads to around a one-third of a standard deviation increase in effort. Thus in these three countries the magnitude of the earnings effect is very similar to that found in the British results. The size of the effect in Norway is smaller than in these other countries. Thus efficiency wage considerations appear to be relevant in a range of countries, consolidating the results obtained for Britain above. The only exceptions in the study are the (then) Communist country of Hungary, the Netherlands and Ireland.

Local unemployment is entered in annual change form, to mirror the British specification, where possible. In two countries, however, Norway and Hungary, consecutive years of regional unemployment data could not be obtained, so the level is entered for these countries.[17] The results do not provide any clear patterns, however, with the unemployment effect, however measured, being positive in four countries, and negative in four countries. Both one of the positive coefficients and one of the negative coefficients are statistically significant, the countries being Italy and Austria respectively. As with the earnings result, the largest effect amongst the eight countries is obtained for Italy, where a one percentage point increase in the rate of change of unemployment leads to an increase of roughly one-sixth of a standard deviation in effort (similar to the magnitude of the British unemployment effect found earlier). It would therefore seem that effort does not respond to local unemployment rates in a consistent way, and where there is an effect, it is much smaller than, for example, the influence of earnings. As far as efficiency wage considerations are concerned, it appears as though individuals are influenced more by the wage that they receive than by outside variables such as the unemployment rate.

Briefly considering the remaining coefficients in Table A1, the result that females supply more effort than males *ceteris paribus* is replicated in every single country, seven of the eight coefficients being statistically significant (the exception being Hungary). The significant coefficients are mostly of a similar size, which in turn is similar to the effect found in the British data, although Italian women, and to a lesser extent German women, do seem to supply a higher level of effort relative to men than in other countries.

With respect to age, a single linear age variable is found to perform better than a

[17] The notes following Table A1 give details of the derivation of the unemployment variable in each country.

quadratic in four of the countries, while the reverse is true in the other four countries. Of the linear age effects, the British result that effort declines with age is replicated in each case, the effect being statistically significant in the United States. In the quadratic cases, a U-shaped age-effort relationship is observed for Italy, Norway and Germany, statistically significant in the first two cases, while in Hungary an inverted U-shape is suggested by the data.

The strongest education results are found for Italy and Ireland, but offer differing conclusions. In Italy, the negative coefficients increase monotonically in absolute size as the level of education rises, re-inforcing the British result that the better educated supply lower effort levels. The coefficients on the top two education levels are statistically significant. In Ireland, however, the reverse is found, whereby all of the education coefficients are positive, the highest two again being statistically significant. In Austria, the lowest education level is associated with significantly lower effort, relative to the 'no qualifications' category, but no other significant education coefficients are found. The insignificant coefficients in the Netherlands, consistently, and in Hungary and the United States, on the whole, do at least fit with the British result of lower effort among the better educated.

Turning to the two household variables, the results here also show some differences across countries. The variable indicating individuals with a non-working spouse attracts a positive coefficient in three countries, and a negative one in four, although both of the significant coefficients observed, in Ireland and Italy, are positive, in line with the British results. The responsibility for household duties variable never comes close to obtaining a significant coefficient in any of the eight countries.

Finally, considering the working condition variables, reports of hard physical work or stressful work are strongly associated with the dependent variable in every country, as indeed was the case for Britain. The likely reason, as advanced above, is that such conditions will influence reported exhaustion levels, and so represent important control variables in the estimated equations. The remaining working condition variables occasionally attract statistically significant coefficients, for example, unhealthy conditions having a significantly positive effect in three countries. The working condition results for the Netherlands are the closest to those found for Britain, in that dangerous work and unpleasant conditions attract significantly negative coefficients.

In conclusion, the effort equations for the eight countries reveal some degree of accordance as to the determinants of workers' effort. In particular, the key result of this chapter, that effort and earnings are significantly positively correlated, emerges in a number of cases, increasing the faith that can be placed in this result's validity and robustness.

Table 3.A1: Effort Equations for Other Countries

Two Stage Least Squares, Dependent Variable=Effort[1]

Variable	Austria	Germany	Hungary	Ireland
log earnings$_{iv}$	0.383 (0.176)**	0.363 (0.197)**	0.795 (1.190)	-0.127 (0.169)
female	0.200 (0.097)**	0.380 (0.119)**	0.442 (0.367)	0.255 (0.124)**
age	-0.002 (0.004)	-0.023 (0.021)	0.063 (0.039)	-0.001 (0.004)
$(age)^2/100$	-	0.028 (0.026)	-0.087 (0.040)**	-
education 1[2]	-0.250 (0.099)**	-0.130 (0.283)	0.006 (0.280)	0.882 (0.581)
education 2	-0.056 (0.133)	-0.165 (0.294)	-0.108 (0.372)	0.749 (0.582)
education 3	-0.049 (0.219)	-0.436 (0.356)	-0.246 (0.479)	0.995 (0.575)*
education 4	0.063 (0.160)	0.097 (0.323)	-0.385 (0.832)	1.257 (0.579)**
education 5	0.016 (0.195)	0.129 (0.415)	-	-
education 6	-	-0.161 (0.346)	-	-
household duties	0.074 (0.081)	0.015 (0.083)	0.116 (0.141)	-0.029 (0.113)
non-working spouse	-0.120 (0.092)	0.122 (0.100)	-0.008 (0.147)	0.182 (0.110)*
union member	-0.048 (0.064)	-0.110 (0.076)	0.014 (0.100)	0.072 (0.096)
hard physical work	0.275 (0.030)**	0.153 (0.042)**	0.114 (0.045)**	0.233 (0.036)**
stressful work	0.282 (0.033)**	0.294 (0.043)**	0.207 (0.038)**	0.321 (0.048)**
boring work	-0.071 (0.040)*	0.050 (0.045)	0.088 (0.039)**	-0.009 (0.044)
dangerous work	0.006 (0.032)	0.034 (0.044)	-0.004 (0.043)	-0.079 (0.051)
unhealthy conditions	0.024 (0.030)	0.077 (0.040)*	0.039 (0.059)	-0.016 (0.055)
unpleasant conditions	-0.004 (0.034)	0.059 (0.045)	0.043 (0.061)	0.086 (0.051)*
Δ(unemployment)[3]	-0.283 (0.145)*	-0.215 (0.136)	-	-0.042 (0.043)
level of unemp.[3]	-	-	0.055 (0.059)	-
constant	-1.699 (1.462)	-1.062 (1.440)	-6.031 (9.616)	1.798 (1.630)
controls	occupation (6) firm size (4)	occupation (6)	occupation (6)	-
no. of observations	750	390	508	420
R^2	0.316	0.321	0.143	0.263
RSS	403.101	153.997	327.721	250.053
instruments	industry (12) age^2 hours	hours	industry (2)	industry (11) age^2 hours
F test of exclusion restrictions [d.f.]	1.111 [13, 708]	-	0.012 [1, 482]	1.700 [12 ,390]

[123] See notes that follow tables.

Table 3.A1 (continued)

Variable	Italy	Netherlands	Norway	United States
log earnings$_{iv}$	1.265 (0.499)**	-0.178 (0.130)	0.170 (0.101)*	0.254 (0.122)**
female	0.557 (0.152)**	0.211 (0.101)**	0.223 (0.084)**	0.210 (0.105)**
age	-0.086 (0.041)**	-0.003 (0.003)	-0.027 (0.014)*	-0.006 (0.003)*
$(age)^2/100$	0.082 (0.046)*	-	0.031 (0.017)*	-
education 1	-1.025 (0.685)	-0.281 (0.299)	0.070 (0.081)	-0.035 (0.117)
education 2	-1.118 (0.686)	-0.195 (0.300)	-0.077 (0.084)	-0.204 (0.164)
education 3	-1.171 (0.692)*	-0.196 (0.306)	0.084 (0.100)	-0.193 (0.155)
education 4	-1.458 (0.730)**	-0.133 (0.316)	-	0.072 (0.175)
education 5	-	-	-	-
education 6	-	-	-	-
household duties	0.020 (0.153)	-0.038 (0.085)	0.065 (0.063)	0.038 (0.079)
non-working spouse	0.245 (0.143)*	-0.101 (0.068)	-	-0.150 (0.132)
union member	-0.129 (0.111)	0.054 (0.066)	-0.109 (0.054)**	-0.096 (0.100)
hard physical work	0.173 (0.055)**	0.171 (0.033)**	0.166 (0.027)**	0.217 (0.038)**
stressful work	0.315 (0.049)**	0.356 (0.036)**	0.316 (0.032)**	0.265 (0.040)**
boring work	-0.017 (0.055)	0.017 (0.041)	-0.017 (0.030)	-0.021 (0.039)
dangerous work	0.028 (0.055)	-0.070 (0.037)*	-0.028 (0.028)	-0.014 (0.041)
unhealthy conditions	-0.072 (0.069)	0.079 (0.031)**	0.064 (0.031)**	0.001 (0.052)
unpleasant conditions	0.059 (0.057)	-0.079 (0.029)**	0.023 (0.027)	0.015 (0.049)
Δ(unemployment)[3]	0.145 (0.062)**	0.081 (0.059)	-	-0.095 (0.073)
level of unemp.[3]	-	-	0.014 (0.009)	-
constant	-4.376 (2.973)	3.956 (1.327)**	0.219 (1.128)	-0.437 (1.150)
controls	occupation (7) firm size (6)	occupation (6)	occupation (6) firm size (4)	occupation (7) firm size (6)
no. of observations	379	538	807	499
R^2	0.171	0.354	0.253	0.198
RSS	294 .972	213.624	354.231	239.169
instruments	hours	hours age^2	industry (7) hours	industry (12) age^2 hours
F test of exclusion restrictions [d.f.]	-	0.183 [1, 513]	0.472 [7, 773]	1.177 [13, 455]

Notes to Appendix Tables

1. Standard errors in parentheses. **=significant at 5% or better, *=significant at 1%.

2. The omitted education category for each country is the lowest level given (typically 'no qualifications') in each case. The next lowest education category is the included variable 'education 1,' with education levels then rising through the included variables, up to a maximum of 'education 6.' Higher numbers therefore represent a higher level of education in each country, though education levels with the same number in different countries are not necessarily comparable. These education levels represent, in detail for each country:

Austria	omitted category	compulsory school
	education 1	compulsory school with vocational training
	education 2	secondary technical or trade school - lower level
	education 3	secondary - higher level without certificate
	education 4	secondary - higher level with certificate
	education 5	university completed
Germany	omitted category	no qualifications
	education 1	lower secondary school qualification
	education 2	middle school qualification / vocational training
	education 3	certification from a secondary technical/trade school
	education 4	abitur
	education 5	higher degree below university
	education 6	university degree
Hungary	omitted category	no qualifications
	education 1	primary education
	education 2	vocational training
	education 3	secondary school completed
	education 4	university

Ireland	omitted category	no qualifications
	education 1	primary education
	education 2	secondary education
	education 3	group/intermediate/leaving certificate, or equivalent
	education 4	diploma or degree
Italy	omitted category	no qualifications
	education 1	elementary school completed
	education 2	lower middle school completed
	education 3	upper middle school completed
	education 4	university graduate
Netherlands	omitted category	no qualifications
	education 1	primary education
	education 2	extended education
	education 3	secondary education
	education 4	university
Norway	omitted category	primary education
	education 1	secondary education
	education 2	high school
	education 3	college or university
United States	omitted category	less than high school
	education 1	high school
	education 2	junior college
	education 3	bachelor
	education 4	graduate

3. Region of residence is given for all respondents in the ISSP survey. Regional unemployment statistics were examined to determine whether or not each respondent lived in a high unemployment region. In the case of Austria, Germany, Italy, the Netherlands and the United States, regional unemployment data was obtained for 1989, the year of the ISSP survey, and 1988, so that unemployment change variables could be calculated. For the remaining three countries, however, regional unemployment data could not be obtained for these particular years. Census data from the nearest available consecutive years to 1989 was used to provide a regional unemployment change variable for Ireland (1987-1986), while for Hungary a single year's data was obtained (1990), so that unemployment was enetered as a levels variable in this case. For the remaining country, Norway, no regional statistics could be found, and therefore the ISSP data itself was used to supply an estimate of the regional unemployment rates, by dividing the number of respondents who claimed to be unemployed by the total of the employed and the unemployed in each region. In the seven countries where data was available, the sources were as follows:

Austria	Statistisches Handbuch fur die Republik Osterreich
Germany	Statistical Office of the European Community
Hungary	1990 Population Census Summary Data, Hungarian Central Statistical Office
Ireland	1991 Census, Volume 4 - Principal Economic Status and Industries, Irish Central Statistics Office
Italy	Statistical Office of the European Community
Netherlands	Statistical Office of the European Community
United States	Geographic Profile of Employment and Unemployment, US Bureau of Labor Statistics

References

Akerlof, G.A. (1982) 'Labor Contracts as Partial Gift Exchange,' *Quarterly Journal of Economics*, 97, 543-569.

Arai, M. (1994) 'Compensating Wage Differentials Versus Efficiency Wages: An Empirical Study of Job Autonomy and Wages,' *Industrial Relations*, 33, 249-262.

Basmann, R. (1960) 'On Finite Sample Distributions of Generalized Classical Linear Identifiability Test Statistics,' *Journal of the American Statistical Association*, 55, 650-659.

Becker, G.S. (1985) 'Human Capital, Effort, and the Sexual Division of Labor,' *Journal of Labor Economics*, 3, S33-S58.

Bielby, D.D. and Bielby, W.T. (1988) 'She Works Hard for the Money: Household Responsibilities and the Allocation of Work Effort,' *American Journal of Sociology*, 93, 1031-59.

Blanchflower, D.G. (1991) 'Fear, Unemployment and Pay Flexibility,' *Economic Journal*, 101, 483-96.

Bluestone, B. (1974) 'The Personal Earnings Distribution: Individual and Institutional Determinants,' Ph.D. thesis, University of Michigan.

Borg, G. (1962) *Physical Performance and Perceived Exertion*, Lund (Sweden), Gleerup.

Brown, C. (1980) 'Equalizing Differences in the Labor Market,' *Quarterly Journal of Economics*, 95, 113-134.

Cappelli, P. and Sherer, P.D. (1988) 'Satisfaction, Market Wages, and Labor Relations: An Airline Study,' *Industrial Relations*, 27, 56-73.

Cappelli, P. and Robinson, M.D. (1991) 'An Interplant Test of the Efficiency Wage Hypothesis,' *Quarterly Journal of Economics*, 106, 768-87.

Clark, A.E. (1992) 'Job Satisfaction and Gender: Why are Women so Happy at Work?' University of Essex, Economics Department Working Paper 415.

Clark, A.E. and Oswald A.J. (1994) 'Unhappiness and Unemployment,' *Economic Journal*, 104, 648-59.

Drago, R. (1991) 'Incentives, Pay and Performance: A Study of Australian Employees,' *Applied Economics*, 23, 1433-1446.

Drago, R. and Heywood, J.S. (1992) 'Is Worker Behaviour Consistent with Efficiency Wages?' *Scottish Journal of Political Economy*, 39, 141-153.

Duncan, G.J. and Holmlund, B. (1983) 'Was Adam Smith Right After All? Another Test of the Theory of Compensating Differentials,' *Journal of Labor Economics*, 1, 366-379.

Fairris, D. and Alston, L.J. (1994) 'Wages and the Intensity of Labor Effort: Efficiency Wage Versus Compensating Payments,' *Southern Economic Journal*, 61, 149-160.

Guest, D.E. (1990) 'Have British Workers Been Working Harder in Thatcher's Britain? A Re-consideration of the Concept of Effort,' *British Journal of Industrial Relations*, 28, 293-312.

Hausman, J. (1983) 'Specification and Estimation of Simultaneous Equations Models,' in Z. Griliches and M. Intriligator, eds. *Handbook of Econometrics*, Amsterdam, North-Holland.

Hogan, J.C. and Fleishman, E.A. (1979) 'An Index of the Physical Effort Required in Human Task Performance,' *Journal of Applied Psychology*, 64, 197-204.

Holzer, H.J. (1990) 'Wages, Employer Costs, and Employee Performance in the Firm,' *Industrial and Labor Relations Review*, 43, S147-S164.

Idson, T.L. (1990) 'Establishment Size, Job Satisfaction and the Structure of Work,' *Applied Economics*, 22, 1007-18.

Lawler, E.E. (1971) *Pay and Organizational Effectiveness*, New York, McGraw-Hill.

Leonard, J.S. (1987) 'Carrots and Sticks: Pay, Supervision, and Turnover,' *Journal of Labor Economics*, 5, S136-S152.

Levine, D.I. (1992) 'Can Wage Increases Pay for Themselves? Tests with a Production Function,' *Economic Journal*, 102, 1102-1115.

Lindbeck, A. and Snower, D.J. (1988) 'Job Security, Work Incentives and Unemployment,' *Scandinavian Journal of Economics*, 90, 453-474.

Lucas, R.E.B. (1977) 'Hedonic Wage Equations and Psychic Wages in the Returns to Schooling, *American Economic Review*, 67, 549-58.

Machin, S. and Manning, A. (1992) 'Testing Dynamic Models of Worker Effort,' *Journal of Labor Economics*, 10, 288-305.

Maddala, G.S. (1983) *Limited-Dependent and Qualitative Variables in Econometrics*, Cambridge, Cambridge University Press

Marin, A. and Psacharopoulos, G. (1982) 'The Reward for Risk in the Labor Market: Evidence from the United Kingdom and a Reconciliation with Other Studies,' *Journal of Political Economy*, 90, 827-53.

McNabb, R. (1989) 'Compensating Wage Differentials: Some Evidence for Britain,' *Oxford Economic Papers*, 41, 327-38.

Meijman, T., Zijistron, F., Kompier, M., Mulders, H. and Broevse, J. (1986) 'The Measurement of Perceived Effort' in Osborne, D. (ed) *Contemporary Ergonomics*, London, Taylor and Francis.

Meng, R. (1990) 'The Relationship Between Unions and Job Satisfaction,' *Applied Economics*, 22, 1635-1648.

Meng, R. (1991) 'Compensating Wages for Long-Term Job Hazards in Canadian Industry,' *Economic Letters*, 36, 331-36.

Moulton, B.R. (1986) 'Random Group Effects and The Precision of Regression Estimates,' *Journal of Econometrics*, 32, 385-397.

Moulton, B.R. (1990) 'An Illustration of a Pitfall in Estimating the Effects of Aggregate Variables on Micro Units,' *The Review of Economics and Statistics*, 72, 334-338.

Quinn, J. (1975) 'The Microeconomics of Early Retirement,' Ph.D. thesis, M.I.T.

Rebitzer, J.B. (1987) 'Unemployment, Long-Term Employment Relations, and Productivity Growth,' *The Review of Economics and Statistics*, 69, 627-35.

Rebitzer, J.B. and Robinson, M.D. (1991) 'Employer Size and Dual Labour Markets,' *The Review of Economics and Statistics*, 73, 710-15.

Schor, J. and Bowles, S. (1987) 'Employment Rents and the Incidence of Strikes,' *The Review of Economics and Statistics*, 69, 584-592.

Shapiro, C. and Stiglitz, J.E. (1984) 'Equilibrium Unemployment as al Worker Discipline Device,' *American Economic Review*, 74, 433-44.

Smith, A. (1976) *An Inquiry into the Nature and Causes of the Wealth of Nations* (first published 1776), ed. Campbell, R.H., Skinner, A.S. and Todd, W.B., Oxford, Clarendon Press.

Smith, R.S. (1979) 'Compensating Wage Differentials and Public Policy: a Review,' *Industrial and Labor Relations Review*, 32, 339-52.

Ulman, L. (1992) 'Why Should Human Resource Managers Pay High Wages?' *British Journal of Industrial Relations*, 30, 177-212.

Vroom, V. (1964) *Work and Motivation*, New York, Wiley.

Wadhwani, S.B. and Wall, M. (1991) 'A Direct Test of the Efficiency Wage Model Using UK Micro-Data,' *Oxford Economic Papers*, 43, 529-548.

Chapter 4

Workers' Effort in Unionised Workplaces:

Effort Choices and Establishment Characteristics

4.1 : Introduction

The aim of this chapter is to investigate the influences on, and determinants of, workers' effort choices within the union sector. Does effort respond positively or negatively to the degree of unionism? How do managers react to the presence of a union, and how do management techniques to elicit effort from their employees interact with union policies? The Workplace Industrial Relations Survey (WIRS) contains a rich set of establishment characteristics with which to answer some of these questions.

Despite extensive secondary analysis of the WIRS data sets, no study has yet, to the best of my knowledge, made use of the effort variable. Union effects on wages are very well documented, while a growing literature has examined the effects of unions on variables such as wage dispersion, employment, productivity, profitability and investment.[1] While such work has improved our understanding of the production process in unionised firms, an analysis of the determinants of workers' effort would be a valuable addition to such knowledge. In particular, the literature as it stands concentrates on the links between unionism and outcomes of the production process, as the list of variables above shows. The intention of this study is to look more closely at an input to the production process: effort.[2] Allen (1988) makes this point, describing studies that examine the union impact on productivity as 'black box' studies, which can tell us nothing about the mechanisms

[1] Metcalf (1993, 1994) reviews this literature.

[2] I believe it is important to distinguish between effort, which is an input to the production process, and productivity, which is output per head. Some studies have used the terms interchangeably, or have used productivity data to proxy effort. Many variables besides effort, however, will determine the final level of output, and hence productivity.

through which unions affect productivity, be they through effort or some other process.[3] Therefore this analysis will consider the internal workings of the firm, rather than just what comes out at the end.

The previous chapter was mainly concerned with characteristics of the individual and how these were associated with his or her effort choices. As the above discussion suggests, this chapter will consider the workplace characteristics which influence effort choices. The original intention was simply to see how the average effort level in unionised workplaces compared with that in similar non-union establishments. This proved to be impossible to analyse, given the data available in the WIRS data set, however. The effort variable was found in the Worker Representative's Questionnaire. The worker representative who answered the question was either a union official, or, in the absence of a recognised union, a representative of the workforce who sat on management committees. In practice, this turned out to be a union official in the vast majority of cases, only twelve non-union workplaces in the entire WIRS sample having the relevant committees to supply such a worker representative. Once further observations had been deleted because of missing values on other explanatory variables of interest, only five useable non-union observations remained, and so a union-non-union comparison was not possible.

The focus of the study therefore changed to an analysis of how effort varies with the degree of unionism, as measured by union density at the establishment level. Consideration was also given to the existence of closed shops and whether management recommends union membership to its employees. To provide more detail concerning the

---

[3] Allen hypothesises that union-nonunion productivity differences could be related to human resource policies. The results presented below cast some light on this issue, describing how union presence and the use of human resource management techniques interact to affect effort.

internal workings of the firm, this union effect was allowed to vary with certain characteristics of the workplace. In particular, consideration was given as to how management techniques to elicit effort, such as attempts to raise the cost of job loss, and human resource management (HRM) methods, interact with the presence and influence of trade unions. The chapter therefore adds to the efficiency wage literature concerning the link between wages or cost of job loss and effort, as well as to the literature describing the impact of HRM.

The next section describes in greater detail the hypotheses to be tested. First it looks at why cost of job loss and HRM variables should influence workers' effort choices, and then goes on to detail how such influences could be affected by the presence of a trade union. The third section describes the data, followed by a results section. The chapter ends with a summary of the main findings.

4.2 : Hypotheses

As described in the introduction, the aim of this study is to acquire some evidence concerning the impact of unions on workers' effort choices,[4] and how this effect interacts with management policies for extracting effort from their workers, through raising the cost of job loss or employing HRM techniques.

The first thing to make clear is how these management policies are supposed to provide incentives for workers to increase their effort. The effect of raising the cost of job loss is perhaps best described in terms of the shirking version of the efficiency wage hypothesis.[5] The cost of job loss is the difference between the utility value of being in the current job, and that of the next best alternative. The utility gained from the current job is primarily determined by the wage received, while the alternative level of utility is affected by the wages on offer, the chances of securing another job and whatever utility is derived, both in terms of replacement income and leisure time, from unemployment. Hence, assuming a linear utility function, the cost of job loss at firm i can be represented by cjl_i, where

$$cjl_i = w_i - [ub + (1-u)\bar{w}_i] \qquad (1)$$

W_i is the current firm's wage, u the probability of being unemployed, as given by the local unemployment rate, b the utility derived from being unemployed, and $\bar{w}_i$ the wage the individual expects to receive in alternative employment.

[4] Implicit in this sentence is the assumption that workers choose their own effort levels, rather than have effort levels set for them by management. This will be the case under the reasonable assumption that effort is unobservable or unverifiable. However, at various points in this chapter it will be acknowledged that the results obtained could be reflecting management forcing new effort levels onto their employees.

[5] See, for example, Shapiro and Stiglitz (1984).

Raising the cost of job loss provides an incentive to supply effort, on the assumption that the firm has a certain probability of observing those employees who are not supplying effort, or shirking, and will dismiss any such individuals it finds. Then, the higher the cost of job loss, the more workers stand to lose if they are caught shirking, and hence the more likely they are to supply effort. The firm's principal means of raising the cost of job loss is through increasing the wage that it pays.

This 'cost of job loss effect' can in turn be influenced by the presence of a trade union. Discipline and dismissal procedures can make sacking a worker a lot more difficult for a firm, even if the individual in question is guilty of shirking. It is reasonable to assume that such procedures will be more deeply entrenched, formalised and used in unionised firms, such activity being one of the primary functions of unions. Hence the threat of dismissal is not so potent when applied to employees in a unionised firm, leading to the hypothesis that variables reflecting cost of job loss should have a smaller effect on effort as union presence becomes stronger.

There has been a growing awareness in the industrial relations literature of the need to understand the effects of HRM. Theory would tend to predict that there would be a beneficial effect on performance, through the utilisation of employee knowledge concerning the intricacies of the production process, which can often be more detailed than management could ever obtain otherwise. Thus, the planning and organisation of work should be significantly improved. However, since the object of the current study is the input, effort, rather than the output, firm performance, the second benefit of HRM frequently mentioned in the literature is likely to be more relevant here. The hypothesised effect is that the presence of HRM techniques raises employees' commitment to the organisation, which manifests itself in a higher level of motivation and hence work effort.

An alternative interpretation of the effort response to HRM is that it is demand led rather than a supply response. Thus, it has been argued that HRM techniques are designed specifically by management to intensify the workload on employees, 'squeezing' more effort out of them. With the data set and type of analysis being used, it will be difficult to say whether HRM works through forcing employees to work harder, or creating a situation where they optimally choose to supply more effort. However, some comments will be made at various points in the chapter.

Not all authors discuss only a positive link between HRM and economic outcomes. With respect to the argument that the utilisation of employee knowledge will improve the production process, a counter-argument is that the increase in available information will simply slow down the bureaucratic decision-making process further, resulting in less efficient production plans. As for the employee motivation argument, in response it could be argued that the introduction of peer pressure from colleagues or employment losses resulting from additional production may reduce employee commitment to the firm and hence their effort.

In the introduction to their own study, Fernie and Metcalf (1995) survey the empirical HRM literature. They describe the generally positive, but often modest, effect of the adoption of HRM techniques on various economic and industrial relations outcomes. Their own results, using the 1990 WIRS data, reveal that communication between management and workers, use of employee involvement, and the existence of profit sharing or merit pay schemes are all positively correlated with either productivity levels or productivity growth at establishments. However, industrial relations variables, such as the quit rate, the degree of absenteeism and the climate of management-employee relations, seem to be little affected by the use of HRM techniques; if anything they are negatively affected. Using these results, Fernie and Metcalf calculate typical values for these outcome variables, for three 'benchmark' establishments; a union workplace, an HRM workplace, and an authoritarian workplace (a workplace with neither of these characteristics). The results show that the HRM workplaces have the best productivity performance of the three, but the worst industrial relations. These results would tend to suggest that the favourable economic outcomes associated with HRM are not an optimal supply response by workers, but are squeezed from them, this being the management's original motivation for introducing such schemes in the first place. If they were a supply response, caused by greater organisational commitment on the part of the employees, then it would be reasonable to expect industrial relations to also benefit from the presence of HRM methods. With the alternative hypothesis, it is easy to understand how industrial relations could worsen, if work pressure has been intensified.

The work by Fernie and Metcalf (1995) therefore considers both union and HRM workplaces, but not the possible interaction of the two, although an appendix does briefly consider this point. They report that employee involvement techniques have beneficial effects on performance indicators in both union and non-union sectors, although slightly larger in the latter. However, the use of communication between management and employees only seems to significantly increase productivity in unionised workplaces. Merit pay similarly only influences economic outcomes in the union sector, although profit-sharing seems to have beneficial effects in all workplaces. Finally, the establishment of joint consultative committees seems to have no effect in any workplaces.

Why should HRM techniques have different effects in unionised and non-unionised firms?[6] Cooke (1992) outlines possible reasons. As with the effects of HRM itself, the theoretical arguments concerning the interaction with unionism can go either way. Assuming, on balance, that HRM procedures have a positive influence on effort, it could be argued that traditional, adversarial unions would negate such benefits, through restrictive practices which prevent the optimal strategies being pursued. In addition, unions may persuade their members not to participate in any management schemes, unless other demands are met, thus holding the HRM plan 'hostage.' On the other hand, the 'collective voice' role of unions, which describes how unions provide an efficient means of transmitting employee ideas and feelings to management, may create a more favourable situation for HRM methods to be successful.[7] If management is better informed of employee opinions, this is likely to lead to a more efficient design for any programme developed. As for the motivation impact of HRM, this is likely to be greater if employees are convinced that their participation is worthwhile and their ideas are really being listened to, which seems more likely when a union is providing an efficient voice mechanism.

Only three empirical studies of which I am aware have explicitly analysed these ideas to obtain estimates of the interaction between unionisation and the use of HRM. Kelley and Harrison (1992) estimate regression equations for the production time per unit of output. Controlling for the skill requirements of the job, they find that production time is actually significantly longer in the presence of joint labour-management problem-solving

---

[6] Given that the WIRS data does no more than reveal whether a workplace uses a particular HRM technique or not, and does not go into details concerning content, this study, through necessity, assumes that such procedures are the same in union and non-union workplaces. Of course, this need not be the case. HRM in a non-union workplace may be implemented as part of a whole strategic package by management, whereas its introduction into a union workplace may be a tactical move by managers, adding something to the existing industrial relations set-up in a piecemeal fashion, to gain an advantage. It is not possible to control for such potential differences, though the reader should be aware of them.

[7] See Freeman and Medoff (1984) for a full description of this role for unions.

committees. This is the case in all plants, although particularly in non-unionised ones. The shortest production times are found in unionised branches of multi-unit firms that do not have such committees.

Cooke (1992) makes use of management responses to surveys which include a question as to the degree of product quality improvement over a five year period, and finds that the presence of a 'work group participation programme' is associated with significantly improved product quality, the effects being very similar in magnitude in union and non-union firms. However, the effect is removed in unionised firms where the scheme is administered solely by management rather than jointly with the union, such firms achieving quality improvements insignificantly different from those in non-union firms with no participation schemes. The union effect is also mitigated, even for a jointly administered programme, if the firm subcontracts out any jobs, or has engaged the union in concession bargaining, which are taken by Cooke to represent situations in which union trust, and hence commitment to the participation scheme, will be reduced. He therefore concludes that the voice effects described above offset any negative impact which unions may have on the success of HRM programmes, but if unions are not involved in their administration, or if union trust in management is low, the voice benefits are lost and unions prove to be a hindrance to the establishment of effective HRM schemes.

In a following paper, Cooke (1994) analyses differences in performance, defined as value added minus wage costs, between firms, across different permutations of the presence or otherwise of unions, work teams and profit-sharing plans. Comparing non-HRM firms first, performance appears to be significantly higher, by 13%, in unionised firms relative to their non-union counterparts. The adoption of work teams widens this differential, significantly improving performance in unionised firms, but having an insignificant effect in similar non-union firms. The use of profit-sharing as a sole HRM technique has very similar positive effects in union and non-union firms, although it seems to interact more successfully with the work teams when no union is present. Overall, of the eight permutations, the most successful, in terms of performance, are unionised firms with work teams but no profit-sharing.

The main conclusion that can be drawn from these studies is therefore that there does not appear to be any reduction in the effectiveness of HRM programmes when introduced in union firms relative to non-union firms, with the proviso expressed by Cooke (1992) that the scheme must be jointly administered with the union. The usual finding is a similar impact whether unions are present or not, with the exception, in the Cooke (1994) study, of the use of work teams, which are found to be more effective when adopted in a unionised firm.

Given the small number of studies examining the interaction between HRM and unionisation, and the fact that the results have yet to provide unambiguous answers, it was felt that this would be a useful question to consider here. In addition, while the previous studies provide information as to whether, and in what setting, HRM techniques are successful, the current study, by focusing on an input, effort, rather than an output, typically some measure of productivity or performance, can provide some indication as to the mechanisms through which HRM works. The next section describes the data and estimation technique to be used for this analysis.

4.3 : Data

a.) The Data Set

The data used come from the third (1990) Workplace Industrial Relations Survey (WIRS), which is the largest and most detailed interview-based survey of industrial relations issues available. It is a nationally representative survey of 2061 establishments in Britain with at least 25 employees. All sectors are covered, with the exceptions of agriculture, forestry and fishing and coal mining. The response rate to the survey was 83%. Further details of the survey can be found in Millward *et al* (1992).

b.) The Effort Variable

The effort variable is based on the following question from the WIRS Manual Worker Representative Questionnaire:

'Generally, how does the intensity or the pace of work for most manual workers here compare with what it is in other similar establishments.'

Answers were given on a five-point scale, ranging from 'a lot lower' to 'a lot higher.' A similar question was asked of the non-manual representative, but no use was made of this in the main part of the analysis, since non-manual workers are defined by WIRS to include such categories as managers and supervisors. Since a major component of the study was to investigate the impact of management schemes on the motivation of their employees to supply effort, it would not be clear how to interpret the results for non-manual workers in this light, when the dependent variable would partly consist of the effort levels of management themselves.

The distribution of the effort variable across respondents is as follows.

Response	Percentage of Respondents
a lot lower	1
a little lower	4
about the same	46
a little higher	25
a lot higher	23

Since the question is concerned with relative effort, there should be an even distribution, assuming the sample is random, with as many workplaces reporting effort levels above the average as there are reporting below the average. Clearly, there appears to be some over-estimating of effort in this sample, given that so many more workplaces report effort above the average compared to below the average. It would therefore seem appropriate to spend some time justifying the use of this effort variable.

Perhaps the most important piece of evidence is the correlation between the effort variable and a productivity question, asked of the various respondents to the WIRS questionnaire. The productivity question is of exactly the same form as the effort question above, and so it would be expected that where effort is high relative to other workplaces, productivity should also be relatively higher. While this correlation will be far from perfect, given that a multitude of factors other than the effort of the workforce will influence the level of productivity in a particular workplace, it would seem reasonable to expect the correlation coefficient to be at least significantly positive, if the effort variable is to be accepted. Looking first at the opinions of the manual worker representatives concerning productivity, these correlate significantly with their views concerning the levels of effort (r=0.46).[8] This may be to be expected, however, since those with an inflated view of their effort may also overestimate the level of productivity at the workplace. Of more importance are the significantly positive correlations, at the 1% level, between the worker representatives' opinions of their effort levels, and the opinions of the managers (r=0.20) and financial managers (r=0.19) concerning the relative productivity performance of their respective workplaces. It therefore seems that, although there appears to be an upward bias in the effort data, the effort variable still behaves as we would wish it to, so that when a worker representative claims that effort is high in a workplace, both a senior manager and the financial manager say, totally independently of each other and the worker representative, that productivity is also high in that workplace. Hence, although the effort variable is biased upwards, it seems to retain its ordinal properties, and since all the estimation below is by ordered probit analysis, it is only the ordering of the data that is important, not the actual numbers. Ordered probit analysis using the effort variable as a dependent variable should therefore give unbiased estimates.

A second piece of evidence comes from comparing the answers given, independently, by the manual and non-manual worker representatives to the effort question. If variations in these answers across respondents are purely random measurement error, then there need be no reason why the two respondents in any particular workplace should be subject to the same degree of mis-reporting, and hence the answers given by the two respondents in each workplace should not be correlated with each other. A finding of a significant positive correlation between manual and non-manual reported effort would therefore seem to indicate, for whatever reason, a high-effort

[8] The correlation coefficient reported here, and the ones which follow, are Spearman rank-order correlation coefficients.

workplace. The reasons should be determined by the empirical analysis below. In fact, the correlation does turn out to be significantly positive at the 1% level (r=0.18). The table below illustrates this correlation by revealing that, in general, when the manual worker representative reports high effort, the non-manual representative does likewise.

How Effort Compares to Similar Workplaces: Manual and Non-Manual Workers

		non-manual				
		a lot lower	a little lower	about the same	a little higher	a lot higher
	a lot lower	0	0	2	2	2
	a little lower	0	2	7	1	3
man-	about the same	1	7	87	40	36
ual	a little higher	0	0	24	29	27
	a lot higher	0	1	29	25	23

The χ^2 statistic for this table is 28.56, which is significant at the 5% level, thus revealing a relationship between the responses given by the two worker representatives across all workplaces.

The next piece of evidence concerns the correlation between the effort variable described so far, and another asking the worker representatives' opinions as to how the level of effort has changed at their workplaces compared to three years previously, answers being given on the same five-point scale.[9] The relationship between the two effort variables is revealed in the table below.

[9] Use was not made of this alternative effort variable in the main analysis below, because it is representing the *change in* effort, and so would not be appropriate as a dependent variable in a regression against variables mostly in *levels*, as are available in the data set.

Effort Levels Relative to Similar Workplaces and the Position 3 Years Ago

		relative to 3 years ago				
		a lot lower	a little lower	about the same	a little higher	a lot higher
	a lot lower	0	2	2	3	1
rel. to	a little lower	3	2	17	6	2
work-	about the same	5	24	125	102	62
places	a little higher	1	9	37	57	67
	a lot higher	2	1	29	22	106

The χ^2 statistic for this table is 149.19, which is clearly highly significant. If the manual worker representative reports a high level of effort relative to similar establishments, they are also much more likely to report a larger increase in effort, relative to three years previously. A perfect correlation between the two effort variables would not be expected, since a workplace could have improved its own effort position, and still be lagging behind other workplaces. Overall, however, it would be expected that where effort had significantly improved, it would often be at a higher level than at similar workplaces. This is indeed what the table reveals. The correlation coefficient between the two effort variables is also highly significant and positive (r=0.39). The fact that the chosen effort variable behaves in this predicted way further adds to its credibility. This piece of evidence is perhaps less persuasive then the previous ones, however, since an alternative interpretation of the results is simply that the respondents confuse changes and levels information.

The final piece of evidence in the data is revealed by the plausibility of the results on the control variables in the estimated equations. Firm predictions can be made as to how effort should vary with respect to certain control variables. If the reported effort level is only representing measurement error and not actual effort, then there is no reason why it should vary in any predictable way with any of the explanatory variables. If the predictions are confirmed, however, this would lend support to the dependent variable being a good measure of effort, thus increasing the confidence in the results for the key variables of interest. In general the predictions for the effects of the control variables are confirmed, as will be discussed further in the results section.

All of the points above suggest that the reported effort level is behaving as would be expected, if it is an accurate indicator of actual effort. When these results are considered together with the literature describing the legitimacy of using self-reported effort data, as outlined in the previous chapter, it seems reasonable to accept the validity of the WIRS effort variable, that is the 'work intensity' question.

c.) Explanatory Variables[10]

As described in Section 2, the primary focus of the study is to examine how effort varies with an indicator of union power, and to see whether this union effect interacts with management schemes to increase effort, such as raising the cost of job loss or HRM techniques.

The indicator of union power used is union density. As mentioned above, the analysis throughout is only conducted with reference to manual workers, and so the density variable is the proportion of manual workers at the establishment who belong to a trade union.[11] If, as hypothesised, the union effect does vary with characteristics of the workplace, then, in a standard specification which does not allow for such interactions, this effect is unlikely to be linear. Therefore, the square of union density is also entered.

---

[10] Data on all explanatory variables are taken from the main (manager's) questionnaire. Any subjective variables therefore represent the opinions of the manager who responded to the survey.

[11] This applies to all other variables in the study. Where questions are asked separately about manual and non-manual workers, only the manual data is used.

The cost of job loss is given by equation 1 above. Unfortunately, not all of the components of that equation are available in the WIRS data set, for example, the level of utility received while unemployed, b, and the alternative or 'outside' wage rate, $\bar{w}_i$. The variable actually used is the log of $w_i.u$, the product of the wage being earned and the unemployment rate. While this is clearly a major simplification of equation 1, it at least has the desirable property that it is increasing in current wages and unemployment, both of which raise the cost of job loss. Thus it is expected, given the theory outlined in Section 2, that this variable will obtain a positive coefficient in the estimated effort equations.[12]

Both components of this cost of job loss variable are available in the WIRS data set. Detailed unemployment rates are supplied at the level of the local labour market within which the establishment is located. The wage data are taken from the following question, asked of various occupational groups providing there are at least five employees in that group at the workplace:

'If all the employees in this group were listed individually in order of their gross earnings (including any bonus or overtime), which of the ranges on this card would apply to the employee in **the middle** of such a list?' (emphasis in original).

Thus the question is attempting to elicit the median earnings for particular occupational groups. The responses are banded, and so, as in the previous chapter, the mid-points of these bands are taken, with checks being made that the results are not sensitive to the value chosen for the open-ended category. Of interest to this study are

[12] Other cost of job loss variables were created that were intended to more accurately reflect equation 1, for example letting w_i equal the average wage in the standard region in which the individuals lives, and setting the utility while unemployed to be an arbitrarily chosen proportion, 50%, of w_i. Such experiments produced results very similar to those using simply $w_i.u$, however, and only the latter are reported here.

the median earnings of unskilled, semi-skilled and skilled manual workers. Since the effort variable is for the group of manual workers as a whole, the average wage for this group was calculated as the employment-weighted average of the wages for the three skill subgroups. Multiplying this number by the relevant local unemployment rate then gives the cost of job loss variable. Taking logs produces the variable actually entered into the estimated equations.

As the question reproduced above makes clear, the recorded earnings include bonus and overtime payments made to workers, and thus the results are subject to the same caveat that was identified in the previous chapter, which readers should be aware of. If managers use effort-inducing policies such as performance related pay schemes, a proportion of earnings received in such workplaces will be contingent upon performance, and will show up in the earnings variable used. Any positive correlation observed between effort and this measure of earnings may therefore be due to high effort being rewarded by high earnings. The high effort observed would therefore be attributable to the performance related pay scheme, rather than an efficiency wage effect. Equally, the presence of overtime payments in the earnings variable may lead to a misinterpretation of the results. It is hoped that respondents will consider effort exerted in a typical hour's work at their establishment, when answering about the intensity of work. If respondents consider working long hours to be working 'intensely,' however, then the use of overtime could increase the effort measure and the earnings measure, as defined, thus leading to a positive link between them that is again unrelated to efficiency wage effects. To check how relevant this argument might be, use was made of the hours variable, which asks how many hours are worked to earn the reported level of wages. The correlation coefficient between this variable and the effort variable was highly insignificant, suggesting that long hours do not lead to a report of high intensity of work. All of the equations reported below were also re-run with the hours variable included as an explanatory variable. However, none of the results were altered, and hours worked always attracted a statistically insignificant coefficient. It therefore does not appear that the presence of overtime payments, at least, in the earnings variable should affect the results to be presented below.

It can be seen that entering a cost of job loss variable as the log of the product of wages and unemployment is, in effect, imposing the restriction that these two variables take the same coefficient. This restriction was tested by running an alternative specification with the logs of wages and unemployment entered separately, thus providing a test of the appropriateness of this particular cost of job loss variable.

Turning now to the HRM variables, a number are used to signify an establishment with such policies. Following the discussion in Section 2, all are anticipated to be associated with effort. The first asked whether there had been any attempts to increase employee involvement over the previous three years. While the interpretation of this question was left to each respondent, employee involvement is usually taken to imply the empowerment of the workforce in the decision-making process; for example, giving employees a degree of autonomy to make choices as and when the need arises on the shopfloor.

Communication between managers and workers is another typically quoted aspect of HRM. WIRS provides quite detailed information on this subject, offering the respondent seven different forms of communication, with the instruction to choose as many as apply to that establishment. For the purpose of the analysis, these were grouped into three categories; two-way communication, one-way communication downwards from managers to workers, and one-way communication upwards from workers to managers.[13] Given the structure of the original question, these categories are not mutually exclusive, and a single establishment could well use all three types of communication.

The remaining two HRM variables identify establishments which have a joint consultative committee (JCC) at the workplace, comprising representatives of management and employees, and establishments which use individual-based payment by results schemes.[14] The latter ties in with the general thrust of HRM, which is to emphasise the individual and his or her relationship with the firm, although for the purpose of estimating effort equations it should be recognised this variable is slightly different. Individuals have a clear incentive, in terms of monetary reward, to input more effort when a workplace employs such schemes, whereas the link between the other HRM variables and effort is not so direct, and is based on feelings of morale and organisational commitment which the various programmes can create in the employee.

The remaining variables in the estimated equations comprise controls for other possible influences on the average effort of employees. One group of variables comprises indicators of whether the establishment has introduced organisational change in the previous three years. Such changes to the way work is organised may affect the effort required of employees, or may affect the potential effort choices open to them. The

---

[13] The two-way communication variable takes the value of one if the workplace holds at least one of the following types of meetings; management meeting employees in their individual workgroups, meeting all employees together on a monthly basis, or meeting all employees on a yearly basis. One-way communication downwards comprises use of the management chain or company newsletters, while upwards communication takes the form of suggestion schemes or surveys of employees opinions.

[14] WIRS also asks about group- and establishment-based payment by results schemes, but it was decided to focus on individual-based schemes, because of the stress placed by the HRM literature on the individual. A further question asked about the existence of merit pay, whereby payments are made on the basis of supervisors' *subjective* opinions of workers' performance. This variable was analysed, but found to have a much weaker association with effort than individual payment by results schemes, and so is not included in the subsequent analysis.

variables included signify whether or not the workplace has introduced new work practices, microelectronic technology, or new plant and machinery, as well as an indicator as to whether the manager considers that changes have been made which have increased flexibility.

A further variable identifies workplaces which have, in the opinion of the manager respondent, good or very good relations between management and employees. The existence of such relations may affect the commitment that employees feel towards their organisations, and so the level of effort that they choose to supply.

The degree of supervision can also influence workers decisions as to how much effort to supply. A direct measure of the amount of supervision conducted in the establishment, available in the WIRS data set, is the proportion of employees who are categorised as supervisors. A more indirect measure, which has been used in the literature as a proxy for supervision when a more direct measure is not available, is workplace size, on the assumption that as the workplace grows, it becomes more difficult to monitor the effort levels of individual employees. Five dummy variables, representing increasing workplace size, are therefore also included in the specification.

The remaining variables in the effort equations attempt to control for the composition of the workforce, with the limited data available in WIRS. Controls for the proportion of manual workers who are male, unskilled and full-time are included, as well as a dummy variable which takes the value of one if more than five per cent of the workforce belong to an ethnic minority.

The means of all the explanatory variables, together with a brief description, are presented in Table 1, first for the sample of 375 observations for which data on all of the variables in the table are available,[15] and then for the subsamples of these 375 observations where union density is above or below 90%, which are used in the subsequent analysis. The main reason why there are far fewer useful cases than the full WIRS sample of 2061 observations is that the effort variable is only asked where a manual worker representative is available, which is a minority of establishments in the sample. Missing wage data also have the effect of significantly reducing the available sample. As a check, the mean of each variable was calculated for the WIRS sample as a whole. The dependent variable, effort, actually has the same mean in the full sample as it has in the reduced sample of 375 observations (3.64). The other variables of interest are also very similar in the full and reduced samples. The one exception to this is the union density variable. As described earlier in the introduction, the vast majority of establishments with a worker representative, who could answer the effort question, are unionised workplaces. When the sample is limited to those establishments where the effort question is asked, therefore, almost all non-union workplaces are excluded. Thus, whereas average union density in the full sample, which includes a significant number of non-union workplaces with zero density, is 54%, this figure is 84% for the reduced sample actually used in the analysis. Hence, inferences concerning non-union workplaces should be guarded against, and the results are best interpreted as the effects of high, relative to low, union power.

The next section contains the formal statistical analysis of the relationships between effort and the explanatory variables described. Table 2 presents some simple descriptive statistics, revealing how average effort, as measured by the intensity variable, differs according to certain characteristics of establishments. With respect to the HRM

[15] All analysis which follows is based only on these 375 observations.

variables, it can be seen that they generally seem to be associated with higher effort, although the differences are small. Average effort is higher if the establishment has introduced employee involvement schemes, has a JCC, employs any of the three types of communication, or has schemes whereby manual employees are paid by results. The difference is significant at the 5% level only in the last case, however.

With respect to organisational change, effort is, on average, higher in establishments which have introduced new plant or machinery, or changed working practices, but lower where the establishment has brought in microelectronic technology or has made the work more flexible. Again, the difference is significant at the 5% level only once, for the new machinery variable.

Further analysis of the effort variable reveals that effort is highest on average in Wales (3.78) and lowest in the West Midlands (3.43) with respect to the standard regions of Great Britain. This difference represents just over one-third of a standard deviation on the effort scale. Average effort increases with establishment size, except for the two extreme sizes of establishment. Effort is highest, at 3.76, in the 500-999 employees category, although this is only one-fifth of a standard deviation on the effort scale higher than the size with the lowest average effort (50-99 employees).

With respect to industry, it is not clear how to interpret variations in the effort level. Presumably when respondents are asked to compare their own effort levels to those in similar establishments, they make their comparisons with workplaces in the same industry. It turns out that respondents in distribution, hotels and catering report effort levels which are the highest relative to other such workplaces. However, this reveals nothing about how effort in that industry compares with effort levels in other industries, if the respondents were, in effect, holding industry constant when giving their answers.

Of course, this point could apply to any variable describing characteristics of the workplace. Thus, respondents could have been comparing their effort levels to those in establishments of the same size, with the same amount of new machinery, with the same HRM programmes, and so on. If this was the case, the data would tell us little about the relative effort levels in workplaces displaying different combinations of these characteristics. However, it seems unlikely that respondents will know the effort levels in other workplaces across all permutations of such characteristics, so the data should still be useful in describing how effort varies across categories of most explanatory variables.

4.4 : Results

a.) Full Sample Results

Including all of the variables described in the previous section provides a starting point as a standard specification. Estimation is by ordered probit, which is the appropriate technique to use when the dependent variable is grouped into a small number of categories that are ranked on an ordinal scale, the actual numbers attached to the variable having little meaning. A full description of the technique can be found in the previous chapter, and in references such as Maddala (1983). The results are shown in Table 3.

Column 1 contains the results when earnings and unemployment are entered into the equation separately; that is, the cost of job loss restriction is not imposed. It can be seen that both of these variables attract the positive coefficients that are predicted by the efficiency wage literature. Both have the effect of increasing the cost of job loss, and so employees, on average, supply more effort, to make sure that they do not in fact lose their jobs due to dismissal.

The effects of the earnings and the unemployment variables are similar in size, although neither attracts a significant coefficient, at conventional significance levels. Column 2 imposes the restriction that their coefficients are actually equal, by replacing the two variables with the single cost of job loss variable, as defined above. The null hypothesis of the restriction being valid cannot be rejected using a likelihood ratio test[16],

---

[16] The likelihood ratio statistic is equal to $-2(L_1 - L_0)$, where L_0 is the unrestricted value of the maximised log likelihood, and L_1 is the equivalent maximum log likelihood when the cost of job loss restriction is imposed. Under the null hypothesis that the restriction is valid, this statistic has a χ^2 distribution, with the degrees of freedom equal to the number of restrictions (which is 1 in this case).

the χ^2 statistic being extremely small. The gain in efficiency which results from imposing a valid restriction, revealed by the lower standard error on the cost of job loss variable, means that cost of job loss is now positively and significantly associated with effort, at the 5% level.

The efficiency wage interpretation of this result is that making it more costly to shirk, in terms of a larger fall in utility if dismissed, will increase the probability that the work group choose a high effort level, on average. The estimated coefficient suggests that raising the cost of job loss by one standard deviation above its mean, holding all other variables at their mean values, will increase the probability of employees working at the highest effort level (effort=5), by over 5 percentage points, from 20.1% to 25.2%.[17] Thus, raising the cost of job loss appears to be an effective motivational tool available to management, if this interpretation of the results is correct. However, causality has not been proved, and an equally valid interpretation of a positive correlation between earnings and effort can be provided by compensating differential theory. This theory predicts that individuals who supply high levels of effort will receive high earnings in compensation, thus arguing that the causality runs from effort to earnings. The evidence presented so far does not prove in which direction the causality runs, and so cannot discriminate between the two theories. This matter will be considered further in subsequent sections.

The next key group of variables, for the purposes of this study, are the HRM variables. Examining either column 1 or column 2 of Table 3,[18] it can be seen that all have insignificant coefficients, except for the variables representing upward communication

[17] Marginal effects are given for all the key variables of interest, that is the cost of job loss and the HRM variables, in Table 4.

[18] The estimated coefficients are very similar between the two specifications for all variables, as would be expected if the imposed restriction is valid.

from employees to management, and the existence of payment by results schemes. If HRM techniques work through higher morale effects, then it would be expected that, of the communication variables, upward communication would have the strongest effect on morale, and hence effort, since workers are likely to feel more valued when they are actually being listened to, rather than simply being talked at. Holding all variables at their mean, the introduction of upward communication to a workplace previously with none, is estimated to increase the probability of the highest effort level being supplied by almost 8 percentage points, from 16.1% to 24.0% Providing information, via two-way or one-way downward communication, does not seem to have any statistically significant motivational impact, although the estimated coefficient on downward communication is almost as large as that on upward communication, its insignificance in a statistical sense being caused by its larger standard error.

Payment by results schemes seem to have the strongly positive effect on effort expected of them; work groups whose pay depends explicitly on their performance have a significantly higher probability of supplying a high average effort level. There is no specific way to compare the effects of the continuous and the dummy variables, but comparing the effects of a one standard deviation increase above the mean for the continuous variables, and the effects of the presence of the various characteristics for the dummy variables, payment by results schemes are found to have the most impact on effort, on this basis, of all the variables in Table 3. The probability of effort being at its highest value increases by almost 12 percentage points, on average, from 18.0% to 29.7%, if a payment by results scheme is present, relevant to a situation where one is absent. The motivational impact of such schemes should be clear, and would be predicted by most psychological theories of motivation. For example, expectancy theory[19] states that any scheme which links a desirable outcome, such as pay, explicitly to good performance, will increase the motivation to work hard. The fact that the results are consistent with such theory can be used to justify the use of the dependent variable, as described in Section 3: payment by results schemes having their predicted effect suggests that the dependent variable is accurately measuring effort.[20]

Turning to the variables representing organisational change and their effect on effort, previous study of this topic is included in the survey by Gallie and White (1993). They devote a lot of attention to describing the increase in skills necessary to work in modern workplaces, and comment on how 'the massive increase in the use of new technology...is closely related to increased skill requirements' (p.28). Given that they also find a strong positive link between skill requirements and effort[21], their analysis suggests that establishments that have introduced such new technology should, other things being equal, have higher effort levels, on average. On the other hand, an argument could be advanced as to how the introduction of new technology could reduce the effort necessary to perform jobs. If a worker previously did a task manually, and then is given a simple machine to operate that can perform the same task, then this is likely to reduce the physical effort required to perform the task, without increasing mental effort to any great

[19] For a review of this theory, see Lawler (1971).

[20] It should perhaps be noted that payment by results schemes could actually reduce effort, for example if employees are working to a quota system, and do not want to work too hard for fear of the quotas being raised. Roy (1952) describes just such a situation of 'goldbricking.' While accepting that such behaviour can occur, I would still argue that the results provide justification that the dependent variable is measuring effort, since a positive effect on effort is consistent with theory, as described, and would probably be the impact most people would predict should be observed.

[21] Of those respondents in their survey who had experienced in increase in necessary skill levels, 76% also reported an increase in effort, whereas this figure was only 35% among respondents whose skill levels had remained unchanged.

extent, once the operation of the machine has been mastered. Similarly, the introduction of some technology can be deskilling, rather than upskilling. For example, new assembly lines may reduce an individual's tasks to loading the inputs, unloading the output, and observing whether the machinery is running smoothly, whereas previously a skilled craftsman may have been required to produce the good.

The results in Table 3 are consistent with Gallie and White's (1993) hypotheses; the introduction of new general plant or machinery is associated with significantly higher effort. The probability of effort being reported at the highest level is 11.5 percentage points (30.1% compared to 18.6%) higher in establishments that have introduced new machinery, relative to those that have not, holding other variables at their mean values. Whether the Gallie and White mechanism, whereby new machinery increases necessary skills which in turn increases necessary effort, is at work is unclear, however. The result could alternatively be a reflection of the argument that new machinery, and in particular production lines, are introduced in attempt by management to control the pace of work. It would be useful if the data supplied information concerning who was using the new machinery, so revealing whether the process involved unskilled workers being upskilled, and thus casting some light on whether Gallie and White's arguments are correct. This is not the case, however. All the data say is in which workplaces new technology was introduced, and since the majority of workplaces have both unskilled and skilled manual workers, it cannot be determined whether one particular group or another is using the new machines. Either way, however, it seems clear that the impact of this variable works through a different mechanism to the variables analysed above, in that the new technology has increased the effort *necessary* to do a lot of jobs, either through upskilling or a more regulated work pace. Thus, it is likely that the higher effort reported by worker representatives in 'new machinery' workplaces is an aspect of the job that has been forced upon them, whereas the variables analysed above work through a motivation effect, whereby employees individually *choose* higher effort levels.[22]

The introduction of specifically microelectronic technology or new work practices appear to have no statistically significant impact on effort. Further, in establishments where job flexibility has been raised, worker representatives actually report significantly lower effort on average. The introduction of increased flexibility lowers the probability that the highest level of effort will be reported from 25.9% to 17.1%, holding other variables at their mean values. Possible reasons for this can only be speculated on. Perhaps flexible work procedures are still in their infancy, and the full beneficial effects for employee effort have yet to be exploited. For example, at present, maybe the practice of employees doing different jobs leaves spare time between jobs, thus reducing average effort, while establishments adjust to the new forms of work organisation.

Union density, whether added linearly or quadratically, fails to attract a significant coefficient.[23] A possible reason for this result is the cumulative effects of the interactions between union power and the other variables of interest in the study, as described in Section 2. As union power rises, it has been hypothesised that the effect of the cost of job loss variables on effort will be diminished, and the effect of the HRM variables may be increased. It could be that not controlling for these interactions will lead to biases in the union coefficient which offset each other, thus explaining the insignificant effect. Indeed,

[22] See, however, footnote 4, which acknowledges the possibility that the HRM variables are also management tools for 'squeezing' more effort from their employees, and may not, in fact, work through employees optimally *choosing* more effort.

[23] The t-statistics on the quadratic terms are higher than the t-statistic when union density is entered linearly, and it is the former coefficients that are reported in Table 3.

when interaction terms are included in the analysis, the union density term becomes significant.

The proportion of supervisors and the workplace size variables were included to control for the effects of monitoring on effort. The shirking version of the efficiency wage theory predicts that workers are less likely to shirk as the degree of supervision increases, since they are more likely to be observed and suffer the costs of job loss, should they do so. The coefficients on all of these variables are insignificantly different from zero, however. Of interest is the negative coefficient on the variable measuring the proportion of employees classified as supervisors, which, although insignificant here, is significant in the subsample of establishments with union density less than 90%, as revealed in the next section. This surprising negative result suggests that workgroups who are tightly supervised supply *lower* effort on average, and deserves comment. Previous analyses of self-reported effort have provided conflicting evidence as to the impact of monitoring. Belman *et al* (1992) find that the availability of a monitoring technology reduces shirking, but a higher of degree of supervision actually employed leads to more shirking. Drago and Heywood (1991) find that the higher the supervision, as measured by respondents saying that their supervisor can insist that they work hard, the higher is effort, while Drago (1990) finds the reverse; that the higher the degree of monitoring, measured on a 0-4 scale, the lower is effort.

The effects of supervision on effort are therefore inconsistent across studies, and the negative impact found here has been found by other researchers. It would seem that supervision may have another, negative, effect on effort, as opposed to the positive, motivating one hypothesised above. Drago and Perlman (1989) discuss just such a possibility. Their argument is that 'supervision may decrease the possibilities for shirking, but enhance employee awareness of shirking and stimulate employee interest in it' (p.45). They back up their claim that supervision can reduce effort by referring to psychological theories such as cognitive dissonance theory, which predicts that individuals will react to the dissonance between their self-belief of their own worth and the signal which they are receiving that they are untrustworthy, by shirking in an attempt to prove supervisory incompetence and so discredit the signal.

Alternatively, efficiency wage models could be adapted to explain the negative effect of supervision on effort. For example, in Akerlof's (1982) Gift Exchange model, an increase in supervision may be viewed by the workgroup as a reduction in the 'gifts' on offer from the firm, which will therefore be reciprocated with a reduction in their effort. Although the shirking model typically predicts a positive effect of supervision on effort, an alternative interpretation is that more supervision is an unwelcome job attribute which reduces the value of that job to the workers, thus reducing the cost of job loss, which in turn increases the likelihood of their shirking.

In general, therefore, supervision can increase effort by its effect on the probability of being caught shirking and subsequently punished, but it can also have a negative effect whereby the relationship between management and workers is soured, with the result that the latter are less inclined and motivated to work hard for the former. It seems as though this negative effect is dominating in the WIRS sample.

An alternative way of explaining the negative supervision result is that the proportion of supervisors is an endogenous variable, and the negative coefficient is reflecting the reverse causality to the one hypothesised. Thus, it could be the case that a workplace hires a larger proportion of supervisors, precisely because effort is low. Because of this potential endogeneity problem, column 3 of Table 3 presents the results for a specification which excludes the supervision variable. It can be seen that there is very little effect on the remaining coefficients, suggesting that any simultaneity bias in the first two columns is minimal. All estimations that follow therefore maintain the supervision variable in their specifications.

Of the remaining control variables, the proportion unskilled and the ethnic minority variables both attract significantly positive coefficients. Employing more than 5% of the workforce from ethnic minorities has considerable impact on the probability of the highest effort level being reported, which rises to 28.9%, from 17.9% in establishments where individuals from ethnic minorities comprise less than 5% of the workforce. The same probability rises by over 5 percentage points, from 20.1% to 25.5%, if the proportion of unskilled workers is increased by one standard deviation above its mean, holding all other variables at their mean values. On the assumption that both of these variables represent individuals at some disadvantage in the labour market who may find it difficult to find another job, these results can be rationalised by efficiency wage theory, since the cost of job loss for such workers will be higher and so they are more likely to supply high effort to avoid losing their jobs. On the other hand, an alternative interpretation is that such workers are typically required to expend more physical effort, because of the nature of the jobs that they perform.

b.) Splitting the Sample into High- and Low-Density Workplaces

The next step of the analysis is to run separate ordered probit regressions for high- and low-density workplaces. The significance of particular variables in the two equations would then show immediately which variables strongly influence effort when union power is high, and which when unions have less power, thus providing evidence for some of the hypotheses advanced in Section 2.

Clearly, it had to be decided at what density level to split the sample. Experimentation revealed that the strongest results were obtained when the split was made at 90% density, in the sense that there was the clearest difference in the effects of the explanatory variables between the two sectors when defined by this cut-off point. The data therefore suggests that 90% density is a crucial level of union strength, beyond which the variables under consideration can have a very different motivational impact. Table 5 presents the results of the two ordered probit effort equations when the sample is split at this density level. [24]

The cost of job loss variable was included in both sub-equations, rather than its constituent parts, earnings and unemployment, because likelihood-ratio tests fail to reject the restrictions implicit in the use of this variable, in either equation. A high cost of job loss has a much larger impact on effort in the low-density sector, where the threat of being dismissed for shirking is real, compared to the high-density sector, where powerful unions with discipline procedures can insulate employees to some extent from this threat, so that it is not so necessary for such workers to supply high levels of effort to keep their jobs. Indeed, there is no evidence of any significant effect on effort at all when union power is high, the point estimate suggesting an increase of only 0.4 percentage points in the probability of the highest effort level being supplied, following a one standard deviation increase in the cost of job loss above its mean value. In the low-density sector, on the

[24] Other experiments cut the sample at different union density levels from 85% to 95%, or on the basis of the existence of a union membership agreement (a closed shop or management recommending membership). While the cost of job loss results to be described remained robust across all of these splits, the HRM variables did not reveal such a clear pattern across the various subsamples as they do at the 90% density split. This would suggest that it is indeed a density level of 90% that is crucial for obtaining the benefits of a strong union for HRM policies, and splitting the sample at any other density level distorts the results somewhat, or that the HRM results are not particularly robust.

other hand, the cost of job loss coefficient is significantly positive, at the 1% level, and suggests that a one standard deviation increase in the cost of job loss above its mean will increase the probability of the highest effort level being supplied by almost 14 percentage points, from 17.5% to 31.4%, holding other variables constant at their mean values.[25] Comparing one standard deviation changes in the continuous variables with the simple presence of the dummy variables, as before, this cost of job loss effect is the third largest amongst all the variables in the low-density sector.

Although these numbers are very suggestive, a formal test of the proposition that the cost of job loss will have a stronger effect on effort in the low-density sector than in the high-density sector was performed. A test of the difference between the estimated coefficients in the two subsamples was performed, with the result that they are significantly different from each other, at the 5% level (t=2.45).[26] Therefore, the hypothesis put forward in Section 2 concerning the influence of unions on cost of job loss effects would seem to be validated by the data. Indeed, an even stronger result is suggested than was originally proposed: not only is the cost of job loss effect on effort stronger in the low-density than in the high-density sector, but there is no evidence of any effect at all when unions are strong. It would therefore appear that the union wage does not act as an efficiency wage.

These results add to a literature that debates the relevance of the shirking model to the primary or secondary sector. This literature was described in detail in the literature review to this thesis. To summarise, the available empirical evidence, such as Oster

---

[25] Marginal effects are given for all the key variables of interest, for each subsample, in Table 6.

[26] The test is a simple t-test on the difference between the coefficients, with the standard error of this difference being the square root of the sum of the variances of the two estimated coefficients, on the assumption that the covariance between them is zero.

(1980), Rebitzer (1987) and Green and Weisskopf (1990), suggest that the shirking model is more relevant to the secondary sector of the economy than the primary sector, contrary to the theoretical predictions of Bulow and Summers (1986). The evidence is that the 'unemployment effect' on effort, however measured in their studies, is negatively correlated with such indicators of primary sector status as high unionisation, long-term employment relations, a high degree of specific skill-usage and high concentration ratios. The evidence provided by WIRS is consistent with these other studies, in that the full cost of job loss effect, rather than just the unemployment effect, is weaker in highly unionised, presumably primary sector, workplaces.

Turning now to the HRM variables, joint consultative committees, upward communication[27] and payment by results schemes are all associated with significantly higher effort when accompanied by a powerful union. The effects are all large, the probability of the highest effort level being reported rising by 8, 9 and 19 percentage points respectively, compared to a similar average workplace without such schemes. The last of these effects, that of payment by results schemes, is the largest of all the effects in the high-density equation. In the low-density sample, joint consultative committees attract the only significant coefficient among the HRM variables, and this is negative, so that the probability of the highest level of effort being reported is actually over 12 percentage points lower when such committees are present, compared the average low-density establishment without such committees.

Comparing the two subsamples, employee involvement, JCCs, upward communication and payment by results schemes all attract a more positive coefficient in

[27] Which was previously revealed to be the most relevant form of communication for eliciting high effort responses.

the high- rather than in the low-density sector. Only one of these differences, that of JCCs, is statistically significant at conventional levels, although the differences for employee involvement and payment by results schemes also attract t-statistics of around 1.4. The variables which are associated with higher effort in the low- rather than the high-density sector are two-way and one-way downward communication, with the difference being statistically insignificant in both cases. Table 1 shows that these two forms of communication are highly prevalent among all establishments, and it would seem that it is upward communication from employees to management which is the principal new innovation of HRM, as far as communication is concerned. Thus all of the *key* HRM variables seem to have more effect on effort when accompanied by a strong union, although it should be stressed again that only one of these differences between the high- and low-density sectors is statistically significant at conventional levels.

The fact that a joint consultative committee is associated with higher employee effort only in workplaces with strong unions may be explained by employees having more faith in, and responding more positively to, such committees, when staffed by worker representatives backed by a strong union or, more likely, officials from that union participating themselves.

As for payment by results schemes, the usefulness of strong unions for their success in motivating effort may be due to employees accepting the fairness of such schemes, in the knowledge that grievance procedures of the strong union would be invoked if management abused the system and did not provide their employees with the rewards they deserved. A point against this interpretation is that the variable is concerned with payments for actual produced results, making it difficult for such abuse or unfair treatment to occur. Had these results been for merit pay, which is based on supervisors' subjective appraisals and is therefore more open to abuse, then the above interpretation may have been more relevant. However, even with pure payment for output schemes, there is still room for management to behave unfairly, for example by deliberately allocating, or threatening to allocate, certain employees to particular tasks that are less productive than others, or by not setting fair rates for tasks, particularly when a new task has been allocated and insufficient allowance is made for the learning of that task. If a strong union is more capable of preventing such unfairness, then this could be a possible reason why employees respond more positively to payment by results schemes in high union density workplaces.

The benefits of strong unions for upward communication schemes and employee involvement programmes are likely to be related to the 'voice' role of unions, which describes the usefulness of strong unions in transmitting information from employees to management. If employees believe that their message is actually getting through and being listened to, this is likely to increase the success of such schemes in increasing organisational commitment amongst the workforce.

The fact that employee involvement schemes and joint consultative committees are associated with lower effort in the absence of a strong union may be somewhat of a surprise to advocates of such schemes, but the finding is not without precedent, as described in the literature review in Section 2. Kelley and Harrison (1992) offer potential explanations for the similar result which they obtained. First they suggest that there may be a reverse causality affect, whereby workplaces with low productivity are more likely to initiate employee involvement schemes in an attempt to rectify the problem. They then dismiss this as a potential explanation, however, on the grounds that approximately 70% of the firms in their sample have such schemes, and it seems unreasonable to assume that they had all been panicked by poor results into establishing employee involvement schemes. Similarly in the WIRS sample being used, over half of the establishments, 57%, had introduced participation in the three years prior to the survey, and by the same argument, it is unlikely that they were all motivated by poor performance.

A second suggestion advanced by Kelley and Harrison is that employee involvement is still a relatively new concept, and may still be displaying 'teething problems.' This argument cannot be fully assessed until more time has passed and new data have emerged. The final possible explanation, which the authors seem to prefer, is that employee involvement schemes are typically found in large organisations in which, rather than simplifying the bureaucratic structure, they add another complication, increasing the decision-making time and reducing the time spent by the workforce in actual production. Given that the dependent variable in the Kelley and Harrison study is a measure of the time taken to perform production tasks, this explanation ties in with their empirical results. With respect to the current study, however, it does not explain why employee involvement schemes should lead to workers reducing their effort. The only *demotivating* aspect of HRM mentioned in the hypotheses section was that if HRM is successful in raising firm performance through improved production techniques, to such an extent that the firm does not need as large a workforce as previously, then any lay-offs undertaken may have a demoralising effect on the remaining employees, who may then reduce their commitment to the firm, and hence their effort. Unfortunately, the current data set makes it very difficult to identify the mechanisms behind the results obtained.

There remain some differences in the estimated coefficients between the two subsamples which deserve comment. The negative impact of the proportion of supervisors on effort that was found for the sample as a whole, is here found to exist significantly in the high union density sample, the low-density coefficient remaining insignificantly different from zero. The effect is not particularly large in either sample, however, in comparison to the effects of the other variables under consideration. A one standard deviation increase in the proportion of supervisors above its mean value reduces the probability of the highest effort level being reported from 17.5% to 15.4% in the low-density sample, and from 17.7% to 14.2% in the high-density sample, holding all other variables at their mean values.

It was argued above that supervision can have two, offsetting effects on effort, with more intensive supervision increasing the risks of shirking, but also souring worker-management relations and reducing employees' commitment to the firm. The negative coefficient on the degree of supervision was then rationalised by the second effect dominating. It has also been argued earlier, when discussing the cost of job loss results, that employees can be protected against a dismissal threat by a strong union. In this case, extra supervision may not reduce shirking if workers do not fear dismissal anyway, but will still have the negative, demotivating effect on morale and hence effort. When unions are not strong, however, the dismissal threat is real, and so the positive influence of supervision on effort should offset, at least to some extent, the negative, demotivating effect. Thus the coefficient on the supervision variable would be expected to be more negative in the high union density sample, as observed. It should be noted, however, that the coefficients in the two subsamples are not significantly different from each other, because of the large standard errors on this variable.

The existence of good relations between managers and employees is associated with significantly higher effort in the low-density sector, while the equivalent coefficient in the high-density sample is negative and insignificant. It therefore appears that maintaining good industrial relations is more likely to be associated with higher effort when union power is low, the difference between the estimated coefficients being statistically significant at the 10% level (t=1.87).

Two contrasting interpretations can be put on these results, although the available data cannot discriminate between them. One is that the presence of unions can harm the transformation of good feeling and morale within the company into greater effort. Thus, even if managers foster good relations with their employees, the adversarial style of unionism may convince employees that they could still be better treated, and that the firm is not deserving of their higher effort. A similar argument is that restrictive practises set by the union may make workers unwilling to increase their effort for fear of angering the union and appearing to behave as the managers' 'poodle.'[28] A different interpretation is that the managers who responded to the survey have different expectations of employee behaviour, depending on whether they operate in a union setting or not, so the answer that they give is endogenously determined. Thus, in a non-union workplace, managers may report good relations with their workforce only when the individuals in question totally acquiesce to all demands made of them, which could explain the positive correlation between the existence of good relations and effort observed in non-union workplaces. A union can offer its members protection against exploitation by managers, however, and hence managers may not expect a total lack of resistance to any demand from their workforce, but may still consider relations to be good, thus removing the link between effort and reported relations.

---

[28] Given that the variable used represents the opinion of management as to the state of employee relations, these arguments assume that the respondent in question differentiated between relations with the union, and relations with individual workers, so that he or she would still report good employee relations, even when faced with an adversarial union of the type described.

Examining the influence of introducing new machinery, the positive impact on effort observed in the full sample is replicated in the two subsamples, but the effect is larger when union power is low, and indeed is insignificantly different from zero in the high-density sample. However, the effect of new machinery seems to be economically significant in both subsamples, and in fact the two coefficients are not significantly different from one another (t=0.82). The introduction of new machinery is associated with a 18 percentage point rise in the probability of the highest effort level being reported in the low-density sample (the second largest of all effects in this subsample, under the rules of comparison described above), and a 7 percentage point rise in the high-density sample, holding all other variables at their mean values.

It was argued above that the introduction of new machinery imposes greater skill demands on workers, which in turn tends to increase their effort, or that it is a management technique for controlling, and increasing, the pace of work. Thus, this variable has its impact through forcing workers to supply more effort, rather than influencing their effort choices. It seems reasonable to assume that more powerful unions would try to limit this forced increase in effort, perhaps through negotiations with management over how the organisational change was to be introduced. Thus, a smaller increase in effort following the introduction of new machinery could be expected in the presence of strong unions, the estimated coefficient in fact suggesting that there is no statistically significant increase at all. The coefficients on the other organisational change variables, new work practices and new microelectronic technology, show a similar pattern, increasing effort when union power is low, and actually reducing effort when union power is high, although none of these coefficients are significant and so less attention should be given to them.

With respect to the workplace size dummy variables, their coefficients are consistently much larger in the low-density sample, two almost achieving significance in that equation, although the differences between the two subsamples are not statistically significant, due mainly to the high standard errors on these variables. The original reason for including workplace size was to control for ease of monitoring, on the assumption that monitoring is more difficult in large workplaces, which should therefore witness an increased incidence of shirking. The finding of higher effort in large workplaces is not consistent with this hypothesis. It may be that higher levels of mechanisation and increased regimentalisation in large workplaces impose extra effort demands on workers, with this effect being reduced, or removed altogether, by a strong, protective union. While this argument is only conjecture, it is at least consistent with the 'new machinery' story told above.

The remaining variables in the estimated equations are those that control for the composition of the workforce at the various establishments. The 'proportion male' variable attracts a negative and insignificant coefficient when union power is low, as was found in the full sample, but is associated with significantly higher effort in the high union density sample. Furthermore, the difference between the two estimated coefficients is significant at the 5% level (t=2.07). The variable representing establishments with more than 5% of their workforce belonging to ethnic minorities attracts a significantly positive coefficient in the low-density sector, but a smaller, and statistically insignificant, coefficient when union density is high. This effect is the largest amongst all the variables in the low-density sample, with a probability of effort being reported at the highest level that is almost 20 percentage points higher (32.4% compared to 12.9%) if ethnic minorities are represented by more than 5% of the workforce, holding all other variables at their mean values. It could be tentatively suggested that unions moderate the effort demands made of non-whites, although the effect remains large, if statistically insignificant, even in the high-density sector, where the same probability is 11 percentage points higher when ethnic minorities constitute more than 5% of the workforce. The two coefficients are not significantly different from each other (t=0.75). Finally, the proportion unskilled variable is one of the few in Table 5 to attract a similar-sized coefficient in the two sectors, the coefficients being significantly positive at the 10% level in both cases, while the proportion full-time coefficients are statistically insignificant in both sectors, as is the difference between them.

c.) Are Earnings Endogenous?

One of the key results of the previous section was that a high union density reduces the motivational impact of earnings on effort; that is, the union wage does not seem to act as an efficiency wage. However, it may be that earnings are endogenous and the coefficient is biased, invalidating this result. Two possible forms of endogeneity can be identified. First, it may be that earnings and effort are being simultaneously determined by another variable which is not controlled for in the analysis. For example, if demand for the firm's product is rising, then we may expect managers to entice more effort from their existing employees, as well as offering higher wages to attract more, or better skilled, workers. Not controlling for this effect could then explain the observed positive correlation between earnings and effort.

Table 7 therefore includes, for the full sample and the split samples, an indicator of workplaces which report an expanding market for their product as an additional explanatory variable.[29] Comparing the results with those in Tables 3 and 5, it can be seen that the cost of job loss variable retains its significant coefficient in the full sample and the low-density sample, even when changes in product demand are controlled for. It therefore does not seem that variations in product market conditions across establishments can explain the positive correlation between earnings and effort found earlier. Furthermore, the coefficient on the expanding market variable is statistically insignificant in each column, suggesting that management cannot extract more effort when product demand is rising anyway. As a final check, the equations of Table 7 were re-estimated using a variable representing respondents who considered that their sales were rising as an indicator of demand conditions, rather than the expanding market variable. Once again, the cost of job loss variable attracted a significant coefficient in the low-density sample, and an insignificant one in the high-density sample, while the new sales variable attracted very small and statistically insignificant coefficients in every equation.[30]

Another potential form of endogeneity bias on the earnings coefficient is that earnings may be determined by effort, rather than the mechanism hypothesised above of higher earnings motivating individuals to supply more effort. The reverse causality argument would be implied, for example, by the theory of compensating wage differentials. On the assumption that effort is an unwelcome job attribute to most individuals, higher required effort levels would lead to demands for higher wages in

[29] This variable was taken from the Employers' Manpower and Skills Practices Survey (EMSPS), which surveyed the same workplaces as found in WIRS, a year later in 1991, and included a question asking the respondents whether the market for their product had expanded, contracted or remained stable over the previous three years, thus referring to a period which included the original 1990 WIRS. This information was matched into the data set by means of the establishment identification number, and a dummy variable was created to identify workplaces for whom the product market had expanded.

[30] The 'expanding market' results were reported rather than the 'rising sales' results, because use of the latter variable reduced the usable sample sizes by more than the former, and also led to a more uneven split between low-density and high-density establishments.

compensation, thus providing an alternative explanation for the observed positive correlation between earnings and effort. However, it could be argued that unions would be more successful than individual workers with little bargaining power in obtaining differential payments, particularly for such a potentially unverifiable variable as effort. In this case we would expect the correlation between earnings and effort to become stronger as union power increases. The results, however, reveal exactly the opposite. The results for the split sample are therefore consistent with efficiency wage theory, but are much more difficult to rationalise in terms of compensating differential theory. It is therefore suggested that efficiency wage payments are the cause of the positive correlation between earnings and effort that has been observed.

As a final control against possible endogeneity of earnings, Table 8 re-estimates the effort equation for the full and split samples, with earnings excluded. Comparing these results with the relevant columns of Tables 3 and 5, it can be seen that all estimated coefficients are very similar to their previous values, and their significance is unaltered in almost all cases. The exceptions are the good employee relations and the proportion unskilled variables in the low union density equation, where only slight changes in the estimated coefficients takes them from being barely significant before to insignificant now. It therefore does not appear that the results presented earlier were subject to any simultaneity bias as a result of endogeneity of earnings.

Table 8 also shows that the unemployment variable does not attract a significant coefficient when entered on its own, rather than as part of the cost of job loss variable, in any of the samples. Given that the cost of job loss was earlier shown to be significantly associated with effort in the full sample and the low-density sample, it would appear that it was the earnings component, rather than the unemployment component, that was driving this result. It would therefore seem that individuals are more affected by the wage they receive than the outside unemployment rate, when responding with effort to the cost of job loss. A similar result was obtained with the individual-level BSAS data in the last chapter.

There remains one way in which earnings may be endogenously determined that has not been controlled for in the analysis. The problem is that the results as presented make no allowance for the possibility of wages and effort being jointly bargained over by the managers and the union. Standard effort bargaining models, as reviewed in Andrews and Simmons (1995), predict that the more powerful a union becomes, the higher will be the negotiated wage and the lower the negotiated level of effort. Thus across a sample of unionised establishments with varying degrees of unionisation, a negative relationship would be expected between wages and effort, which may offset any positive relationship between the two variables due to compensating differential payments. If the extent of effort bargaining differs between the two subsamples, then the observed results could be explained by such a compensating differentials and effort bargaining story. For example, if establishments in the high-density sector are more likely to conduct effort bargaining, then the positive relationship between wages and effort will be offset by more in that sector than in the low-density sector, where compensating differentials will remain the principal source of a relationship between wages and effort. This would then be a consistent rationale for a more positive wage, or cost of job loss, effect in the low-density than in the high-density, sector, without efficiency wage payments being mentioned.

However, that effort bargaining should imply a higher wage and lower effort as union power rises is by no means certain. Andrews and Simmons (1995) show that a sequential bargaining model, where effort is bargained over first, followed by wages, will lead to a positive relationship between wages and effort across establishments with different relative bargaining strengths. Such a model may be more appropriate, since the 1980s saw a period of both rising wages and rising productivity, the latter usually attributed to more effort, as union power fell. [31] The reason most often advanced is so-called 'concession bargaining,' whereby as union power fell, firms were able to obtain more effort from their union workforces, but for which they had to pay higher wages in the second stage of the bargain. In other words, unions conceded the higher effort as long as they were suitably recompensed. If, as argued in the previous paragraph, high- density firms are more likely to engage in effort bargaining, so that high-density unions are more likely to get to the 'concessionary bargaining table,' then this would lead to the prediction of a stronger positive association between wages and effort in the high-density sector, a prediction not validated by the facts. What this discussion reveals most is that it may require a more structural approach to modelling the effort equation, if the results are to be used to successfully distinguish efficiency wage theory from effort bargaining.

Had the results in Table 8, when the unemployment rate was entered on its own, been stronger, then this may have offered some indications as to the relevance of efficiency wage theory, given that the local unemployment rate should be much less susceptible to the problem of endogeneity than earnings. A more positive unemployment coefficient in the low-density sector would be consistent with the hypothesis that efficiency wage considerations are relevant, except where dismissal threats are not credible, as in the high-density sector. Looking at the results in Table 8, it could be argued that the unemployment coefficient is over four and a half times larger in the low-

[31] It should be noted, however, that this fact is considering the effect of changes in union power over time, and therefore may not be relevant to a model containing only cross-sectional variation in union power.

density equation than in the high-density equation. However, neither of the coefficients, nor the difference between them, is statistically significant. This lack of significance need not necessarily be taken as evidence against the efficiency wage hypothesis, though, since evidence in this chapter and the preceding one has shown that the earnings effect usually dominates the unemployment effect. A possible reason for this is that the unemployment rate is not an accurate indicator of the individual's personal opinion as to his or her re-employment probability, which is the more theoretically relevant variable. The fact that there remains such a large difference in the relative size of the threat variable coefficient in the two equations, even after the dominant effect of earnings is taken out, may suggest, though of course not statistically prove, that the hypothesis advanced in Section 2 is relevant.

d.) Interacting Union Density with the Variables of Interest

Splitting the sample according to whether union density at each establishment is above or below 90% is an informative, but very particular, way of investigating the interaction between union strength and the relationships between the variables of interest and effort. A more general investigation is to create interaction terms between these variables and effort. Such an analysis considers the whole continuum of union density, rather than a single, arbitrary cut-off point. Table 9 therefore presents the results for when union density is interacted with the cost of job loss and HRM variables.

Looking at the cost of job loss results in column 2, the results described above would appear to be robust to allowing the most general interaction with union power. Thus when union density is zero, the coefficient on the basic cost of job loss variable shows that there is a significantly positive association with effort. As union density rises, however, the significantly negative interaction term reveals that this cost of loss effect is reduced. In the extreme, when union density is 100%, the cost of job loss effect is almost totally removed, the basic and interaction coefficients being similar in absolute size.

Using the cost of job loss variable now imposes two restrictions, the extra one being caused by the presence of the interaction term. The likelihood ratio test reveals that these restrictions are rejected at the 10% level. If this is taken as sufficient evidence against the restrictions, then column 1 enters the log of earnings and unemployment separately, together with their interactions with union density. The results reveal the same point that was made in the previous subsection; the fact that the significant coefficients are on the earnings variables rather than the unemployment variables shows that it is the former which influence individuals' effort choices more. The conclusion regarding the interaction of earnings with union density is the same as that of the combined cost of job loss variable; rising union power reduces the motivating impact of earnings, and totally removes any effect in the limit when union density is 100%. As before, it seems that the union wage does not act as an efficiency wage.

Turning to the HRM variables, the interactions between union density and employee involvement, JCCs, upward communication and payment by results schemes are all positive, suggesting that such schemes have more impact as the union becomes stronger. These are exactly the same four HRM variables which had a stronger positive association with effort in the high-density than in the low-density sector in the previous analysis. In fact, in the absence of a union, the basic coefficients reveal that all four schemes are negatively associated with effort. However, all of these basic coefficients are insignificantly different from zero, and only one of the four positive interactions is statistically significant in column 2; that on the variable indicating the existence of payment by results schemes. The two HRM variables to be more positively associated with effort in the low-density sector previously, namely two-way communication and downward communication, both have positive basic coefficients and negative interaction terms, consistent with the earlier results. The coefficients on the two-way communication variable are statistically significant at the 5% level. Therefore, the pattern of results obtained with the split panel analysis is robust to the use of interaction terms, although even fewer variables obtain statistically significant coefficients. This lack of significance may be the result of multicollinearity between all of the interaction terms, leading to high standard errors.

Finally, among the controls, the same variables that achieved statistical significance in the standard specification of Table 3 do so again when the interaction terms are included.[32] In addition union density now attracts a large, significantly positive coefficient in column 1. It would seem that once all of the alternatively signed interactions with union power have been controlled for, there is a residual positive relationship between union density and effort. However, if the cost of job loss restrictions are imposed, as in column 2, the union density variable once again attracts a statistically insignificant coefficient. It is not clear why there is such a large change in the density coefficient when moving from column 1 to column 2, although the cumulative effect of the obvious relationship between unions and wages, and the imposition of restrictions involving union density and wages, which are invalid at the 10% significance level, is a likely cause.

[32] These are new machinery (+), increased flexibility (-), the proportion of manual employees who are unskilled (+), and the indicator of establishments where more than 5% of the employees belong to ethnic minorities (+).

e.) Formal Written Documents and the Union Effect

It has been noted that the motivating influence of earnings, or cost of job loss, is reduced as union power rises, with the proposed reason being that employees protected by a strong union do not fear dismissal, so that they do not consider that they are putting their high earnings at risk by shirking. This subsection aims to investigate this effect further.

It is hypothesised that employees will feel most protected, and so be more likely to shirk, when there are formal written documents covering dismissals, negotiated by the union and management. Rather than interacting earnings and unemployment with the measure of union power, Table 10 therefore presents the results when the same variables are interacted with the presence of a formal written dismissal document, on the assumption that it is this, rather than simply a large membership, which makes employees feel secure. In column 1, the new interaction terms are added to the standard specification of Table 3. Looking at the coefficients involving cost of job loss, it can be seen that, in the absence of a formal written document covering dismissals, cost of job loss is positively, but insignificantly, associated with employee effort. The interaction term, is exceedingly small, and clearly highly insignificant.[33] Column 2 considers the interaction specification of Table 9, with the difference that cost of job loss is interacted with the existence of a written dismissal document, rather than union density.[34] A similar finding

[33] Comparing the remaining coefficients in Table 10 with the previous results, it can be seen that the other significant relationships discovered above are not affected by the inclusion of the written document interaction term, while two of the firm size dummy variables now additionally attract statistically significant coefficients, in column 2.

[34] The HRM effects should be unrelated to the existence or otherwise of a written document covering dismissals, since they do not rely on dismissal threats to obtain their motivating influence. Column 2 therefore continues to interact the HRM variables with union density. The results for these interaction terms are very similar to those of the original interaction specification in Table 9.

is obtained, with both the cost of job loss variable and its interaction with the existence of a written document for dismissals attracting statistically insignificant coefficients, the latter being tiny.

It would seem, therefore, that the most relevant split for distinguishing the cost of job loss effect is the 90% density split, rather than the existence or otherwise of a written document for dismissals. An explanation of the results in Table 10, consistent with this conclusion, would be that when no such document exists, there are some workplaces with below 90% and some with above 90% density, with those in the latter group having no cost of job loss effect, so dragging the coefficient downwards, leading to the insignificant result. This brief analysis would therefore seem to suggest that it is simply a high union membership amongst the workplace's employees, rather than the existence of any formal written documents, which provides workers with the security from dismissal which removes the motivating impact of a high cost of job loss. Of course, what is crucial for workers' effort choices is whether they themselves feel secure, regardless of how secure they actually are. A possible explanation for these results is then that rank-and-file members are unaware of written documents negotiated by their union, and that the most observable feature of union power on which to base their decisions is simply the proportion of members at the establishment.

4.5 : Summary

This chapter uses a sample of 375 observations from the 1990 Workplace Industrial Relations Survey to analyse workers' effort choices in a unionised setting. The effort variable adopted is based on a question in the Manual Worker Representative Questionnaire, which asks about the intensity or pace of work among manual workers at the respondent's workplace, relative to similar workplaces. This is the first time that this variable has been studied, and so considerable time is spent justifying its use, and describing its cross-sectional relationship with other variables of interest.

The most convincing evidence for the suitability of this variable is its significant correlation with both managers' and financial managers' opinions regarding productivity at their workplaces. Further support is provided by the significant correlations between the manual and non-manual ratings, and between the variable described above and a similar one concerning intensity of work relative to three years previous.

Analysis of this intensity variable reveals that effort is, on average, higher when any of the human resource management (HRM) techniques under consideration are used, when new plant or machinery has been used, or when new work practices have been introduced. Effort is lower, however, when microelectronic technology is introduced, or when the work regime is made more flexible. With respect to firm size, effort seems to be highest in medium sized firms, while effort is highest in Wales with regard to the standard regions of Great Britain. It should be pointed, out, however, that most of these one-way cross-tabulations do not show significant differences across the various categories of the variables.

Estimating ordered probit equations to allow for the ordinal categorisation of the effort variable reveals the relationship between each explanatory variable and effort in turn, holding the effects of all the other variables constant. A significantly positive relationship is revealed between the cost of job loss, defined to be the log of the product of earnings and unemployment, and effort, which is consistent with both efficiency wage theory and compensating differential theory. Amongst the HRM variables, both upwards communication from employees to management and payment by results schemes are associated with significantly higher effort. Other significant relationships observed are that effort is higher where new plant or machinery have been introduced, lower in establishments that have increased their work flexibility, and higher in establishments with a greater proportion of unskilled and non-white employees.

The analysis then goes on to divide the sample into low union density and high union density establishments, with 90% density being the cut-off point. Very few of the variables under consideration have similar effects in the two sectors. Cost of job loss reveals a very stark difference, with a large, positive, significant coefficient in the low-density sector, and a very small, insignificant coefficient in the high-density sector. The difference between the coefficients is statistically significant, with the coefficients themselves suggesting that a one standard deviation increase in the cost of job loss at the average workplace will increase the probability of the highest level of effort being reported by almost 14 percentage points in the low-density sector, but by only 0.4 percentage points in the high-density sector. It is suggested that this difference is due to the ease with which managers can dismiss their employees. If a strong union presence makes such dismissals more difficult, then employees will not respond to a high cost of job loss, if they feel they are unlikely to be dismissed for shirking. Thus the results for the split sample analysis can be rationalised using efficiency wage theory, but this is more difficult using compensating differential theory. Under the reasonable assumption that strong unions would be more adept at extracting compensating payments from management, the observed correlation between earnings and effort should be stronger in the high-density sector, if compensating differentials are indeed the cause. That this is not the case is one of the few pieces of efficiency-wage evidence to be offered in the literature that can not be explained by competitive compensating differential theory.

The other important results to come from the split sample analysis concern the HRM variables. Employee involvement schemes, joint consultative committees, upward communication and payment by results schemes all are more positively associated with effort in the high-density rather than the low-density sector, the last three all obtaining significantly positive coefficients when union power is high, but not when union power is low (joint consultative committees actually attract a significantly negative coefficient in this case). While this evidence is suggestive, it is not conclusive, given that only one of the differences between the coefficients in the two subsamples is significant at conventional significance levels (joint consultative committees). Also, two HRM variables, downward communication and two-way communication, attract more positive coefficients in the low- than in the high-density equation, although the coefficients and their differences across subsamples are all insignificant. What evidence there is to discriminate between HRM effects across sectors is therefore in favour of a stronger influence in the high-density sector, but this evidence is not decisive.

If HRM policies are indeed more likely to be associated with high effort when unions are strong, the likely reason would seem to be that employees will be more committed, and hence will respond in a more positive way, to schemes which have been negotiated, agreed and maybe partly administered by their union. In addition, the 'voice' aspect of union activities creates well-used channels of communication between officials and management which can increase the effectiveness of participation schemes, as well as guaranteeing to workers that their opinions and ideas are being heard, which should further increase their commitment to the schemes. Finally, employees will be more likely to put forward to top managers new ideas which question existing practices, if they know that a strong union can protect them against any reprisals from immediate supervisors or middle managers who may have been directly or indirectly criticised.

The key results of this chapter are thus the offsetting effects which unions can have on the motivation of their members to supply effort. It has been shown that employees seem to respond with more effort to higher wages or local unemployment, particularly the former, only when *not* under the protection of a strong union which can reduce dismissal threats. However, the study then went on to show that the presence of a strong trade union can have a beneficial influence on employees' effort. In particular, employee involvement schemes, communication schemes, joint consultative committees and payment by results schemes all seem to be more successful in terms of motivation when introduced into a workplace with a strong trade union. While this analysis cannot quantify the offsetting effects, and thus say whether a strong union will have a positive or negative overall effect on effort, it could be suggested that, as team work, flexibility and quality become more important in the modern workplace and new styles of management are introduced, it may well be that unions have an important role to play in this change. Such a role has already been identified as a means by which unions can retain their influence in modern workplaces (Guest, 1995), and indeed, unions themselves have begun to recognise this potential. A 1994 TUC document on HRM says;

'...if the employer is genuinely concerned to improve the performance of the enterprise, is committed to involving the workers in the running of the company and is seeking to develop a real partnership with recognised trade unions then HRM and collective bargaining can work in harmony. Unions are not opposed to HRM strategies which respect their collective bargaining rights and are consistent with the development of high quality, well paid employment.' (TUC, 1994, p.6).

Thus it may be that the introduction of HRM and the continuing existence of trade union representation will prove to be mutually beneficial to both sides of the industrial relation.

Tables

Table 4.1 : Variable Means

Variable	Description	Mean: full sample	Mean: density <90%	Mean: density ≥90%
effort	average effort among manual workers (1-5)	3.640	3.692	3.603
	threat variables			
average earnings	average manual earnings (mid-points of groups taken)	182.4	160.5	198.1
local unemp. rate	unemp. rate in the establishment's local labour market (%)	6.073	5.727	6.320
cost of job loss	average manual wage * local unemployment rate	10.95	8.92	12.39
	HRM variables			
employee involv't	attempts to increase employee involvement in last 3 years	0.568	0.526	0.598
JCC used	existence of a joint consultative committee at establishment	0.637	0.647	0.630
2-way comm.	2 way communication between management and employees	0.803	0.782	0.817
comm. (down)	communication methods from management to employees	0.888	0.878	0.895
comm. (up)	communication methods from employees to management	0.533	0.455	0.589
paid by results	any manual workers paid by results proportion of manual	0.195	0.237	0.164
	control variables			
union density	proportion of manual employees members of a union	0.842	0.648	0.980
new machinery	new plant/machinery introduced for manuals in last 3 years	0.147	0.147	0.146
new work practices	changes in work practices for manual workers in last 3 years	0.587	0.519	0.635
new micro techn.	microelectronic tech. introduced for manuals in last 3 years	0.531	0.442	0.594
good relations	good relations between management and employees	0.403	0.449	0.370
increased flexibility	any changes to increase flexibility in last 3 years employees	0.635	0.635	0.635
proportion male	proportion of manual employees male	0.791	0.723	0.840
prop. unskilled	proportion of manual employees unskilled	0.413	0.505	0.348
prop. full-time	proportion of manual employees work full-time	0.871	0.811	0.914
>5% ethnic	more than 5% of manual workforce from an ethnic minority	0.216	0.288	0.164
prop. supervisors	proportion of total workforce engaged in supervision	0.048	0.046	0.050
50-99 employees	50-99 employees at establishment	0.072	0.071	0.073
100-199 employees	100-199 employees at establishment	0.125	0.135	0.119
200-499 employees	200-499 employees at establishment	0.195	0.167	0.215
500-999 employees	500-999 employees at establishment	0.251	0.224	0.269
1000+ employees	more than 1000 employees at establishment	0.328	0.372	0.297

Table 4.2a : Average Effort Levels for HRM and Organisational Change Variables

HRM Variables	Used or Introduced	Not Used/Introduced
employee involvement	3.66	3.61
joint consultative committee	3.66	3.61
2-way communication	3.65	3.58
1-way communication (down)	3.67	3.45
1-way communication (up)	3.67	3.60
paid by results	3.81	3.60
Organisational Change	Used or Introduced	Not Used/Introduced
plant or machinery	3.84	3.61
work practices	3.66	3.62
microelectronic technology	3.63	3.66
increased flexibility	3.62	3.67

Table 4.2b : Average Effort Across Regions and Establishment Size

Region	Average Effort
Scotland	3.63
North	3.76
North West	3.71
Yorkshire	3.45
West Midlands	3.43
East Midlands	3.72
East Anglia	3.58
South West	3.74
South East	3.72
London	3.59
Wales	3.78
Establishment Size	Average Effort
25-50 employees	3.67
50-99 employees	3.57
100-199 employees	3.62
200-499 employees	3.64
500-999 employees	3.76
1000+ employees	3.59

Table 4.3 : Effort Equations -Full Sample

Variable	column 1	column 2	column 3
threat variables			
log average earnings	0.433 (0.305)	-	-
log local unemp. rate	0.279 (0.190)	-	-
cost of job loss	-	0.323 (0.159)**	0.306 (0.159)*
HRM variables			
employee involvement	0.144 (0.131)	0.143 (0.131)	0.126 (0.130)
jcc used	0.008 (0.142)	0.006 (0.142)	0.006 (0.142)
2-way communication	0.056 (0.160)	0.061 (0.160)	0.045 (0.160)
1-way comm. (down)	0.246 (0.211)	0.239 (0.211)	0.262 (0.210)
1-way comm. (up)	0.287 (0.139)**	0.286 (0.139)**	0.267 (0.138)*
paid by results	0.388 (0.163)**	0.380 (0.162)**	0.380 (0.161)**
control variables			
union density	1.435 (1.234)	1.392 (1.229)	1.440 (1.231)
(union density)^2	-1.190 (0.961)	-1.149 (0.956)	-1.175 (0.956)
new machinery	0.374 (0.191)**	0.372 (0.191)*	0.384 (0.190)**
new work practices	0.024 (0.143)	0.032 (0.142)	0.021 (0.142)
new micro technology	0.161 (0.156)	0.165 (0.155)	0.147 (0.155)
increased flexibility	-0.305 (0.137)**	-0.305 (0.137)**	-0.292 (0.137)**
good employee rel.'s	0.085 (0.125)	0.083 (0.125)	0.089 (0.125)
proportion male	0.148 (0.327)	0.192 (0.310)	0.155 (0.309)
proportion unskilled	0.514 (0.198)**	0.504 (0.197)**	0.505 (0.197)**
proportion full-time	-0.123 (0.414)	-0.044 (0.368)	-0.136 (0.363)
>5% ethnic	0.357 (0.172)**	0.362 (0.171)**	0.372 (0.171)**
proportion supervisors	-2.971 (1.934)	-2.965 (1.935)	-
50-99 employees	0.351 (0.411)	0.336 (0.410)	0.284 (0.408)
100-199 employees	0.292 (0.397)	0.278 (0.396)	0.225 (0.394)
200-499 employees	0.211 (0.386)	0.209 (0.386)	0.131 (0.383)
500-999 employees	0.402 (0.388)	0.396 (0.388)	0.325 (0.385)
1000+ employees	-0.044 (0.387)	-0.045 (0.387)	-0.109 (0.384)
controls	industry (8) region (10)	industry (8) region (10)	industry (8) region (10)
no. of observations	375	375	375
pseudo R^2	0.077	0.077	0.074
log likelihood	-428.101	-428.189	-429.365
LR test	-	0.18	0.17

Notes: 1. Standard errors in parentheses.
2. **=significant at 5% or better. *=significant at 10%
3. Cost of job loss = ln(average wage * local unemployment rate)
4. LR test = $\chi^2(1)$ test of the restrictions implied by using the cost of job loss variables.
5. Estimation is by ordered probit

Table 4.4 : Marginal Effects for the Key Variables in the Full Sample Equation

	Probability that effort=5 if:	
threat variable	**cjl at its mean**	**cjl 1 s. d. above its mean**
cost of job loss	0.201	0.252
HRM variables	**technique not used**	**technique used**
employee involvement	0.179	0.218
jcc used	0.199	0.201
2-way communication	0.187	0.204
1-way communication (down)	0.146	0.208
1-way communication (up)	0.161	0.240
paid by results	0.180	0.297

Note: In each case, all other variables are held at their mean values.

Table 4.5 : Effort Determinants in High and Low Union Density Establishments

Variable	union density < 90%	union density ≥ 90%
threat variables		
cost of job loss	0.926 (0.297)**	0.030 (0.215)
HRM variables		
employee involvement	-0.144 (0.228)	0.266 (0.186)
jcc used	-0.456 (0.251)*	0.315 (0.192)*
2-way communication	0.370 (0.293)	-0.100 (0.219)
1-way communication (down)	0.484 (0.349)	0.177 (0.304)
1-way communication (up)	0.104 (0.234)	0.352 (0.192)*
paid by results	0.111 (0.271)	0.623 (0.235)**
control variables		
new machinery	0.596 (0.325)*	0.251 (0.264)
new work practices	0.146 (0.260)	-0.042 (0.195)
new micro technology	0.163 (0.283)	-0.002 (0.213)
increased flexibility	-0.334 (0.266)	-0.175 (0.185)
good employee relations	0.361 (0.217)*	-0.156 (0.171)
proportion male	-0.301 (0.474)	1.108 (0.488)**
proportion unskilled	0.611 (0.367)*	0.490 (0.262)*
proportion full-time	0.468 (0.592)	-0.311 (0.536)
>5% ethnic	0.674 (0.272)**	0.387 (0.270)
proportion supervisors	-2.318 (3.257)	-4.584 (2.747)*
50-99 employees	0.685 (0.704)	-0.005 (0.569)
100-199 employees	1.006 (0.665)	-0.096 (0.574)
200-499 employees	1.070 (0.654)	-0.087 (0.548)
500-999 employees	1.079 (0.662)	0.158 (0.555)
1000+ employees	0.595 (0.653)	-0.138 (0.560)
controls	industry (7) region (10)	industry (7) region (10)
no. of observations	156	219
pseudo R^2	0.189	0.107
log likelihood	-148.775	-249.142
LR test	3.29	0.45

Notes: See Table 3

LR test in each column is a test of the cost of job loss restrictions, compared to equivalent equations with wages and unemployment entered independently (the results of the latter equations are not reported here).

Table 4.6 : Marginal Effects for the Key Variables in the Split Sample Equations

	Probability that effort=5 if:			
	union density<90%		**union density≥90%**	
threat variable	**cjl at its mean**	**cjl 1 s.d. above mean**	**cjl at its mean**	**cjl 1 s.d. above mean**
cost of job loss	0.175	0.314	0.177	0.181
HRM variables	**technique not used**	**technique used**	**technique not used**	**technique used**
employee involvement	0.195	0.158	0.138	0.206
jcc used	0.261	0.137	0.130	0.209
2-way communication	0.110	0.196	0.200	0.172
1-way comm. (down)	0.087	0.191	0.139	0.182
1-way comm. (up)	0.163	0.190	0.128	0.217
paid by results	0.168	0.198	0.151	0.342

Note: In each case, all other variables are held at their mean values.

Table 4.7 : Controlling for Demand Changes

Variable	full sample	denisty<90%	density≥90%
threat variables			
cost of job loss	0.333 (0.176)*	0.727 (0.344)**	0.176 (0.239)
HRM variables			
employee involvement	0.036 (0.147)	-0.436 (0.282)	0.361 (0.214)*
jcc used	0.138 (0.161)	-0.252 (0.299)	0.405 (0.218)*
2-way communication	-0.186 (0.180)	-0.196 (0.366)	-0.198 (0.244)
1-way comm. (down)	0.194 (0.236)	0.322 (0.436)	0.292 (0.331)
1-way comm. (up)	0.375 (0.154)**	0.205 (0.273)	0.418 (0.212)**
paid by results	0.431 (0.178)**	0.576 (0.322)*	0.477 (0.260)*
control variables			
union density	-0.608 (1.368)	-	-
(union density)^2	0.609 (1.067)	-	-
new machinery	0.399 (0.210)*	0.816 (0.355)**	0.474 (0.303)
new work practices	0.033 (0.158)	-0.022 (0.309)	-0.160 (0.217)
new micro technology	0.102 (0.171)	0.327 (0.336)	-0.028 (0.235)
increased flexibility	-0.326 (0.154)**	-0.235 (0.323)	-0.223 (0.209)
good employee rel.'s	-0.011 (0.140)	0.254 (0.254)	-0.274 (0.191)
proportion male	-0.231 (0.349)	-0.829 (0.549)	1.072 (0.588)*
proportion unskilled	0.669 (0.220)**	0.720 (0.443)	0.740 (0.297)**
proportion full-time	0.124 (0.401)	0.560 (0.693)	-0.212 (0.601)
>5% ethnic	0.381 (0.191)**	0.582 (0.302)*	0.512 (0.313)
proportion supervisors	-4.993 (2.131)**	-4.767 (3.775)	-6.208 (3.060)**
50-99 employees	0.281 (0.489)	1.651 (1.290)	0.014 (0.608)
100-199 employees	0.335 (0.475)	1.731 (1.236)	0.076 (0.599)
200-499 employees	0.197 (0.454)	2.030 (1.190)	0.120 (0.568)
500-999 employees	0.391 (0.456)	1.992 (1.188)	0.313 (0.582)
1000+ employees	0.042 (0.457)	1.490 (1.159)	0.157 (0.592)
expanding prod. mkt	0.172 (0.139)	0.413 (0.276)	0.111 (0.185)
controls	industry (8) region (10)	industry (7) region (10)	industry (7) region (10)
no. of observations	317	130	187
pseudo R^2	0.093	0.220	0.139
log likelihood	-343.404	-114.650	-198.507
LR test	0.05	2.04	1.84

Notes: See Table 3

LR test in each column is a test of the cost of job loss restrictions, compared to equivalent equations with wages and unemployment entered independently (the results of the latter equations are not reported here).

Table 4.8 : Omitting Wages from the Model

Variable	full sample	denisty<90%	density≥90%
threat variables			
log local unemp. rate	0.288 (0.190)	0.540 (0.338)	0.117 (0.263)
HRM variables			
employee involvement	0.127 (0.130)	-0.162 (0.226)	0.272 (0.186)
jcc used	0.002 (0.142)	-0.468 (0.249)*	0.314 (0.192)*
2-way communication	0.066 (0.160)	0.321 (0.290)	-0.095 (0.219)
1-way comm. (down)	0.235 (0.211)	0.378 (0.347)	0.171 (0.303)
1-way comm. (up)	0.284 (0.139)**	0.080 (0.234)	0.345 (0.193)*
paid by results	0.356 (0.161)**	0.080 (0.269)	0.618 (0.235)**
control variables			
union density	1.265 (1.226)	-	-
(union density)^2	-1.012 (0.951)	-	-
new machinery	0.361 (0.190)*	0.601 (0.323)*	0.246 (0.264)
new work practices	0.056 (0.141)	0.187 (0.259)	-0.041 (0.195)
new micro technology	0.166 (0.156)	0.185 (0.282)	-0.001 (0.213)
increased flexibility	-0.298 (0.137)**	-0.404 (0.263)	-0.178 (0.184)
good employee rel.'s	0.075 (0.125)	0.344 (0.216)	-0.153 (0.171)
proportion male	0.301 (0.308)	-0.051 (0.472)	1.116 (0.482)**
proportion unskilled	0.477 (0.197)**	0.562 (0.363)	0.490 (0.261)*
proportion full-time	0.157 (0.363)	0.916 (0.580)	-0.273 (0.525)
>5% ethnic	0.366 (0.171)**	0.616 (0.271)**	0.383 (0.269)
proportion supervisors	-2.853 (1.932)	-1.738 (3.218)	-4.630 (2.746)*
50-99 employees	0.285 (0.409)	0.523 (0.696)	-0.015 (0.569)
100-199 employees	0.217 (0.394)	0.717 (0.652)	-0.102 (0.573)
200-499 employees	0.182 (0.386)	0.939 (0.650)	-0.090 (0.548)
500-999 employees	0.358 (0.387)	1.007 (0.658)	0.153 (0.554)
1000+ employees	-0.061 (0.387)	0.564 (0.649)	-0.146 (0.560)
controls	industry (8) region (10)	industry (7) region (10)	industry (7) region (10)
no. of observations	375	156	219
pseudo R^2	0.075	0.169	0.107
log likelihood	-429.108	-152.475	-249.052

Notes: See Table 3

Table 4.9 : Interacting Union Density with Key Variables of Interest

Variable	column 1	column 2
threat variables		
log average earnings	2.390 (0.777)**	-
density*log av. earnings	-2.362 (0.843)**	-
log local unemployment rate	0.574 (0.547)	-
density*log unemp. rate	-0.243 (0.595)	-
cost of job loss	-	1.100 (0.489)**
density*cost of job loss	-	-0.873 (0.526)*
HRM variables		
employee involvement	-0.558 (0.534)	-0.549 (0.526)
density*EI	0.850 (0.617)	0.844 (0.607)
jcc used	-0.604 (0.643)	-0.663 (0.635)
density*JCC	0.710 (0.727)	0.772 (0.718)
2-way communication	1.427 (0.650)**	1.365 (0.643)**
density*2-way communication	-1.649 (0.753)**	-1.579 (0.745)**
1-way communication (down)	0.877 (0.766)	0.609 (0.750)
density*1-way comm. (down)	-0.547 (0.912)*	-0.300 (0.898)
1-way communication (up)	-0.594 (0.519)	-0.340 (0.499)
density*1-way comm. (up)	1.027 (0.599)*	0.737 (0.578)
paid by results	-0.319 (0.553)	-0.443 (0.550)
density*paid by results	0.898 (0.651)	1.058 (0.646)*
control variables		
union density	11.284 (4.456)**	1.684 (1.420)
new machinery	0.332 (0.193)*	0.351 (0.193)*
new work practices	0.045 (0.145)	0.055 (0.143)
new micro technology	0.107 (0.160)	0.149 (0.158)
increased flexibility	-0.268 (0.141)*	-0.292 (0.140)**
good employee relations	0.082 (0.127)	0.063 (0.126)
proportion male	0.167 (0.331)	0.237 (0.314)
proportion unskilled	0.583 (0.200)**	0.538 (0.198)**
proportion full-time	-0.060 (0.417)	-0.031 (0.373)
>5% ethnic	0.359 (0.176)**	0.387 (0.174)**
proportion supervisors	-3.101 (1.982)	-3.062 (1.973)
50-99 employees	0.514 (0.423)	0.519 (0.419)
100-199 employees	0.401 (0.403)	0.379 (0.399)
200-499 employees	0.261 (0.392)	0.302 (0.389)
500-999 employees	0.502 (0.394)	0.515 (0.391)
1000+ employees	0.065 (0.392)	0.058 (0.390)
controls	industry (8) region (10)	industry (8) region (10)
no. of observations	375	375
pseudo R^2	0.099	0.093
log likelihood	-418.007	-420.671
LR test	-	5.33*

Table 4.10 : Does the Union Cost of Job Loss Effect Work Through the Existence of a Formal Written Document Covering Dismissals?

Variable	column 1	column 2
threat variables		
cost of job loss	0.285 (0.255)	0.317 (0.261)
written doc.*cost of job loss	0.012 (0.198)	0.006 (0.200)
HRM variables		
employee involvement	0.135 (0.131)	-0.591 (0.524)
density*EI	-	0.891 (0.605)
jcc used	0.007 (0.142)	-0.688 (0.631)
density*JCC	-	0.818 (0.715)
2-way communication	0.039 (0.160)	1.156 (0.635)*
density*2-way communication	-	-1.367 (0.735)*
1-way communication (down)	0.229 (0.211)	0.750 (0.747)
density*1-way comm. (down)	-	-0.522 (0.894)
1-way communication (up)	0.292 (0.139)**	-0.478 (0.499)
density*1-way comm. (up)	-	0.907 (0.580)
paid by results	0.394 (0.162)	-0.557 (0.551)
density*paid by results	-	1.206 (0.648)*
control variables		
union density	-0.087 (0.298)	-0.311 (0.874)
new machinery	0.388 (0.192)**	0.357 (0.193)*
new work practices	0.026 (0.142)	0.053 (0.143)
new micro technology	0.157 (0.155)	0.143 (0.158)
increased flexibility	-0.326 (0.138)**	-0.327 (0.140)**
good employee relations	0.088 (0.125)	0.083 (0.127)
proportion male	0.251 (0.310)	0.253 (0.314)
proportion unskilled	0.563 (0.196)**	0.587 (0.198)**
proportion full-time	-0.004 (0.368)	0.061 (0.372)
>5% ethnic	0.380 (0.171)**	0.379 (0.175)**
proportion supervisors	-2.924 (1.936)	-2.889 (1.974)
50-99 employees	0.559 (0.427)	0.793 (0.438)*
100-199 employees	0.533 (0.413)	0.649 (0.418)
200-499 employees	0.465 (0.404)	0.593 (0.409)
500-999 employees	0.646 (0.405)	0.801 (0.410)*
1000+ employees	0.218 (0.404)	0.358 (0.409)
controls	industry (8) region (10)	industry (8) region (10)
no. of observations	374	374
pseudo R^2	0.078	0.095
log likelihood	-426.141	-418.517
LR test	0.25	0.19

Notes: See Table 3
LR test in each column is a test of the cost of job loss restrictions, compared to equivalent equations with wages and unemployment entered independently (the results of the latter equations are not reported here).

References

Akerlof, G.A. (1982) 'Labor Contracts as Partial Gift Exchange,' *Quarterly Journal of Economics*, 97, 543-569.

Allen, S.G. (1988) 'Human Resource Policies and Union-Nonunion Productivity Differences,' NBER Working Paper no. 2744.

Andrews, M. and Simmons, R. (1995) 'Are Effort Bargaining Models Consistent with the Facts? An Assessment of the Early 1980s,' *Economica*, 62, 313-334.

Belman, D., Drago, R. and Wooden, M. (1992) 'Workgroups, Efficiency Wages and Work Effort,' *Journal of Post Keynesian Economics,* 14, 497-521.

Borg, G. (1962) *Physical Performance and Perceived Exertion,* Lund (Sweden), Gleerup.

Bulow, J.I. and Summers, L.H. (1986) 'A theory of Dual Labor Markets with Application to Industrial Policy, Discrimination, and Keynesian Unemployment,' *Journal of Labor Economics*, 4, 376-414.

Cooke, W.N. (1992) 'product Quality Improvement Through Employee Participation: The Effects of Unionization and Joint Union-Management Administration,' *Industrial and Labor Relations Review*, 46, 119-134.

Cooke, W.N. (1994) 'Employee Participation Programs, Group-Based Incentives and Company Performance: a Union-Nonunion Comparison,' *Industrial and Labor Relations Review*, 47, 594-609.

Drago, R. (1991) 'Incentives, Pay, and Performance: a Study of Australian Employees,' *Applied Economics,* 23, 1433-1446.

Drago, R. and Perlman R. (1989) 'Supervision and High Wages as Competing Incentives: a Basis for Labour Segmentation Theory,' in Drago, R. and Perlman, R. (eds) *Microeconomic Issues in Labor Economics: New Approaches*, Hemel Hempstead, Harvester.

Drago, R. and Heywood, J. (1992) 'Is Worker Behaviour Consistent with Efficiency Wages,' *Scottish Journal of Political Economy*, 39, 141-153.

Fernie, S. and Metcalf, D. (1995) 'Participation, Contingent Pay, Representation and Workplace Performance: Evidence from Great Britain,' *British Journal of Industrial Relations*, 33, 379-415.

Freeman, R.B. and Medoff, J.L. (1984) *What Do Unions Do?* New York, Basic Books.

Gallie, D. and White, M. (1993) *Employee Commitment and the Skills Revolution. First Findings from the Employment in Britain Survey*, London, PSI Publishing.

Green, F. and Weisskopf, T.E. (1990) ' The Worker Discipline Effect: a Disaggregative Analysis,' *Review of Economics and Statistics*, 72, 241-249.

Guest, D.E. (1995) 'Human Resource Management, Trade Unions and Industrial Relations,' in Storey, J. (ed) *Human Resource Management: a Critical Text*, London, Routledge.

Hogan, J.C. and Fleishman, E.A. (1979) 'An Index of the Physical Effort Required in Human Task Performance,' *Journal of Applied Psychology*, 64, 197-204.

Kelley, M.R. and Harrison, B. (1992) 'Unions, Technology, and Labor-Management Cooperation,' in Mishel, L. and Voos, P.B. (eds) *Unions and Economic* Competitiveness, Armonk, New York, M.E. Sharpe, Inc.

Lawler, E.E. (1971) *Pay and Organisational Effectiveness: a Psychological View*, New York, McGraw-Hill.

Maddala, G.S. (1983) 'Limited-Dependent and Qualitative Variables in Econometrics,' Cambridge, Cambridge University Press.

Meijman, T., Zijistron, F., Kompier, M., Mulders, H. and Broevse, J. (1986) 'The Measurement of Perceived Effort,' in Osborne, D. (ed) *Contemporary Ergonomics*, London, Taylor and Francis.

Metcalf, D. (1993) 'Industrial Relations and Economic Performance,' *British Journal of Industrial Relations*, 31, 255-283.

Metcalf, D. (1994) 'Transformation of British Industrial Relations? Institutions, Conduct and Outcomes 1980-1990,' in Barrell, R. (ed) *The UK Labour Market*, Cambridge, Cambridge University Press.

Millward, N., Stevens, M., Smart, D. And Hawes, W. (1992) *Workplace Industrial Relations in Transition: the ED/ESRC/PSI/ACAS Surveys*, Aldershot, Dartmouth.

Oster, G. (1980) 'Labour Relations and Demand Relations: a Case Study of the "Unemployment Effect",' *Cambridge Journal of Economics*, 4, 337-348.

Rebitzer, J.B. (1987) 'Unemployment, Long-Term Employment Relations, and Productivity Growth,' *Review of Economics and Statistics*, 69, 627-635.

Roy, D. (1952) 'Quota Restriction and Goldbricking in a Machine Shop,' *American Journal of Sociology*, 57, 427-442.

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

Trades Union Congress (1994) *Human Resource Management - a Trade Union Response*, London, TUC.

Chapter 5

Collective Sanctions as the Basis of Group Norms for Effort

5.1 : Introduction

In the literature review above, an efficiency wage paper by Akerlof (1982) was described, in which the author specifies a group norm for effort function, which he uses to argue that effort will respond positively to the wage received. At no point is a formal derivation of the group norm for effort function offered, however. In an attempt to provide stronger foundations for Akerlof's model, this chapter formally models, within a traditional choice-theoretic framework, the emergence of, and adherence to group norms for effort. This is done by introducing the idea of collective sanctions, already found in the sociological literature[1], into the economic analysis of the internal workings of firms.

The basic principle is that a workgroup adopts a single identity, when it comes to the reward or punishment of behaviour. All members of the group are treated equally, so that a single member's behaviour affects the outcomes for all his or her colleagues. This chapter looks at the particular case where collective sanctions affect workers' effort decisions. Thus, a single worker caught shirking has repercussions for the workgroup as a whole, who all suffer the consequences.

A number of points are worth making at this stage. First, it is not being claimed that such an arrangement is an accurate description of the internal workings of all firms. Rather, this chapter is asking what the consequences would be, *if* collective sanctions were in place. It could be argued, however, that even if collective sanctions are not explicitly used by a firm's management, workers may still behave as if they are. For example, an individual's negligence may increase the risk of an accident to all workers.

[1] See Heckathorn (1988, 1990)

There are probably many other examples where workers perceive that their own actions affect the payoffs of all group members.[2]

Assuming that the firm does adopt a policy of explicit collective sanctions, it could be asked why it would do this. The results of the model should demonstrate the potential benefits of the system, but some motivation for its introduction can be given at this stage, to justify the assumption of its use. The obvious advantage to the firm is that such a system will create a situation where individual workers do not want their colleagues to shirk, and thus the workers' interests will coincide with those of the firm. Also, using collective sanctions could save the firm money on its monitoring costs. In many situations it may require extensive monitoring to detect particular workers shirking, but it may be relatively easy to deduce if *someone* has been shirking, and this is the only information required in a system of collective sanctions. This principle is well known and well used by school teachers, who, rather than watch all the pupils all of the time, often threaten the whole class with a punishment when a misdemeanour is detected, in the knowledge that pressure from classmates will soon make the culprit confess.

Third, the phrase 'collective sanctions' has been borrowed from the sociological literature, but in the setting of this chapter the firm need not always adopt a punishment strategy. The natural implication of the term 'collective sanctions' is that if one worker is caught shirking, all members of the workgroup are punished, financially or otherwise. However, it could equally be a 'collective reward' for non-shirking, which is guaranteed, unless one member is caught shirking, in which case the reward is forfeited by all

[2] It should be noted that for any such example to accurately reflect an actual collective sanctions situation, it must be the case that the size of the payoff to all group members is independent of the total number of individuals shirking, since the sanctions analysed below have this property. Thus, arguing that an individual's shirking can harm the well-being of the firm, and thus all who work there, is not appropriate in this context, since the degree of harm done to the firm is likely to increase with the number of shirking employees.

members. Traditional effort extraction models, for example Shapiro and Stiglitz (1984), assume that firms always sack workers they catch shirking.[3] This chapter does not assume that all firms rule with an iron rod, rather at least allowing for the possibility that some firms may prefer to offer their workforce 'carrots.'

Of course, whether the sanction is positive or negative, the group members are losing something when evidence of someone shirking is found, and so this distinction would be irrelevant in strict expected utility theory. However, it will be shown below that it may make a difference whether the system is seen by the workers as a reward for non-shirking or a punishment for shirking.[4]

This point relates to one made by Rebitzer (1993), in a paper criticizing the existing effort extraction literature. Rebitzer saw it as a flaw that economic models in this area had narrow psychological foundations, in particular their assumption that all workers simply maximize the present value of expected utility, paying no attention to whether the incentive scheme in question is one that threatens punishment or one that offers rewards. While the present model does not use psychological theory either, it will be shown that such a distinction between incentive schemes can be important, if it influences the workers' decision on whether to cooperate with the scheme or not.

Another criticism made by Rebitzer (1993) similarly does not apply to the current chapter. He saw it as a fault that most papers attempting to model workers' effort decisions have no role for group norms, which he argued can strongly influence the effort

---

[3] This is an assumption which managers would argue does not accurately describe how such situations are resolved in practice - see Agell and Lundborg (1993).

[4] Luft (1994) provides evidence that a contract specifying a bonus will be preferred to a contract offering identical monetary payoffs, but expressed in the language of penalties. See also Kahneman and Tversky (1979) for evidence that gains and losses relative to some reference point are evaluated differently by individuals.

choice made by individual workers. The approach adopted below is to stay within the economist's traditional choice-theoretic model of the individual, yet it will be shown that group norms still arise as individually rational, without having to assume their presence in the first place.[5]

Of course, this is not the first paper to introduce such group norms. For example, Akerlof (1980) was one of the earliest attempts to model social customs. The key to his model is that economic agents are concerned about a loss of reputation should they ignore the social customs of society. To illustrate such an effect, he imagines an economy of labourers and capitalists trading their respective endowments with each other. He assumes that a social custom exists whereby agents are expected to trade capital for labour at an exchange rate $\bar{\omega}$ (with capital as the numeraire, this can be interpreted as a 'fair wage' for labour). He specifies a utility function of a labourer of type ϵ as

$$U = a_L + b_L K - c_L \epsilon d^R \mu \bar{R} - d^R d^C \bar{C}$$

where K is the number of units of capital held after trading, d^R is a dummy variable taking the value of one if the worker disobeys the code, and d^C is similarly a dummy indicating a worker who believes in the code. $\bar{C}$ is the loss of utility suffered by a believer who disobeys the code (caused by, for example, loss of self-esteem), while $\bar{R}$ is the loss of reputation suffered by one who disobeys. The loss of reputation is proportional to the fraction of agents who believe in the code, μ. The parameters of the model are given by a_L, b_L and c_L.

After defining a similar utility function for capitalists, Akerlof analyses the agents'

[5] This point refers to a criticism made by Marsden (1986) who argues that previous attempts to model adherence to group norms on an individualistic basis led to unstable equilibria, unless it was assumed that the group norm existed in the first place.

decisions, in particular determining whether or not it will be in their interests to obey the code. Given such decision-making, he derives a short-run equilibrium for the economy, in which the effective supply of labour is equal to the effective demand for labour at a wage $\omega \neq \bar{\omega}$, for a given fraction of believers, μ. The long-run equilibrium, derived subsequently, has the additional requirement that the fraction of the population who obey the code must equal the fraction who believe the code.

Akerlof shows that the resulting solution will typically involve multiple equilibria.[6] In particular, there will always be a possible long-run equilibrium in which no agents believe in the code ($\mu=0$). Other equilibria can exist, however, depending on the particular values taken by the various parameters. If reputation effects are unimportant, then equilibria with a proportion of the population believing in the code can only emerge if the self-esteem effects are large enough; in particular, if $\bar{C} > b_L$. As reputation becomes more important to the agents, it becomes more likely that equilibria with believers will emerge, and in the limit, if agents are sufficiently concerned about reputation (if $\bar{R}$ is sufficiently large), then an equilibrium can emerge with all agents believing in and obeying the social code.

Akerlof (1982)[7] returned to the idea of social customs, this time specifically looking at group norms for effort within a workplace, the subject of the present chapter. In the introduction to his paper, he says that workers choose an effort level that maximizes their utility, taking into account the group norm for effort, e_n:

---

[6] Multiple equilibria for a given set of parameter values is also typical of the solution to the model to be described below.

[7] This paper is described in detail as part of the literature review in Chapter 2.

$$\max\ u(e, w, e_n, \epsilon)$$

where ϵ is the individual's type. However, at this point, the paper departs from traditional economic modelling, by simply stating, rather than deriving, a group norm for effort function, using some psychological evidence in support:

$$e_n = e_n\{[w(e,\epsilon)], e_{min}, u_1, \ldots\ldots, u_J; w_0, u, b_u\}$$

where $w(e,\epsilon)$ is the wage scheme, e_{min} the work rules, u_1 ..... u_J the utility levels of colleagues, w_0 the outside or alternative wage, u the unemployment rate and b_u the level of unemployment benefits. There is no solid theory offered as to why such a group norm for effort should form, and similarly no explanation as to why workers should conform to the group norm (except for the arbitrarily stated utility function above which contains the group norm as one of its arguments).

Jones (1984) also offers models in which workers conform to the effort levels of their colleagues, the models resting on a conformism parameter, c, in the individuals' utility functions.

$$u_i = w(e_i) - b_i e_i^2 - c(e_i - e_j)$$

Hence utility is increased by choosing a similar effort level to that of fellow workers. Again, however, it is simply assumed at the beginning that the utility functions will include the conformism element, and so Jones' models cannot be described as theoretical justifications for conformist behaviour, rather only as descriptions of situations *should* workers display conformist tendencies.

One more paper that examines peer pressure and group norms is that by Kandel and Lazear (1992). Unlike the other papers mentioned, in this case the authors do try to offer a rationale for why such behaviour should evolve, by assuming the specific case of a profit-sharing firm. In such a set up, each worker's utility depends on the effort supplied by their colleagues, so that they have an incentive to influence the effort choices of fellow workers. Kandel and Lazear assume that there is no basic wage, and that all remuneration is in terms of shared profits, so that the utility received by individual i is

$$U_i = \frac{f(\underline{e})}{N} - c(e_i) - p(e_i, e_j.....e_N)$$

where $f(\underline{e})$ is a production function with a vector of effort levels, $\underline{e}$, as its argument, $c(e_i)$ is the utility cost of effort, and p(.) is a peer pressure function, with $\partial p/\partial e_i < 0$. The authors then define a specific peer pressure function, so that utility for individual i becomes

$$U_i = \frac{f(\underline{e})}{N} - c(e_i) - p(\bar{e}-e_i)$$

where $\bar{e}$ is the group norm for effort, defined as

$$\bar{e} = \frac{1}{N-1}\sum_{i \neq j} e_j$$

that is, the average effort level among the other workers. Hence, in this formulation, workers are penalised by their colleagues for working at an effort level below the group norm level. Then the peer pressure function could be set to $p=(\bar{e}-e_i)\gamma$, assuming that peer pressure can be manipulated, where γ is a measure of the penalty imposed should effort fall below the group norm. γ could then be set to the value that would lead to the efficient choice of effort, that is, the one a social planner would choose, once the individuals have maximized the above utility function with this value of γ included.

Having described other attempts to model group norms, the chapter continues by developing a new approach. The next section outlines the model in general, followed by a section that describes the effort equilibrium of a workgroup, under a system of collective sanctions. Section 4 analyses how a group norm for effort can form, and includes an appraisal of how the theoretical description of norm-related behaviour compares with that in the papers just referred to. The chapter ends with a discussion.

5.2 : The Model

The model of collective sanctions to be described has three stages. First, the optimum effort level, chosen individually by each worker, is determined. Given these individual choices, the next stage is to calculate the probability of collective sanctions being imposed. Once this probability is known, the third stage analyses the decisions concerning the establishment of a group norm for effort.

To proceed with the analysis, the following assumptions are made:

(i) Workers receive a fixed wage, w. Piece-rates, or any other similar individual-based incentive schemes, are therefore not considered. All workers in a given workgroup earn the same wage, either because of equity constraints, or the extra negotiation and administrative costs involved in deciding a particular wage for each individual worker.

(ii) Effort is an either/or choice; each worker can either work or shirk.

(iii) There is an implied monetary cost, c, associated with working rather than shirking. The model initially assumes that this cost of effort parameter is the same for all workers, then later allows workers to be heterogeneous with respect to their dislike of effort.

(iv) The firm operates a collective sanctions system, whereby all members of the specified workgroup each face sanctions, s, equally valued by all workers.

(v) Utility is defined over wages, effort and possible sanctions, is assumed to be additively separable and is normalised to $u(w,e,s)=w-c-s$. Thus individuals are assumed to be risk neutral, and the utility function expresses the returns to the job.

(vi) Collective sanctions are triggered when a single member of the workgroup is caught shirking. The model is not game theoretic since only the workers have any decisions to make; the firm does not make any. Specifically, it always imposes the sanctions when, and only when, shirking is observed. Assume that reputation effects prevent the firm dishonestly imposing sanctions when no shirking has in fact been detected, while sanctions that take the form of a transfer from workers to firm are sufficient to ensure that the latter has no incentive to renegotiate with the workers after a detection. Thus, although the workers would be willing to pay a sum of money, m, where $0<m<s$, to avoid sanctions, the firm will prefer simply to take the sanctions, s, and will not be interested in any renegotiation.

(vii) The probability that the shirking of any particular individual is observed by the firm is given by p. This does not necessarily mean that the culprit is identified, only that the firm is aware that someone has shirked. It will be important for the results derived that this probability is constant, and does not depend on the total number of shirkers. Therefore, there is just as much chance of a *particular* individual's shirking being discovered when there is no-one else shirking as when everybody else is shirking (though, of course, the *overall* probability that *somebody's* shirking is discovered will be higher in the second case). Thus, it is being argued that random surveillance by supervisors may turn up evidence that someone has been shirking, for example a low quality output, a machine not properly maintained, or a sweet paper on the floor. The assumption is that the chances of coming across a *particular* one of these is unrelated to the total number that exist; each one has an equal and independent probability, p, of being discovered.

Given the situation described by the above assumptions, each worker decides whether to work or to shirk in each period, given his or her perceptions of what other workers in the group will choose to do. The solution to the model is a non-cooperative Nash equilibrium in effort choices and expectations, that is where each group member's effort choice is the optimal choice, given what they expect their colleagues to choose, and these expectations about fellow group members' effort choices turn out to be correct. The chapter will be concerned with analysing the properties of this equilibrium, rather than examining the dynamic path towards it.

No attempt is therefore made to calculate an optimal lifetime path for effort, on the assumption that too much about the future is unknown to the workers, in particular the number of group members who will shirk in the future. Also, to keep the model simple, no attempt is made to consider strategic behaviour by workers, such as taking into account the effects of their own effort choices on the future choices of their colleagues. Thus, the effort decision of the individual is independent of the future.

The next section begins the analysis of such effort choices.

5.3 : The Individual Effort Choice

a.) Homogeneous Workers

The first stage of the model is to examine the effort decision of each worker, taken individually in a collective sanction regime. In later stages, once all decisions have been made, the probability of sanctions being triggered (by the detection of just one shirker) can be determined, which in turn will influence the workgroup's decision whether or not to introduce a group norm for effort in an attempt to reduce this probability.

Each worker must decide whether to work or to shirk. This decision will be affected by the cost of effort (initially assumed to be c, the same for all workers), and the worker's subjective opinion as to the *extra* likelihood of sanctions being triggered, if she were to shirk. The latter is in turn influenced by the worker's opinion concerning how many of her colleagues she expects to shirk, n^e.

If a worker shirks, then the probability of personally triggering sanctions is p, the probability of being observed shirking. Obviously, the probability of a non-shirker triggering sanctions is zero. Thus the probability of *not* triggering sanctions is (1-p) and 1 for a shirker and a worker respectively. Assuming independence, the expected probability that no-one triggers sanctions is given by

$$(1-p)\times.........\times(1-p)\times1\times.........\times1$$

where there will be n^e (1-p) terms, and $N-n^e$ unity terms, N being the total number of workgroup members. Thus

$$Pr(\textit{no-one triggers sanctions}) = (1-p)^{n^e}$$

and

$$Pr(\textit{someone triggers sanctions}) = 1-(1-p)^{n^e}$$

Therefore sanctions are considered to be more likely, the higher the probability of shirking being detected, and the greater the number of expected shirkers in the workgroup. Thus, although the probability that any *one* individual triggers sanctions remains at p however many shirkers there are, the probability that *someone* triggers sanctions rises with the total number of shirkers, as discussed in the previous section.

The expected returns of a non-shirker are then

$$E(U_n) = w - c - [1-(1-p)^{n^e}]s \qquad (1)$$

Sanctions still enter the non-shirker's returns function with probability $[1-(1-p)^{n^e}]$, because of the risk of one of the n^e expected number of shirkers triggering them.

A shirker will obviously not incur the cost of supplying effort, but she will increase the number of shirkers in the group to n^e+1, where n^e is now interpreted as the expected number of 'other shirkers,' before the individual in question makes her decision. Thus, a shirker increases the probability that sanctions are triggered. The expected returns of a shirker are then given by

$$E(U_s) = w - [1-(1-p)^{n^e+1}]s \qquad (2)$$

Each worker individually decides whether to shirk or not. The no-shirking condition, $E(U_n)>E(U_s)$ is therefore

$$w - c - [1-(1-p)^{n^e}]s \geq w - [1-(1-p)^{n^e+1}]s$$

$$\Rightarrow [(1-p)^{n^e} - (1-p)^{n^e+1}]s \geq c \qquad (3)$$

The left-hand side of this no-shirking condition is the cost of shirking, expressed as the lower probability, $(1\text{-p})^{n_e+1}$, that no-one triggers sanctions (that is, the increased likelihood of sanctions), of value s, when the individual shirks, compared with when she chooses to supply effort. If this cost of shirking outweighs the cost of supplying effort, c, then that worker will choose not to shirk. Further simplifying the condition gives

$$[(1-p)^{n^e}][1-(1-p)]s \geq c$$

$$\Rightarrow (1-p)^{n^e}ps \geq c \qquad (4)$$

Thus the previous probability of sanctions not being triggered, $(1\text{-p})^{n_e}$, is multiplied by p and therefore reduced, when another group member decides to join the n^e shirkers. Again, this simply means that the likelihood of sanctions is increased when another individual decides to shirk.

Since all workers are assumed to be identical, the equilibrium must have all group members either working or shirking. The effort level chosen will clearly depend on c, the cost of effort. The left-hand side of equation 4 is the 'cost of shirking' (COS), and this can be plotted as a downwards sloping function of the number of 'other shirkers.' If this curve lies everywhere below c, the common cost of effort, then all members of the workgroup will shirk, since supplying effort will always be more costly than the extra risk of sanctions resulting from shirking, no matter how many members of the group are expected to shirk. On the other hand, if the cost of shirking curve lies everywhere above c, then all members of the workgroup will supply effort, since such an action will be less costly than the increase in expected sanctions that would result from each individual's decision to shirk, irrespective of the number of her colleagues who are already shirking. Figure 1 illustrates these 2 situations.

An increase in the level of sanctions imposed when a shirker is detected will clearly increase the cost of shirking for any given number of 'other shirkers,' shifting the COS curve outwards. Thus a sufficiently large increase in the proposed degree of sanctions can turn a shirking workgroup into a working one (see Figure 2).

Of course, there is no reason why the cost of effort line cannot cut the cost of shirking line, as illustrated in Figure 3. In this case, two Nash equilibria exist; an all-shirking and an all-working group. Between these two extremes, no proportion of the workgroup shirking can be sustained as an equilibrium. If the expected number of shirkers is any number fewer than $\bar{n}$ (other than zero), then it will not be in the interests of any worker to shirk, hence the expectation is not consistent with the outcome, implying that such a number of shirkers cannot be an equilibrium. Similarly, any number of shirkers above $\bar{n}$ (but fewer than the whole group) cannot be an equilibrium, since if such a number were shirking, then the best response of all group members would be to shirk, so that an all-shirking group is the only possible Nash equilibrium above $\bar{n}$. Not surprisingly, therefore, in a group of identical workers, the only possible Nash equilibria involve everyone behaving the same; either all working or all shirking.[8]

[8]These are actually the possible pure strategy Nash equilibria. If mixed strategies are considered then another equilibrium emerges, with all group members shirking with probability $\bar{n}/N$, and not shirking with probability $1-\bar{n}/N$. In this case the expected number of shirkers will be $\bar{n}$, which is a possible equilibrium

The final variable that can influence effort choices is the probability of detection, p. An increase in p will rotate the COS schedule about a point along its length, as illustrated in Figure 4. As drawn, initially the only possible equilibrium has all group members shirking. Following an increase in the detection probability, the COS schedule rotates to COS′, where it now cuts the cost of effort line, c. As described above, there are then two possible Nash equilibria, with either the whole group working or the whole group shirking. Increasing the detection probability thus at least has the potential of transforming a shirking group into a working group. Which equilibrium position actually emerges will depend on how such changes affect expectations concerning colleagues' effort choices. This fact will be demonstrated algebraically in the section on heterogeneous workers that now follows.

b.) Heterogeneous Workers

From this point onwards it will be assumed that workers are heterogeneous with respect to their cost of effort, c_i. A number of reasons could exist for such a divergence in effort parameters across individuals, for example psychological differences implying various degrees of aversion to work, or that the more skilled or intelligent may find work less of a burden. The result of making this, more realistic, assumption is that intermediate numbers of shirkers can exist; the group do not have to all work or shirk as a whole.

Each worker will decide individually to supply effort if

$$(1-p)^{n'} ps \geq c_i \tag{5}$$

position, since at that number of shirkers, all workers are indifferent between working and shirking, so that their best response set will include both these strategies, and all mixed strategy probability distributions across them. Hence, the mixed strategies described are in each player's best response set, if all other players play a similar mixed strategy, and so this position constitutes a Nash Equilibrium.

As before, an increase in the size of the sanctions clearly has the effect of reducing the likelihood that each individual will shirk.

There are now two factors specific to the individual that affect her effort decision. Not surprisingly, the individual is more likely to shirk the greater her cost of supplying effort. The second individual factor, as before, is the number of other group members whom the individual expects to shirk, n^e. However, it is no longer sufficient for the individual to predict whether she belongs to a working or a shirking group in which all members are identical, since any division of the group between workers and shirkers is also possible. Differentiating the cost of shirking with respect to the expected number of shirkers:

$$\frac{\partial COS}{\partial n^e} = (1-p)^{n^e} \ln(1-p) ps < 0$$

Since p lies between 0 and 1, ln(1-p) must be negative, hence the cost of shirking falls with n^e. In other words, the more of her colleagues the individual expects to shirk, the more likely she is to shirk as well. Thus, even before group norms have been explicitly introduced, the influence of work culture on the individual's decisions is clear. The intuition, in terms of the model, is that if few co-workers are shirking, then the individual faces a high risk of triggering sanctions herself, if she decides to shirk. However, in a situation where most of her colleagues are shirking, there is a good chance that one of the others will trigger sanctions anyway, so the individual may as well shirk herself, i.e. the increase in the probability that someone triggers sanctions would be negligible, should the individual decide to shirk herself. Thus following behaviour, or conformism, emerges easily in this model.

This result will be strengthened, if each individual's cost of effort parameter is a positive function of the number of shirkers. This could be the case when there is a fixed amount of work to be performed by the group, so that those who choose to put in the effort will have to work harder to compensate for their shirking colleagues. Such a scenario can easily be imagined in the service sector where there are a fixed number of customers to be served, for example among sales assistants in shops or waiters in restaurants. Then, if an individual expects most of her colleagues to shirk, she is more likely to shirk herself for two reasons; first, because the cost of shirking will be low, as before, but now also because her cost of supplying effort will be high. Modifying equation 5, the no-shirking condition in this case is

$$(1-p)^{n^e} ps \geq c_i(n^e)$$

As n^e rises, the left-hand side falls, as before. Now, however, the right-hand side also rises, making it increasingly unlikely that the condition will be satisfied, and so even more likely that the individual will conform to the effort choice of her colleagues.

Turning to the probability of detection, p, it was seen that in the case of homogeneous workers, an increase in this probability may or may not convert a shirking workgroup into a working one. This result generalizes in the heterogeneous case, where any number of shirkers is possible within the group, so that an increase in the probability of detection may have a positive or negative effect on the cost of shirking, and hence on the number of shirkers in the group. To show this, differentiate the cost of shirking with respect to p;

$$\frac{\partial COS}{\partial p} = (1-p)^{n^e} s - n^e (1-p)^{n^e-1} ps$$

$$= (1-p)^{n^e} s[1-\frac{n^e p}{1-p}]$$

The terms outside the brackets are positive (assuming sanctions are used), hence raising the probability of detection will increase (decrease) the cost of shirking only if

$$\frac{n^e p}{1-p} < (>)\ 1 \qquad (6)$$

Thus, raising the probability of detection will only be beneficial to the firm in terms of increasing the cost of shirking, when there are few group members expected to shirk. The intuition behind this result is that when the number of 'other shirkers' is low, the individual is concerned about triggering sanctions herself, and so an increase in the detection probability will increase this concern, making the individual less likely to shirk. On the other hand, when most of her colleagues are shirking, the individual reasons that there is a good chance that sanctions will be triggered anyway, so she may as well shirk herself. Increasing the detection probability merely strengthens this reasoning by making it more likely that one of the other shirkers will trigger sanctions, so that there is an increased likelihood that the individual will shirk.

Equation 6 also shows that, for any given number of shirkers, increasing p can only have a positive effect on the cost of shirking over a certain range, beyond which further increases will actually reduce the cost of shirking (since individuals will increasingly reason that someone else will trigger sanctions anyway). Indeed, for any positive n^e, equation 6 shows that further increases in p above 0.5 must have a negative effect on the cost of shirking.

However, this is not quite the full story. Individuals only expect their colleagues to make the same effort decision if all other things remain equal. Therefore, when p changes, individuals may update their expectations. Writing n^e as a function of p,

$$COS = (1-p)^{n^e(p)}ps$$

and hence

$$\frac{\partial COS}{\partial p} = (1-p)^{n^e}s - n^e(1-p)^{n^e-1}ps + \frac{dn^e}{dp}(1-p)^{n^e}\ln(1-p)ps$$

$$= (1-p)^{n^e}s[1 - \frac{n^e p}{1-p} + \frac{dn^e}{dp}\ln(1-p)p]$$

The term outside the brackets is again positive if sanctions are in place, and so increasing the probability of detection will increase (decrease) the cost of shirking if

$$\frac{dn^e}{dp}\ln(1-p)p > (<) \frac{n^e p}{1-p} - 1$$

Previously, it was seen that once $n^e p/(1-p)$ was greater than unity, further increases in p would have a negative impact on the cost of shirking. Now that it is recognised that n^e will vary with p, this need not be the case, as long as the left-hand side of the inequality is large enough to outweigh the difference between $n^e p/(1-p)$ and unity. As ln(1-p) will be negative, this implies that as long as dn^e/dp is also sufficiently negative, increases in p will have a positive effect on the cost of shirking. In other words, as long as the individual believes that the increase in the detection probability will significantly reduce the number of shirkers in the workgroup, then the expected cost should she shirk will be increased, and hence she will be less likely to shirk herself. Thus the effect of raising the detection probability is determined by the following or conformist behaviour noted above; it will increase an individual's effort level to the extent that she believes it will increase the effort levels of the other group members.

Thus the cost of shirking to any individual, i, can be written as a function

$$COS_i = f(p,s,n_i^e) \tag{7}$$

where n_i^e has a negative effect, s has a positive effect, and p can have a positive or a negative effect on the cost of shirking.

The equilibrium number of shirkers in the group, n^*, will be a non-cooperative Nash equilibrium in expectations and effort choices. This requires, first, that the expectations of all members of the group must be correct:

$$n^* = n_i^e \quad \forall i \tag{8}$$

If this condition does not hold, then individuals will update their expectations, which may affect their effort decision in the following period, so that this period's choices cannot be an equilibrium. Next, given that everyone expects n^* workers to shirk, n^* workers must actually choose to shirk, according to the condition derived above. The remaining $N-n^*$ must choose to supply effort. Hence

$$(1-p)^{n^*-1}ps < c_i \quad \textit{for prop. } \frac{n^*}{N} \textit{ of group} \tag{9}$$

and

$$(1-p)^{n^*}ps \geq c_i \quad \textit{for prop. } \frac{N-n^*}{N} \textit{ of group} \tag{10}$$

Therefore, faced by n^*-1 'other shirkers,' n^* members of the group optimally decide to shirk, while the remaining $N-n^*$ members realize this, but still optimally decide to supply effort. In this sense n^* is then the equilibrium number of shirkers in the group, since the effort choice of each individual is optimal, given what they expect their colleagues to do, and these expectations are correct.

Given the cost of shirking function derived above (equation 7) and the fact that anything which increases the cost of shirking to a particular individual makes her less likely to shirk, and so may reduce n^*, a function determining the equilibrium number of shirkers can be written as

$$n^* = f(p, s, n_1^e, \ldots\ldots., n_N^e, c_1, \ldots\ldots., c_N)$$

where the equilibrium number of shirkers will rise with the expected number of shirkers and with all the cost of effort parameters, and will fall with the size of the sanctions. The detection probability can have a positive or a negative effect.

This function is not continuous in its arguments. Also, unique solutions for n^* are not guaranteed, and in fact are likely to be the exception.[9] The following numerical example illustrates both of these points. Assume there are five members of the workgroup, faced with collective sanctions of value 100. Let their cost of effort parameters, c_i, be 25, 20, 15, 10 and 5 respectively. The following table shows which equilibria are possible, for various detection probabilities.

[9] Akerlof (1980) also noticed the likelihood of more than one equilibrium in models involving social effects, writing 'There are multiple equilibria in the sense that many different customs, once established, could be followed in equilibrium. Indeed, such multiplicity is the essence of social custom.' (p.751)

no. of shirkers	detection probability								
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0			✓	✓	✓	✓	✓	✓	✓
1									
2									
3		✓	✓						
4	✓			✓					
5					✓	✓	✓	✓	✓

The explanation of these results is as follows. At very low detection probabilities, there will always be some shirkers within the group in equilibrium. The individuals with the higher cost of effort will always take the risk of triggering sanctions, given the low detection probability, and once the others have realized that such members are shirking, they will also start to shirk, illustrating the following behaviour discussed above. Individual 5, with her low cost of effort, continues to supply effort, until the detection probability reaches p=0.5.

The table shows that the detection probability must be at least 0.3 before a no-shirking equilibrium is possible. At this level of risk, individual 1 no longer considers it worthwhile becoming the *first* shirker in the group, and given her decision to work, everyone else finds it in their interests to supply effort as well. Thus nobody shirks if they each expect all their colleagues to supply effort, and so a no-shirking equilibrium is possible. Note, however, that an equilibrium with individuals 1, 2 and 3 shirking is also possible, in the sense that if, for some reason, expectations are formed so that everyone expects these three to shirk, then they will indeed shirk, with the remaining two still preferring to supply effort.

At detection probabilities of 0.5 and above, the two possible equilibria are the two extremes of an all-working or an all-shirking group (this should not be too surprising, given the inbuilt conformist nature of the model). Thus, at these higher detection probabilities, if everyone expects the other four to supply effort, nobody wants to take the risk of triggering sanctions themselves, and so all group members work. As soon as it is expected that someone will shirk, however, then a particular individual or individuals may decide to shirk themselves, and once they have been observed shirking by their colleagues, then the latter will reason that they may as well shirk also, given the high probability of the former triggering sanctions anyway. Thus the situation soon degenerates towards the all-shirking equilibrium. This degeneration will be faster, the higher the detection probability.

It can be shown diagrammatically or algebraically that as the detection probability rises, the two extremes of an all-working or an all-shirking group become increasingly likely as the two possible Nash equilibria. The former method involves diagrams similar to those depicted in the 'Homogeneous Workers' section above. In the diagrams that follow, the cost of shirking (COS) schedule is based on equation 5 above. To allow for the heterogeneity with respect to effort parameters, it will be assumed, without loss of generality, that the N members of the workgroup can be placed in order of decreasing cost of effort parameters (that is, placing the individual most averse to work in first position), so that the cost of effort line, c_i, can be drawn as a downward sloping function.[10] This line

[10] Thus the group members are placed in order along the horizontal axis, and the height of the c_i line at any point denotes the cost of effort to a particular individual. The height of the line at the right-hand edge denotes the cost of effort to the Nth individual, that is the group member with the lowest cost of effort parameter, giving the line a truncated appearance.

is drawn for the general case of a large group of N workers, rather than for the specific numeric example above, so that the function can be drawn as continuous. It will also be assumed that the distribution of effort cost parameters is uniform, so that this cost of effort line can be drawn as linear. Given these assumptions of continuity and a uniform distribution for the cost of effort parameters, the equilibria that emerge will depend on the slope of the cost of shirking line. It is possible for the cost of shirking line to lie below the cost of effort line at all points so that all group members always shirk, or for the cost of shirking line to lie above the cost of effort line at all points so that all group members always supply effort. These situations could occur when sanctions are very low or very high respectively. The diagrams that follow analyse the more interesting cases in which the curved cost of shirking line cuts the linear cost of effort line. Three situations will be analysed, each distinguished by a different detection probability, causing the slope of the cost of shirking line, and hence the possible equilibria, to vary in each case.

Figure 5 reveals a typical situation for a low detection probability, with a relatively flat COS schedule. The only possible Nash equilibrium is at point E, with n^* shirkers. For any number of shirkers less than n^*, say n_1, it is the best response for more than n_1 group members to shirk, since the cost of effort is greater than the cost of shirking to such individuals, so that number, n_1, cannot be an equilibrium. This will be the case for any number until n^* is reached. Similarly, for any number of shirkers greater than n^*, say n_2, it is the best response for more than $N-n_2$ group members not to shirk, since the cost of shirking is greater than the cost of effort to such individuals, so again, this number could not be a stable equilibrium. Hence, n^* is the only number of shirkers that constitutes a Nash equilibrium, in the sense that if everyone expects there to be n^* shirkers, exactly this number find it in their interests to shirk, and exactly $N-n^*$ do not want to shirk. In terms of the numeric example above, a single equilibrium with an intermediate number of shirkers was indeed found at the low detection probabilities of 0.1 and 0.2.

The intercept of the COS schedule (=ps) and its slope at $n^e=0$ (=ln[1-p]ps) are both increasing, in absolute value, in p. Hence, as the detection probability rises, the intercept of the COS schedule shifts up, and the schedule then falls away more steeply. Given these two facts, and an initial position as illustrated in Figure 5, successive increases in the detection probability must at some point lead to the position depicted in Figure 6.[11] Now there are two possible Nash equilibria; n^*_1, which equals zero, and n^*_2. Thus the all-working group becomes a possibility, since, if no other group member is shirking, then it is in the best interests of each individual to also work, since the cost of shirking is greater than the cost of effort to them all. The only other possible stable equilibrium position is at point E, determined in the same way as point E in Figure 5. The all-shirking group cannot yet exist in equilibrium, since even if everyone else is expected to shirk, it will not be in the interests of those with the lowest cost of effort parameters to shirk, since for them the cost of shirking will still be greater. Again, therefore, an intermediate number of shirkers can exist, in this case together with the possibility of an all-working group. This situation corresponds to the numeric example, above, at detection probabilities of 0.3 and 0.4.

If the detection probability continues to rise, COS will become steeper still. A situation could then be reached where it crosses the cost of effort line only once, from above, with the rate of its descent preventing it from re-crossing before the maximum group size is reached. Such a situation is illustrated in Figure 7. Now the only stable,

[11] An exception occurs if the value of sanctions, s, is so low that the COS schedule never becomes steeper than the cost of effort line, no matter what the detection probability. In this case, the situation will always remain as depicted in Figure 5.

pure strategy Nash equilibria involve either all group members working, or all shirking. Any intermediate number of shirkers cannot be an equilibrium, since if a number fewer than $\bar{n}$, say n_1, are expected to shirk, less than this number n_1 would actually find it in their interests to shirk, while if more than $\bar{n}$ group members, say n_2 are expected to shirk, then more than this number n_2 would actually find it in their interest to shirk. Neither of these numbers can be sustained as an equilibrium, therefore, since they would not be optimal responses to that number of individuals shirking. This would be the case for any number of shirkers greater than or less than $\bar{n}$, except at the extremes of an all working or all shirking group, where the number of shirkers, 0 or N, is consistent with everyone's optimal choice, given what everyone else is doing. Such a situation corresponds to all detection probabilities greater than or equal to 0.5 in the numeric example.

Analytically, it can be shown as follows that a two-equilibria solution, involving the two extremes of an all-working or an all-shirking group, becomes increasingly likely as the detection probability rises. The all-working group can exist if

$$(1-p)^0 ps \geq c_i \quad \forall i$$

If the N individuals are placed in order of their dislike of effort, the individual with the highest cost of effort parameter being placed first (that is, $c_1 > c_i$, $\forall\, i \neq 1$), then as long as this condition holds for the first individual, it will hold for all group members. It can then be written as

$$p \geq \frac{c_1}{s}$$

Similarly, the all-shirking group can exist if

$$(1-p)^{N-1}ps < c_i \quad \forall i$$

If this condition holds for the Nth individual, the group member with the lowest cost of effort parameter, then it will hold for all group members:

$$p < \frac{c_N}{(1-p)^{N-1}s}$$

Thus a two-equilibria solution to the model, with the two equilibria being the two extremes of an all-working or an all shirking group, will exist if

$$\frac{c_1}{s} \leq p < \frac{c_N}{(1-p)^{N-1}s}$$

It is possible for both inequalities to hold simultaneously since, although $c_1 > c_N$, it is also the case that $1 < 1 / (1-p)^{N-1}$. Indeed, as p rises, it becomes increasingly likely that both inequalities *will* hold. Clearly, the left-hand inequality is more likely to be satisfied as p rises, but this is also true of the right-hand inequality, since $c_N / s(1-p)^{N-1}$ is convex in p, and thus will rise at an increasing rate as p gets larger, so that the chances of being greater than p are higher, the larger is p. Thus, as was seen in the numerical example above, a multiple equilibria solution, with the two extremes as possible equilibria, becomes increasingly likely as the probability of detection is raised.

Two further points can be made. First, $c_N/s(1-p)^{N-1}$ is increasing in N, for any value of p. Hence the right-hand inequality will be satisfied at a lower value of p, increasing the likelihood of the two-equilibria solution of an all-working or an all-shirking group. Thus, the larger the group size, the more likely are the group members to be influenced by each others' behaviour and all make the same choices.

The second point is that if the cost of effort rises with the number of shirkers, as described above, then this will increase c_N, and so again the right-hand inequality will be satisfied at a lower value of p than previously. The larger is c_N with all group members shirking, relative to c_1 with no-one shirking, the wider will be the range of detection probabilities that satisfy the above inequalities, making the two-equilibria solution of an all working or an all shirking group more likely.

Which of the two equilibria actually occurs will depend on how individuals expect their colleagues to act. Thus, it can be seen that expectations play a crucial role in this model, with the major influence on individuals' behaviour being what they expect their colleagues to do. This link between the individual and the group is strengthened in the next section when group norms are introduced. First, it remains for this section briefly to consider 'collective rewards.'

c.) Collective Rewards

The analysis so far has been framed implicitly in a negative sanction regime. The purpose of this section is to show that the results obtained so far are unaffected if a positive reward regime is assumed.

'Collective rewards' have been defined to exist when each member of the group receives a bonus, s, unless any shirking is detected, in which case no group members receive the bonus. The probability that none of the expected n^e shirkers are detected was found to be $(1-p)^{n^e}$ above. Under collective rewards, this can be interpreted as the probability that the individual in question will receive the bonus. Thus, if the individual decides to supply effort, she will receive expected returns

$$E(U_n) = w - c_i + [(1-p)^{n^e}]s$$

If, on the other hand, she decides to shirk, she will increase the expected number of shirkers to n^e+1, and hence reduce the probability of receiving the bonus. Her expected returns in this case will be

$$E(U_s) = w + [(1-p)^{n^e+1}]s$$

Using these two equations, the no-shirking condition, $E(U_n) \geq E(U_s)$, reduces to the same condition as that obtained earlier in the case of collective sanctions (equation 5)

$$[(1-p)^{n^e} - (1-p)^{n^e+1}]s \geq c_i$$

or

$$(1-p)^{n^e} ps \geq c_i$$

Therefore, it makes no difference to the individual's effort decision whether the system is positive or negative in nature, that is whether the firm rewards good behaviour or punishes bad behaviour. This in turn implies that all of the comparative static results derived above continue to hold in the collective reward case. Of course, this is to be expected, since in either case the detection of shirking leads to a loss of value s to each worker. The only difference is the interpretation; the loss of the carrot rather than the use of the stick. Although this difference in interpretation can have no effect on the individual's effort choice in an expected returns framework, the next section will show that once group norms are allowed for, the distinction between rewards and sanctions can be important. This is the next stage, towards which the analysis now turns.

5.4 : The Group Norm for Effort

a.) The 'Within-Firm' Choice

Looking again at the numerical example of possible effort equilibria in the previous section, an interesting fact is revealed. In all equilibria in which some shirking exists, the expected level of sanctions[12] is greater than 25, and hence greater than the cost of effort to any of the five individuals. Therefore, in that particular example, whenever the system ends up at an equilibrium involving shirking, each worker would always be individually better off in a situation where all group members are supplying effort, rather than in the existing situation with its risk of sanctions. In other words, each member of the group would individually benefit from the establishment of a group norm for supplying effort. In such a situation, it is reasonable to expect such a group norm to form.

Considering the same example, it can also be seen that for all detection probabilities at least as large as 0.3, a no-shirking equilibrium is a possibility, where the fact that everyone else is supplying effort is sufficient to persuade each group member to do likewise. At these detection probabilities, it should be an easy matter for the establishment of a group norm to transform a shirking equilibrium into a no-shirking one. As long as everyone believes that all their colleagues will follow the group norm, then each group member will individually decide to supply effort. Therefore, no group member would ever have to be pressurised into supplying effort, and the only role that the threat of such peer pressure would play is to make it appear credible to each member that their

[12] The expected level of sanctions equals actual sanctions, 100, multiplied by the probability of them being triggered.

colleagues will all supply effort. Since all group members will be convinced in this way and so decide to supply effort themselves, the peer pressure need never actually be manifested.

At lower detection probabilities (p=0.1 and 0.2), the situation is slightly different, in that a no-shirking equilibrium could not arise without the presence of a group norm for supplying effort. At these detection probabilities, the individual with the highest cost of effort finds it individually optimal to shirk, even if no other group member is shirking. In this case the other group members must increase the cost of shirking to this individual, by applying peer pressure. Such persuasion will usually take the form of verbal abuse or ostracisation from the group, and maybe in extreme cases physical abuse, in the event of the worker shirking. It will be assumed that such tactics will be effective, in that they will always create a sufficient cost to make shirking unattractive. Such an assumption requires support, since it is the first time that this study has broken from usual economic modelling by including a term as vague as 'pressure from others' in individuals' utility functions. Ouchi (1981) provides such support, offering strong evidence that individuals are affected by what their colleagues think of them. He writes 'What we care about most is what our peers think about us....Failure to adhere to the norms of the group can bring loss of group support, approval, and ultimately the group can throw the offender out of the membership. To someone who feels closely integrated with any small group, these are severe penalties. More than hierarchical control, pay, or promotion, it is our group memberships that influence our behavior.' (p.28)

Two further points are worth making concerning peer pressure. First, even at low detection probabilities, it does not mean that group members are constantly ignoring or abusing each other. The mere threat of such treatment, should any individual flout the group norm for effort, should be sufficient to keep her in line, since this would reduce the *expected* returns from shirking, as long as the threat is believed to be credible. Thus individuals will choose the no-shirking option, without actually suffering the peer pressure. Although the threat of such an unpleasant working environment is therefore always there, it will only actually occur in cases where individuals did not consider it a credible threat. After occasional usage, this credibility should not be in question. Second, though at low detection probabilities some individuals find it optimal to shirk even if no other group member is shirking, in the above example, such individuals still find the group norm for supplying effort to be the best solution of all. The problem before the introduction of the group norm is that there is an externality effect, whereby the individuals who want to shirk do not take into account the effect of this decision on other group members, who may decide to follow this decision and shirk themselves, thus increasing the level of expected sanctions to such an extent that the first individuals would have been better off simply supplying effort in the first place. The group norm effectively internalizes this externality, by ensuring all group members choose not to shirk. Thus, since the individual herself is in favour of the group norm and actively supports its enforcement, it is likely that she would have feelings of guilt if she herself ignored it by shirking, which will further enforce the effects of the peer pressure from colleagues.

The discussion so far concerning the establishment of the group norm for supplying effort has been in terms of the numerical example of the previous section, but it can easily be generalised. Before the establishment of any group norm, the returns actually received (as opposed to the expected returns of the previous section) by a non-shirker and a shirker are respectively

$$U_n = w - c_i - [1-(1-p)^{n^*}]s$$

$$U_s = w - [1-(1-p)^{n^*}]s$$

where n^* is the equilibrium number of shirkers, as defined by equations 8, 9 and 10. If a group norm for effort is established, the returns to each group member are

$$U_{GN} = w - c_i - k$$

There is a cost, k, associated with peer pressure, constant across all individuals. In order to keep the argument simple, consideration of who imposes this peer pressure and the possibility of free-riding on the peer pressure of others will be avoided, by assuming that k is the cost of working in a peer pressure situation, suffered by all group members, regardless of whether they are actually exercising the pressure themselves or not.

There is deliberately no allowance made for the possibility of sanctions in the last equation. The group norms under consideration are ones in which peer pressure increases the cost of shirking to such a degree that it is in no individual's interests to shirk, thus removing all shirking and hence the risk of sanctions. In this case, k can be more specifically defined as the cost of working in a situation where sufficient peer pressure is exerted to prevent any group member wanting to shirk.[13] It will be assumed that this cost is a declining function of the proportion of group members who support the norm. This is a somewhat arbitrary assumption, adopted on the basis that the greater is the support

[13] Other group norms may exist whereby the group members exert weaker peer pressure, which might not deter the most determined shirkers. A risk of sanctions would then still exist, but this could be offset by a lower cost of imposing such reduced peer pressure. Since the aim of this work is to demonstrate how a group norm can be derived in a formal model, rather than to provide an exhaustive list of all possible norms, this extension will not be pursued here.

for the norm, the greater will be the moral justification for persuading others to follow this norm, making the exertion of peer pressure less unpleasant, and so less costly.[14]

A non-shirker will therefore support the setting up of a group norm for effort if

$$w - c_i - k(n^f/N) \geq w - c_i - [1-(1-p)^{n^*}]s$$

$$\Rightarrow [1-(1-p)^{n^*}]s \geq k(n^f/N) \qquad (11)$$

Thus, if the expected sanctions are greater than the cost of imposing peer pressure, with a proportion n^f/N of the workgroup in favour of the norm, then a particular non-shirker will add her support to the setting up of a group norm for supplying effort. The no-shirking condition (equation 5), written as

$$[(1-p)^{n^*}p]s \geq c_i$$

is by definition satisfied for a non-shirker. It can be shown that

$$1-(1-p)^{n^*} > (1-p)^{n^*}p$$

for all n^* not equal to zero, which are the situations under consideration when a group norm for effort is required. Assuming that it costs more to an individual to supply effort herself than to persuade someone else to do it, so that $c_i > k$, the following inequalities hold for a non-shirker:

$$[1-(1-p)^{n^*}]s > [(1-p)^{n^*}p]s \geq c_i > k$$

Hence equation 11 is always satisfied for a non-shirker, who will therefore always support

---

[14] The reason for making this assumption is that it facilitates the writing of the group norm as a Nash equilibrium, described below. The assumption is not crucial to whether a group norm will exist or not, and descriptions of such norms can be given where k is a constant amount regardless of the support for the norm.

the setting up of a group norm. This result is not too surprising. If an individual is not shirking herself, she will not be content with a situation where others are shirking and so risking triggering collective sanctions. It is therefore to be expected that a non-shirker would support the setting up of a group norm for effort that removes this risk.

Similarly, an individual who found it individually optimal to shirk, given the other group members' expected behaviour, will nevertheless support the establishment of a group norm for effort if

$$w - c_i - k(n^f/N) \geq w - [1-(1-p)^{n^*}]s$$

$$\Rightarrow [1-(1-p)^{n^*}]s \geq c_i + k(n^f/N) \qquad (12)$$

Hence, since a shirker is going to have to supply effort herself following the imposition of a group norm, the expected level of sanctions must outweigh the costs of both supplying effort and imposing peer pressure on others, if the shirker is to support the setting up of the group norm.

If equations 11 and 12 hold for all individuals, when all other group members support the group norm for effort, then such a norm will be a Nash equilibrium. Formally, the group norm will be a Nash equilibrium if

$$[1 - (1-p)^{n^*}]s \geq c_i + k(1) \qquad \forall i$$

Since it has been shown that the support condition is always satisfied for a non-shirker, this equation is expressed in terms of the shirkers' support condition, the more stringent requirement. Thus, if it is in each group member's interests to support the norm, given that everyone else does (so that the proportion in favour is unity) then such a norm will be a Nash equilibrium and it is reasonable to expect one to form.

However, it is not possible to prove that the 'support condition' will definitely hold for shirkers, as it was in the case of non-shirkers. Although it did hold in all possible equilibria involving shirking in the numerical example above[15], this is clearly not a general result. Is it then necessary for *all* group members to be in favour of a group norm, before an effective one that eliminates shirking can be set up? The answer is 'probably not' in practice, although this is a question that economics cannot answer adequately. I would argue that if only a few group members, those with the highest cost of effort parameters, were not in favour of a group norm for effort, then one could still emerge, as long as *enough* members were in support, so that the peer pressure that they impose on the abstainers would be sufficient to bring them into line and end their shirking. Therefore, although such individuals are better off shirking than *supporting* the group norm, they still *follow* the group norm, because the large support for it within the group implies that they will suffer large costs in terms of peer pressure from the supporters, should they ignore it.

The only question that remains is then, how many is 'enough' - what proportion of the group is needed to support a group norm for supplying effort before one is formed that can invoke sufficient peer pressure to prevent any abstainer shirking?[16] A small majority of the group being in favour is unlikely to be sufficient, since the significant minority will be large enough to form a subgroup of their own with, in effect, a group norm for shirking. The negative returns suffered by the members of this subgroup as a

[15] Assuming that the cost of imposing peer pressure on others, k, which was not included in the numerical example, is not too large.

[16] The continued focus of analysis is group norms that prevent *any* shirking.

result of peer pressure from the 'supporting faction' will then be greatly reduced and likely to be ineffective, since they will be influenced more by the members of their own subgroup. If this is the case, no effective group norm that prevents all shirking can be established, and the situation is likely to remain as described in section 3, with each individual making individually rational effort choices.

Therefore, it is likely that a significantly large majority is required to support a group norm for supplying effort, if it is to generate sufficient peer pressure to successfully prevent all shirking. Denote by n^{fc} this critical proportion required to make the group norm effective . Then $k(n^{fc}/N)$ is the cost of creating sufficient peer pressure to prevent all shirking when the norm has this degree of support. Given this cost, if the support conditions, equations 11 and 12 and in particular equation 12, are satisfied for n^{fc} group members, then a sufficient number will find it to be in their interests to establish a group norm, for that norm to be effective. Formally, an effective group norm for effort can exist as a Nash equilibrium if

$$[1 - (1-p)^{n^*}]s \geq c_i + k(\frac{n^{fc}}{N}) \quad \textit{for } n^{fc} \textit{ members} \tag{13}$$

Thus, n^{fc} group members find it in their interests to support a group norm for effort, given n^{fc}-1 of their colleagues are also in support, and this is the number required for the group norm to be effective. The norm is therefore a Nash equilibrium and it is reasonable to expect one to form. Note that n^{fc} is the minimum number of supporters required. If equation 13 is satisfied for any number greater than this critical amount, then an effective group norm with that degree of support will also be a Nash equilibrium.[17]

If more arguments were to be entered into the cost of peer pressure function, equation 13 makes it clear that anything that reduces the cost of working in a peer pressure environment will increase the likelihood of a group norm for effort forming, since the chances of the condition in equation 13 being satisfied for the required proportion of the workgroup will rise. A natural variable to include in the cost of peer pressure function would be the cohesion of the workgroup. In a naturally cohesive group, with good relations between members, less peer pressure may be required to persuade a colleague of the merits of the norm, compared to a situation where individuals are antagonistic towards each other. In addition, in a strongly cohesive group, feelings of guilt at letting down one's colleagues are likely to reduce the amount of peer pressure from others required to alter that individual's behaviour. Thus, less peer pressure should be required for the norm to be totally effective in a strongly cohesive group, and so the cost of exerting peer pressure will be correspondingly lower. Anything that increases the cohesion of the workgroup should then also increase the likelihood of a group norm for effort forming. Intuitively, if the presence of a trade union is a force for binding work colleagues together, then group norms for effort may then be more likely to exist in unionised, rather than non-unionised, establishments.

The discussion so far has been in terms of collective sanctions, but the results are exactly the same if the firm adopts a system of collective rewards instead. In such a situation, the actual returns received by non-shirkers and shirkers are respectively

---

[17] Equation 13 may be satisfied for proportions of the workgroup between n^{fc} and unity, since the lower level of k that comes with a greater degree of support for the norm may offset the higher cost of effort among the last of the workgroup to offer their support.

$$U_n = w - c_i + [(1-p)^{n^*}]s$$

$$U_s = w + [(1-p)^{n^*}]s$$

If a totally effective group norm for supplying effort is established, then each group member will receive the reward, s, with certainty, having paid the costs of supplying effort and working in a peer pressure environment. The returns received by each group member would then be

$$U_{GN} = w - c_i - k + s$$

Hence a non-shirker will support the setting up of a group norm if

$$w - c_i - k + s \geq w - c_i + [(1-p)^{n^*}]s$$

Rearranging gives the same 'support condition' as derived for non-shirkers in equation 11

$$[1-(1-p)^{n^*}]s \geq k$$

Similarly for shirkers, the same 'support condition' (equation 12) is obtained under a system of collective rewards as under a collective sanctions regime. Shirkers will support the establishment of a group norm for effort in a collective reward regime if

$$w - c_i - k + s \geq w + [(1-p)^{n^*}]s$$

$$\Rightarrow [1-(1-p)^{n^*}]s \geq c_i + k$$

Hence, since the support conditions are identical under the two schemes, the group norm

Nash equilibria under collective rewards will also be identical to those derived for a system of collective sanctions above (equation 13).

b.) Outside Options

The discussion so far in this section has assumed that the group members face a simple choice between supporting a group norm or not. However, even if all group members find it in their interest to support the establishment of a group norm for supplying effort, rather than allowing a situation to persist in which some or all group members are shirking, with its corresponding risk of sanctions being triggered, this still involves the cost associated with peer pressure on others. Thus, even if the returns derived from the group norm situation satisfy their internal support condition, it may be that some individuals would be better off not belonging to the firm at all, because of the peer pressure costs. For a group norm to be supported, therefore, the returns yielded must also exceed the returns available outside the firm. The latter can be expressed as

$$U_e = ub + (1-u)\bar{w}$$

Considering a collective reward scheme first, the returns received in a situation where a group norm for effort is established were shown above to be U_{GN}=w-c_i-k+s, so that a group norm can form if

$$w - c_i - k + s \geq ub + (1-u)\bar{w} \qquad (14)$$

where b represents the returns (monetary and non-monetary) received by the individual if she cannot find another job, which occurs with probability u, the unemployment rate. $\bar{w}$ is the market level of returns, that is the going wage paid in other firms, net of effort costs imposed, thus allowing for compensating wage differentials for more demanding work, where the cost of exerting effort is higher[18]. The market level of returns will be individual-specific, however, if a firm pays the same wage to all workers doing the same job, and bases this wage on the average cost of effort spent working on that job. Then different workers will receive different levels of 'alternative work returns,' since they will all receive the same wage in each job, but their cost of effort parameters will differ from the mean parameter, upon which this wage is based. An individual's personal level of 'alternative work returns,' $\bar{w}_i$, will be the same in *all* firms, however, if it is further assumed that her cost of effort parameter is the same distance from the mean in all firms, so that her returns also differ from the average market returns by the same amount in all firms.[19] Rewriting equation 14 incorporating these individual-specific alternative work returns gives

$$w - c_i - k + s \geq ub + (1-u)\bar{w}_i$$

This expression can be simplified if it is assumed that the firm under consideration also offers the going level of returns. Since incentives are provided by the collective reward scheme, the basic wage is free to perform its labour-allocating role, and so can adjust until returns are at the market level. Hence, for individual i, it can be written that $w-c_i=\bar{w}_i$, using the assumption that her cost of effort parameter will be the same distance from the

[18] Note that there is still an either/or effort choice in each firm, but the cost of choosing the 'effort' option rather than the 'shirk' option can vary across firms.

[19] The assumption that an individual's cost of effort parameter differs from the mean by the same amount in all firms implies that she will have, say, a low cost of effort parameter whatever job she does. This may be criticized on the grounds that her effort cost relative to other individuals may well be affected by how well suited to the type of work she is. It will be assumed that individuals do not know about such differences in advance, so that the *expected* level of alternative work returns for a particular individual will be the same in all firms.

mean in her current firm as in all other firms.

Rewriting the decision expression incorporating this assumption, an individual will support the setting up of a group norm for supplying effort if

$$s \geq k + u(b - \bar{w}_i) \qquad (15)$$

Thus, it is more likely that a group norm for supplying effort will evolve the higher the rewards on offer, s; the lower the cost of imposing peer pressure on colleagues, k; the lower are the returns while unemployed, b, relative to alternative work returns, $\bar{w}_i$ and the higher is the unemployment rate, u.

Referring back to equations 11 and 12 that describe the internal conditions for supporting a group norm for effort, it can be shown that if these conditions hold, then the external support condition given in equation 15 will definitely hold, since the following systems of inequalities are then true[20], for an original non-shirker and shirker respectively;

$$s > [1-(1-p)^{n^*}]s \geq k > k + u(b - \bar{w}_i)$$

$$s > [1-(1-p)^{n^*}]s \geq c_i + k > k + u(b - \bar{w}_i)$$

Therefore a firm need only set the level of the reward to ensure that the internal support conditions, equations 11 and 12, are met[21], and group members will find establishing a group norm for effort preferable to quitting and looking for a job elsewhere.

---

[20]The final inequality on each line is true, on the assumption that the alternative work returns, $\bar{w}_i$, are greater than the returns received while unemployed, b.

[21]In fact, since the internal support condition is always satisfied for a no-shirking individual, as shown in section 4a, the firm need only ensure that the relevant condition for original shirkers is satisfied, that is equation 12.

Hence the condition for a group norm to be a Nash equilibrium derived in the 'Within-Firm' section above, equation 13, continues to be sufficient, since a similar condition that considers outside alternatives will always hold when the former condition is satisfied.

This result does not hold when collective sanctions are examined, so that there is at last a difference between the two schemes. Under collective sanctions, an individual will receive higher returns by supporting the establishment of a group norm than she would find in the external labour market if

$$w - c_i - k \geq ub + (1-u)\bar{w}_i$$

If it is again assumed that the firm pays the going level of returns, before any sanctions are imposed, so that $w\text{-}c_i=\bar{w}_i$, this expression can be rewritten as

$$u(\bar{w}_i - b) \geq k \qquad (16)$$

Therefore, unless the possible drop in returns from $\bar{w}_i$ to b, with probability u, are large enough to outweigh the costs of imposing peer pressure on colleagues, k, then the worker will find it in her interests to quit the firm rather than participate in setting up a group norm for effort, and so will not support such a norm. This condition may or may not be satisfied for a particular individual, independently of whether the internal support condition is satisfied. Formally, a necessary condition, in addition to the internal condition derived above as equation 13, for an effective group norm to emerge as a Nash equilibrium is

$$u(\bar{w}_i - b) \geq k\left(\frac{n^{fc}}{N}\right) \qquad \textit{for } n^{fc} \textit{ members} \qquad (17)$$

Again, n^{fc} is the minimum level of support required for an effective group norm to be established, and if equation 17 is satisfied for any number of group members above n^{fc}, this number would also constitute an effective norm Nash equilibrium (assuming, of course, that the internal condition, equation 13, is also satisfied for that number of supporters).

If equation 17 does not hold for any number greater than or equal to the critical number, n^{fc}, of the workgroup, then there will be insufficient support for the group norm to be effective. In this case a group norm for supplying effort will not form, and the situation will be as described in Section 3, with all individuals making individually rational effort choices. Whether individual workers then do actually leave the firm will depend on how this internal effort equilibrium compares with external choices. One point can be made about such a comparison. If the internal support conditions are satisfied but the external support condition is not, then the group norm situation is preferable to the internal effort equilibrium, but is itself dominated by external choices, which implies that the external choices are preferable to the internal effort equilibrium. In such a situation, the firm can expect a high turnover rate, and may decide to abandon the policy of collective sanctions altogether, in favour of offering the market level of returns and using some other incentive mechanism. The next sub-section considers the firm's choices in more detail.

The effects of the two schemes therefore differ, with a collective reward scheme being more likely to induce the setting up of a group norm for effort, since this only requires the internal support condition, equation 13, to be satisfied, while a collective sanction scheme must satisfy this condition, and overcome an additional hurdle in the form of the external support condition, equation 17. Of course, this result derives from the fact that under collective rewards, group members can receive returns with a value s above the market rate, $\overline{w}_p$, while under collective sanctions, they can at best receive only the market rate. Taking this into account, the analysis now turns towards the choices made by the firm.

c.) The Firm's Choices

The whole discussion in this chapter has been concerned with the decisions made by the workers, rather than those by the firm. What the firm has to decide is whether to use an incentive scheme, and if so, whether to use a collective sanctions or a collective rewards system, as well as the size of the sanctions or rewards. It may also be able to influence the detection probability, p, but this will not be considered here.

Firms may decide to use an incentive scheme when effort is not perfectly observable. Which factors will influence whether a collective reward or a collective sanction scheme is used? The internal condition necessary for a group norm to exist as a Nash equilibrium (equation 13), is based on the internal support condition for shirkers, equation 12, which is identical whichever scheme is used. For either a collective sanctions or a collective rewards scheme to motivate the establishment of a group norm for effort, it must be feasible for the firm to set s so that equation 13 is satisfied. If it is not, then the firm must consider the effort equilibrium which emerges when all group members make individually rational effort choices. If this is not satisfactory, the policy may be abandoned and alternative incentive mechanisms sought.

If it is feasible for the firm to set s so that the internal support condition is satisfied, then the choice between collective sanctions or rewards depends on the external support condition. As demonstrated above, considering external options does not place any extra constraint on the possibility of a group norm emerging under a system of collective rewards. However, another necessary condition, given by equation 17, must be satisfied, in addition to the internal support condition, before a group norm for effort emerges under a system of collective sanctions. Thus a group norm is more likely to be created under a system of collective rewards. This comes at a cost to the firm, however, since it must pay the reward, s, as a bonus, whereas under a collective sanction scheme, it merely pays the standard market rate, $\bar{w}_i$, even when a group norm for effort is formed. Formally modelling this trade off faced by the firm would be an interesting extension, but is beyond the scope of this study. For now, a couple of the issues raised can be mentioned.

There is no variable in equation 17 that can be manipulated by the firm, which is therefore powerless to ensure that the condition is satisfied. Thus a collective sanctions scheme will successfully lead to the creation of a group norm for effort only in certain circumstances, beyond the control of the firm. If a collective sanctions scheme can be successful in inducing a group norm for effort at a particular firm, however (that is, if equation 17 is satisfied), then such a scheme will presumably be adopted. This is because it leads to the same level of effort[22] as a collective reward scheme, but at a lower cost to the firm, since it does not have to supply the bonus, s, to all group members. Hence, it should be expected that groups containing a large majority of members for whom equation 16 is satisfied will face collective sanction schemes. Looking at equation 16, it can be seen that this condition is more likely to be satisfied the higher is the unemployment rate, u. Thus workers with a higher probability of being unemployed should they relinquish their job, for example workers who are low-skilled, less well educated, working in a

[22] Since effort is an either/or choice, a group norm for effort will imply the same effort level whatever incentive regime is used. In more complex models with effort as a continuous variable, group members may be able to respond with different effort levels to different schemes.

declining industry or working in a depressed area, are more likely to encounter a collective sanctions scheme.

If equation 17 cannot be satisfied for a given group of workers, then the firm must accept that its collective sanctions scheme will not lead to the formation of a group norm for effort. In addition, as pointed out above, the returns associated with the external labour market may also dominate those offered by the internal effort equilibrium under a collective sanctions scheme, in which case the firm is likely to face turnover problems. In such a situation, the firm may turn towards a collective rewards scheme. The external support condition in this case is always satisfied whenever the internal support condition holds. Under a reward scheme, however, the firm must decide whether it is worth its while ensuring that the latter condition is satisfied, since it actually pays the amount s above the market rate, rather than receiving that amount in penalty payments. Thus the firm must decide whether the value of the extra output produced when a group norm for effort is established justifies the bonus payments above the market rate. This would again suggest that low-skill workers, with low value-added, may be excluded from sharing in collective rewards, which will therefore be reserved for groups of skilled workers with good outside employment alternatives.

Thus it should be observed that some firms use collective reward schemes, and others adopt collective sanctions schemes. This division could contribute, in part, to the definition of primary and secondary sectors. Firms offering collective rewards will be primary sector firms, supplying jobs to skilled workers, which offer a level of returns over and above the market rate, while collective sanctions firms will constitute the secondary sector, offering unattractive jobs to low-skilled workers and paying no more than the going wage. Although group norms can form under either regime, with the consequent imposing of peer pressure on colleagues, it is possible that this will be psychologically more attractive under a collective reward scheme. Group members may find it more acceptable to monitor each others' effort levels when they are all working towards receiving rewards, rather than in a sanctions system, where they may feel unease at supporting a system where the firm is punishing both them and their colleagues. Finally, it was shown above that workers are more likely under a collective reward scheme than in a collective sanction regime, to stay and create a group norm, rather than quit the firm. This also fits the primary/secondary sector interpretation, since a lower turnover rate is a well-documented feature of the primary sector.

Such a view is opposed to some interpretations of the work-discipline model, in which firms pay a basic wage above the market-clearing level and threaten observed shirkers with the sack. Such incentive schemes are sanction-based in nature (though individual in the basic work-discipline model, rather than collective) and as such, firms using them would constitute secondary sector firms, according to the above arguments. Chapters 2 and 4 above have commented on Bulow and Summers' (1986) paper, in which they argue that work-discipline arrangements will be found in primary sector firms. However, the available empirical evidence, such as Oster (1980), Rebitzer (1987) and Green and Weisskiopf (1990)[23], as well as the cost of job loss results in Chapter 4 of this thesis, suggests that the use of sanctions is more typical of secondary sector firms. This evidence is therefore more consistent with the interpretation offered in the present chapter, than with the model of Bulow and Summers (1986).

[23] All of these papers are described in detail as part of the literature review in Chapter 2.

d.) Previous Attempts to Model Group Norms

As was described in the introduction, this is not the first study that has attempted to model group norms for effort. Comparing the model presented above with, first, that of Akerlof (1982), Akerlof's work can be seen as a particular case of collective rewards, in the language of the current chapter. In particular, the reward, s, can be viewed as Akerlof's 'gift' of extra monetary payments, above that necessary simply to attract workers, $\bar{w}_i$. Akerlof argued that such a gift will be reciprocated by the workforce, through the establishment of a group norm for supplying effort above the minimum necessary. However, although he offers a psychological discussion in support of such 'gift exchange,' he does not attempt to derive such behaviour as optimal in a standard choice theoretic framework. Looking at the current model, as long as the group members expect the firm to withdraw its gift of extra monetary payments if this is not reciprocated with high group effort, then this would be exactly the situation described by the collective rewards sections above, and so this chapter provides an analytical derivation of the group norm for effort as optimising behaviour, absent in Akerlof (1982).

The collective sanctions model possesses a similar advantage over the papers by Akerlof (1980) and Jones (1984). Both begin by basically assuming what they are trying to model, Akerlof assuming that a certain group norm exists which will cause a loss of reputation, and hence utility, to those who disobey it, while Jones simply assumes that individuals derive utility from taking similar actions, in this case effort choices, to others. As such, both papers are interesting descriptions of the effects of such phenomena, but it could be argued that the model presented here takes the analysis one stage further by offering reasons for *why* group norms are created and *why* individuals conform to the effort choices of others.

As mentioned in the introduction, the paper by Kandel and Lazear (1992) does attempt to offer a rationale for why a group norm for high effort might evolve. A number of points can be made about their model, in comparison with the model presented above. First, it is concerned only with the specific case of profit-sharing. A similar criticism can be levelled at the model presented here, however, which looks only at the specific scenario of a firm adopting a policy of collective sanctions or rewards. In fact, both models can be generalised to situations where one person's actions affect the payoffs of all members, which is the underlying mechanism by which individuals care about their colleagues' effort levels in both models.

Second, Kandel and Lazear's group norm is simply specified as the average effort level, and will therefore change as individuals vary their effort levels. The concept of the group norm presented in the current chapter is more in line with the psychologists' view of a group norm, which is something above and beyond a simple aggregation of the individual effort levels. If one group member decided to disobey the norm and, for example, decided to shirk, this should in no way affect the group norm, which would still be to supply effort. The group norm presented above satisfies this property, since it is established as the optimal action for a significant majority of the group, and would not be affected by one deviant. With Kandel and Lazear's definition of a group norm, however, the norm would immediately change following the deviant's decision to shirk, as the average effort level would be recalculated. Surely this does not accurately capture the idea of a group norm, if it is so sensitive to individual behaviour?

The definition of a group norm adopted by Kandel and Lazear could also be crucial to any dynamics in their model, if they wanted to argue that the group norm influences behaviour. Although neither their paper nor this one considers the dynamic approach towards their respective equilibria, I believe that, if such dynamics were to be worked out, the model presented here could offer a more internally consistent description of the group norm's role. The problem with Kandel and Lazear's definition of the group norm is that it is the outcome, rather than the perpetrator, of change. It is therefore not clear what the role of the group norm would be in the dynamics of their model, since by definition it cannot change until individuals' effort choices have changed, and so there must be something else, rather than the norm itself, influencing the change in behaviour. The group norm in the collective sanctions model does not suffer from this timing problem, since it can adopt a value, for example to supply effort, without any initial change in the behaviour of any individual group member. The norm could then comfortably take its place as the main causation variable in the transformation of a shirking group into a working group.

Finally, it is the case in Kandel and Lazear's model that the peer pressure causes everyone to choose an effort level that they would not find individually optimal. This seems to pose a problem, which the authors are aware of, as they write, 'workers in a firm with peer pressure may be worse off than those in one without it. While pressure guarantees higher effort, it does not guarantee higher utility since the pressure itself is a cost borne by all members of the firm. It may well produce higher effort levels, but workers may feel bad about working in an environment that has rampant peer pressure,' (p.805). Akerlof (1980) is aware of a similar problem in his model when he notes that, 'a social custom may be obeyed even though it is to everyone's economic disadvantage to obey it,' (p.751). It is not then clear what forces could ever form the group norms in these two models in the first place. Although Elster (1989) argues that some group norms cannot be rationalised in terms of human self-interest, and presumably come into being for some higher, moral reason, I do not think that this argument applies to the group norms under consideration here. I do not think that workers are under any moral obligation to supply high levels of effort, and believe that a group norm for supplying effort will only evolve if the group members desire one. There is no reason in Kandel and Lazear's model why any individual worker should support the setting up of a group norm for effort, and so it is difficult to imagine how one could arise.

This is less of a problem in the current model, since, as long as a no-shirking equilibrium is a possibility at the given detection probability, then the only role for the group norm is to persuade individuals that their colleagues will choose not to shirk. Given this information, all group members would individually choose to supply effort, even in the absence of peer pressure. Thus the group norm is not forcing anyone to do something that they do not want to do, only it creates the situation in which individuals find supplying effort to be desirable. Even in the cases when a no-shirking equilibrium is not a possibility in the absence of a group norm (that is, at low detection probabilities), it may still be the case that all individuals are better off in a group norm situation than in the free choice effort equilibrium. This can occur because a group member's individually-rational decision to shirk may influence others to follow, leading to a high risk of sanctions. A group norm internalizes this externality, and so may be the preferred situation for all group members. If this is not the case, and a significant number are worse off under the group norm for effort, then it was argued that such a norm would not emerge. Thus a significant majority must be in favour of a group norm, otherwise one would not evolve. Again, therefore, the majority are not doing anything against their wishes (unlike in Kandel and Lazear's model where *everyone* makes a choice they would not otherwise make) and it is only the individuals with the highest cost of effort parameters who are being influenced to make choices that they would not find individually optimal. I believe this to be the principal reason for preferring the group norm as derived in this chapter.

5.5 : Summary

This chapter has analysed the effects, on the behaviour of workers, of collective sanction or collective reward schemes, whereby all group members suffer the consequences of just one of their number being observed shirking.

The first stage examined the effort equilibrium of the workgroup, as a non-cooperative Nash equilibrium in expectations and effort choices. It was seen that multiple equilibria are possible, and indeed, as the probability of shirking being detected rises, the two extremes of an all-working or an all-shirking group become increasingly likely to be two possible equilibria.

The chapter then examined the possible formation of group norms for effort. It was argued that a group norm will form if it is in the interest of a large majority of the group, in terms of comparing the returns that they could expect to receive if a group norm were to form, with both the expected returns derived from the existing effort equilibrium, and the returns they could expect to receive outside the firm should they quit their current job. It was shown that, although the use of collective sanctions or collective rewards made no difference to the former, internal, comparison, a distinction existed when group members made the comparison with alternative, external levels of returns, with a collective reward scheme being more attractive and hence more conducive to the establishment of a group norm for effort. This advantage of collective rewards comes at a cost to the firm, however, since it must pay the promised reward, or bonus, above the going market level of returns. It would be interesting to model this trade off in future work, to determine whether the firm will choose a collective reward or a collective sanction system. More generally the firm's behaviour could be modelled in more depth, with the aim of comparing the incentive schemes suggested in this chapter with others that have been proposed in the literature, for example individual or group-based piece rates and tournaments.

The group norm for effort derived in this chapter appears to possess certain advantages over previous attempts to model such norms. Papers by Akerlof (1980 and 1982) and Jones (1984) examine the consequences of group norms existing, but do not explain *why* they are created in the first place. The group norm analysed in this chapter does attempt to derive the formation of a group norm from first principles, using the standard choice-theoretic techniques used by economists. The fact that the group norm is derived on an individualistic basis from first principles implies that Marsden's (1989) criticism is not relevant in this case. Marsden argued that attempts to model group norms on an individual basis are flawed, since non-norm-guided individuals can always enter the relationship, and do at least as well as the norm-guided agents, if not better. Thus, for the norm to be a stable equilibrium, new entrants must similarly be affected by the norm. Hence, the argument that all individuals follow, and will continue to follow, the group norm rests on the assumption that the norm exists in the first place. In the model presented above, however, the actual forming of the group norm was discussed, and at no point was it assumed that the norm already existed.

The paper by Kandel and Lazear (1992) is probably the closest to the present one in deriving a group norm for effort. However, it was argued that their group norm was inferior to the one developed here, for two reasons. First, their group norm was simply defined as the average effort level, rather than being something over-and-above the constituent effort levels of the workgroup members, as presented above. It was hypothesized that this could cause problems were they to work out dynamics towards the equilibrium in their paper, since it is not clear how the group norm could influence effort decisions, when it is itself merely the average effort level chosen. Second, it is the case in Kandel and Lazear's model that all individuals choose an effort level under the influence of the group norm, which they would not find individually optimal, thus bringing into question why the group norm should form in the first place, or continue to exist. In the present model, however, the formation of the group norm is explicitly modelled, and indeed it is argued that one cannot exist unless a significant majority of the workgroup is better off due to its presence.

Finally, it should be mentioned that collective sanctions have been used to model a very particular type of group norm. Clearly the model presented will not generalise to the modelling of other group norms. Indeed, I would agree with Elster (1989) that some group norms cannot be rationalised purely in terms of economic self-interest, for example society's norms of decent or civilised behaviour. Group norms for effort, on the other hand, are not based upon some moral code, but are created by (presumably rational) economic agents, and as such it should be possible to model them in an economic framework. This chapter has been an attempt to do just this.

Figure 5.1a : COS Everywhere Below the Cost of Effort Line

A Shirking Group

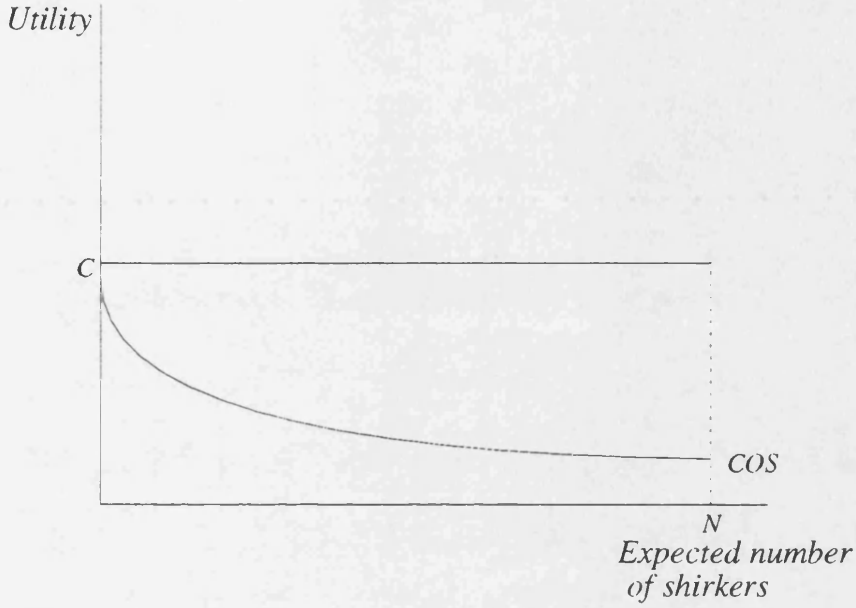


Figure 5.1b : COS Everywhere Above the Cost of Effort Line

A Working Group

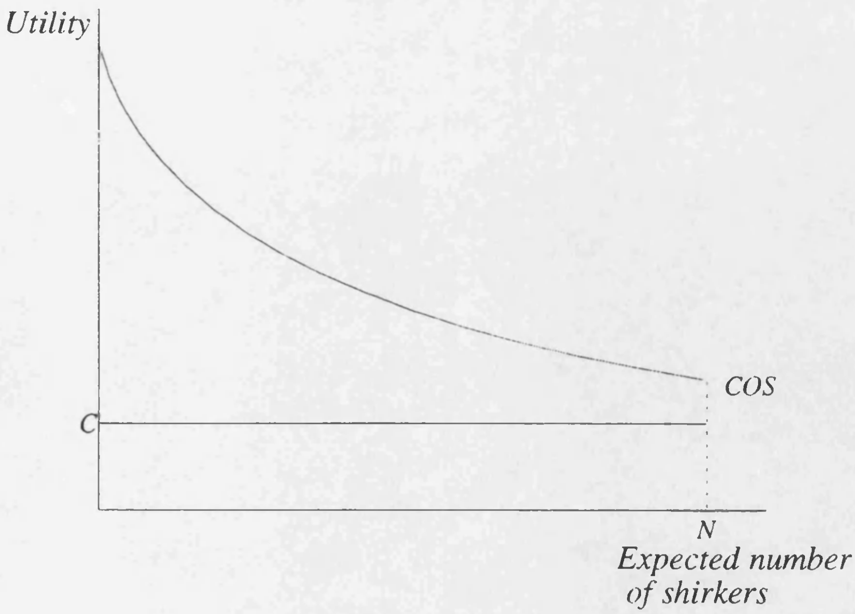


Figure 5.2 : Increasing the Level of Sanctions

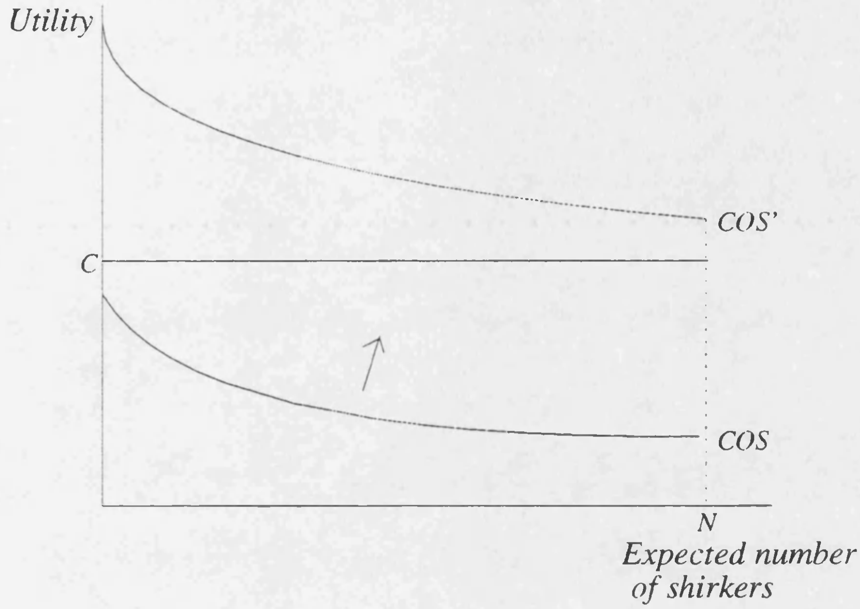


Figure 5.3 : COS Intersecting the Cost of Effort Line

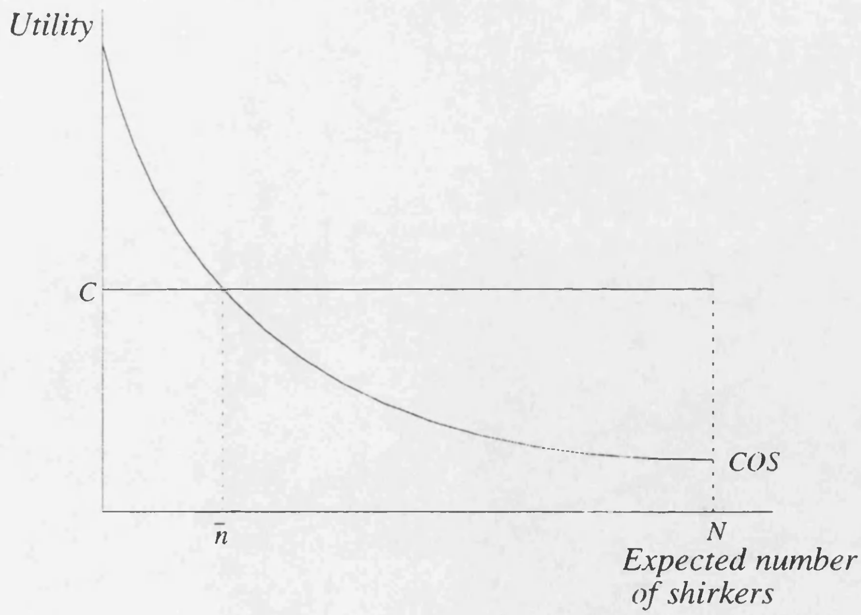


Figure 5.4 : Increasing the Detection Probability

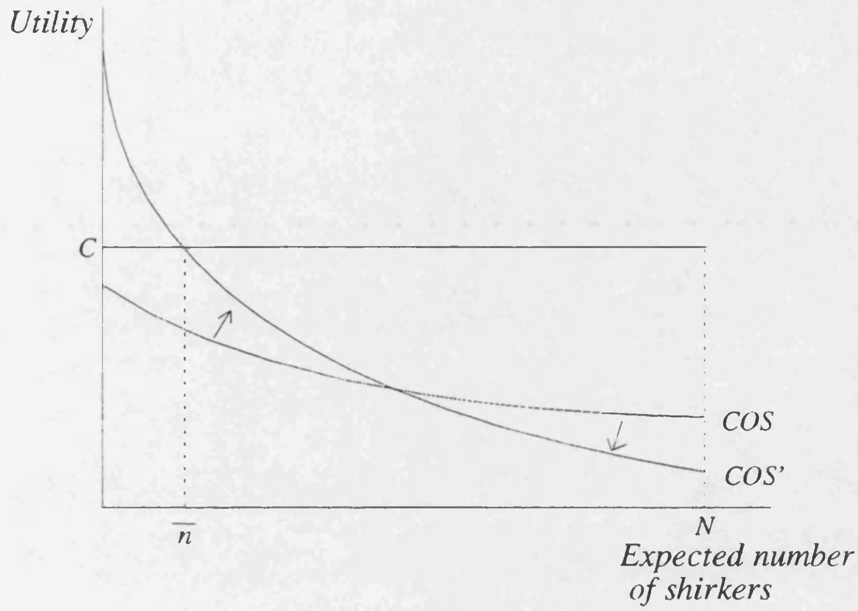


Figure 5.5 : Heterogeneous Workers and a Low Detection Probability

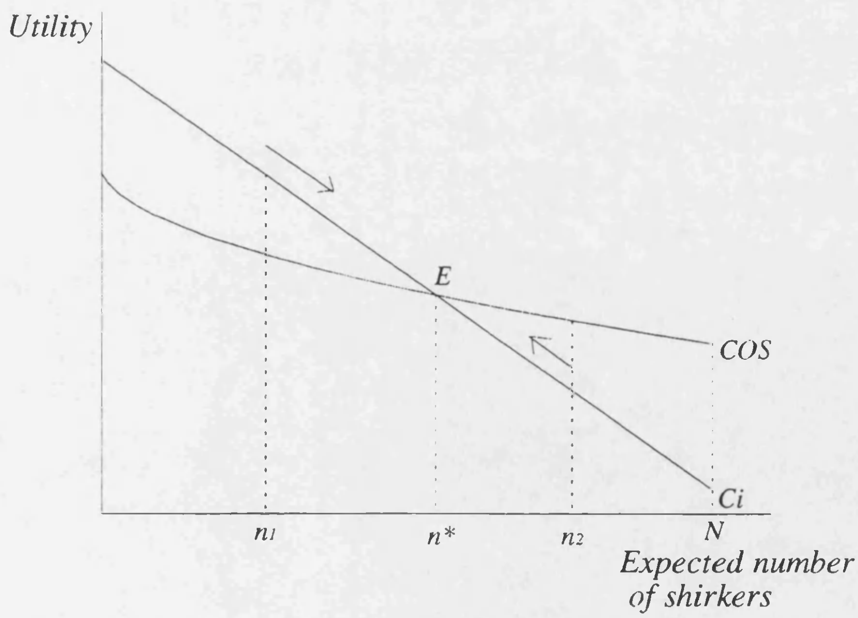


Figure 5.6 : Heterogeneous Workers and an Intermediate Detection Probability

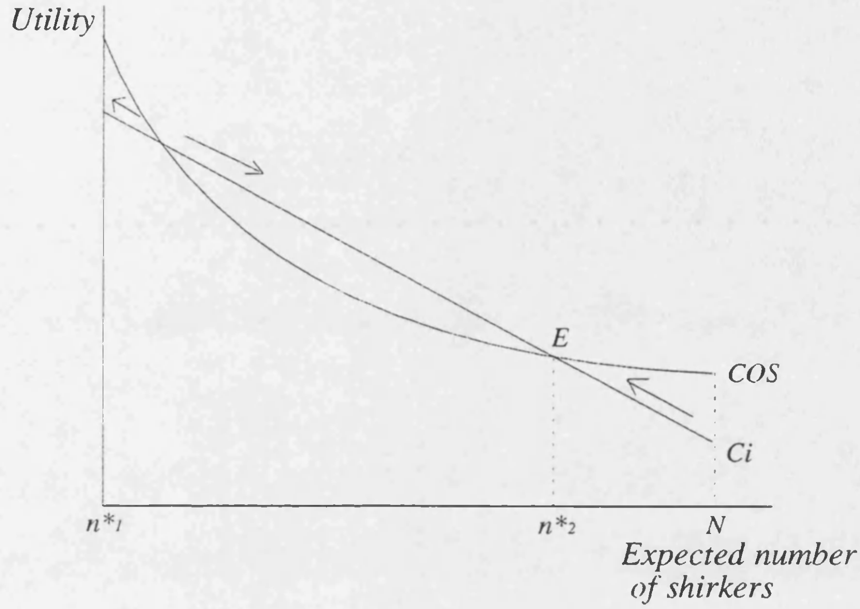
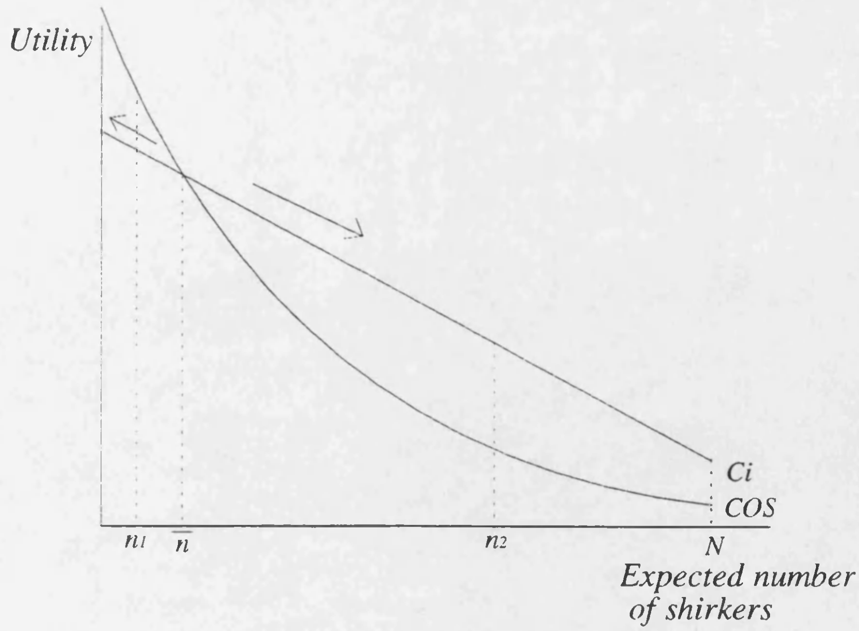


Figure 5.7 : Heterogeneous Workers and a High Detection Probability



References

Agell, J. and Lundborg, P. (1993) 'Theories of Pay and Unemployment: Survey Evidence from Swedish Manufacturing Firms,' Working Paper 1993:8, Department of Economics, Uppsala University.

Akerlof, G.A. (1980) 'A Theory of Social Custom, of which Unemployment May Be One Consequence,' *Quarterly Journal of Economics*, 94, 749-775.

Akerlof, G.A. (1982) 'Labor Contracts as Partial Gift Exchange,' *Quarterly Journal of Economics,* 97, 543-569.

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

Elster, J. (1989) 'Social Norms and Economic Theory,' *Journal of Economic Perspectives*, 3, 99-117.

Green, F.G. and Weisskopf, T.E. (1990) 'The Worker Discipline Effect: A Disaggregative Analysis,' *Review of Economics and Statistics*, 72, 241-249.

Heckathorn, D.D. (1988) 'Collective Sanctions and the Creation of Prisoner's Dilemma Norms,' *American Journal of Sociology*, 94, 535-562.

Heckathorn, D.D. (1990) 'Collective Sanctions and Compliance Norms: A Formal Theory of Group-Mediated Social Control,' *American Sociological Review*, 55, 366-384.

Jones, S.R.G. (1984) *The Economics of Conformism*, Oxford, Blackwell.

Kahneman, D. and Tversky, A. (1979) 'Prospect Theory: An Analysis of Decision Under Risk," *Econometrica*, 47, 263-291.

Kandel, E. and Lazear, E.A. (1992) 'Peer Pressure and Partnerships,' *Journal of Political Economy,* 100, 801-817.

Luft, J. (1994) 'Bonus and Penalty Incentives: Contract Choice by Employees,' *Journal of Accounting and Economics*, 18, 181-206.

Marsden, D. (1989) *The End of Economic Man?* Wheatsheaf Books, Brighton.

Oster, G. (1980) 'Labour Relations and Demand Relations: a Case Study of the "Unemployment Effect",' *Cambridge Journal of Economics*, 4, 337-348.

Ouchi, W.G. (1981) *How American Business Can Meet the Japanese Challenge*, New York, Addison-Wesley Publishing Company, Inc.

Rebitzer, J.B. (1987) 'Unemployment, Long-Term Employment Relations, and Productivity Growth,' *Review of Economics and Statistics*, 69, 627-635.

Rebitzer, J.B. (1993) 'Radical Political Economy and the Economics of Labor Markets,' *Journal of Economic Literature*, 31, 1394-1434.

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

Chapter 6 : Conclusions

The introduction to this thesis gave the limited amount of information on the determinants of workers' effort choices as the motivation for this study. A number of questions were asked concerning how effort varies with certain characteristics. How far have we come in answering those questions?

Chapter 3 used individual level data from the 1989 British Social Attitudes Survey, and thus provided evidence on how effort varies with individual characteristics. That effort does indeed vary across individuals seems clear. The results showed that females, the young and the less well-educated put more effort into their jobs. Another important determinant of work effort is the individual's responsibilities away from work. Thus, having primary responsibility for the care of children at home is seen to significantly reduce effort provided at work, while having a non-working spouse who presumably takes on such responsibilities is associated with higher effort at work. However, once working conditions were controlled for, the significant 'household' effects disappeared. It would appear that individuals with responsibilities outside of their work choose less exhausting jobs, for example ones which do not involve hard physical work or stress.

The analysis of individual characteristics was followed by an empirical test of efficiency wage theory, by including earnings in the estimated effort equation. The results were consistent with the central prediction of efficiency wage theory that an individual's effort responds positively to the wage that he or she receives. The analysis attempted to allow for the endogeneity of earnings by estimating the effort equation in a two-stage least squares framework. The results lack conclusiveness, however, because of the arbitrary choice of the exclusion restrictions needed to identify the effort equation. It is difficult to see how future work can solve this problem. The inability of econometricians to conduct natural experiments means that explanatory variables are probably being affected by a host of other factors, leading to serious endogeneity problems. The closest situation to a natural experiment in this area could be a change in minimum wage laws, which would be exogenous as far as the firm is concerned. If a measure of effort could be found for firms affected by such a law change, and collected both before and after the change, then this would provide clear evidence on how effort responds to wages, without the possibility of a reverse causality argument.

The analysis of Chapter 3 at least attempted to counter the argument that the observed wage-effort correlation can be explained by the existence of compensating differential payments, with causality running from effort to wages. A new test was developed, which made use of a job satisfaction variable in the data set. The results showed that effort significantly reduced job satisfaction at a given level of earnings, but when the earnings variable was removed from the equation and thus allowed to vary, the significant relationship between effort and job satisfaction remained. If compensating differentials were at work, then earnings should have moved to equalise satisfaction or utility across all jobs with varying levels of effort requirements. It was therefore suggested that the efficiency wage interpretation of the positive wage-effort correlation is the more likely explanation.

The final experiment in Chapter 3 considered the effect of local unemployment on effort. The results revealed that effort responds positively to the change in unemployment. However, the effect was small, just failing to achieve statistical significance at the 10% level, and was dominated by the wage effect on effort. When the analysis was repeated for eight countries in the International Social Survey Programme, this same relative magnitude of effects was found in every case, which reveals the benefits of performing analyses on different data sets for improving the credibility of results. A possible reason for the small unemployment effects obtained is that individuals' opinions as to their personal re-employment probabilities may not be accurately reflected in the local unemployment rate. Attempts to analyse this point did not lead to significant results, however.

Unless new data or methods become available to deal with the problem of endogeneity in earnings variables, there is probably not much more work that can be done on the relationship between effort choices and individual characteristics. This is not true of the establishment characteristics in Chapter 4, which pose interesting questions about the mechanisms at work behind some of the results. The key results of that chapter include that effort is positively associated with the cost of job loss, defined to be the product of the average wage earned and the local unemployment rate. HRM variables such as upward communication from employees to managers and payment by results schemes were also positively associated with average effort levels in the workplace. As for organisational change, effort was seen to be higher in establishments that have recently introduced new plant or machinery. Finally, the variables included to control for average workforce characteristics revealed that effort is higher in workgroups that contain a greater proportion of unskilled and non-white employees.

When the sample was split into high and low union density workplaces, some interesting differences emerged. The cost of job loss effect was found to exist only in the low-density sector, while among the HRM variables, joint consultative committees, upward communication and payment by results schemes were all associated with significantly higher effort when accompanied by a strong union, but not when accompanied by a weak union, (joint consultative committees actually attracted a significantly negative coefficient in the low-density equation). As for the organisational change result, the statistically significant positive relationship between effort and the introduction of new machinery, that was obtained in the full sample, was found to exist only in the low-density equation when the sample was split.

Chapter 4 considered possible reasons for all of these results, but the discussion was mostly educated speculation rather than rigorous analysis. It was suggested that a strong union will implement formal dismissal procedures, which reduce the firm's threat of dismissal and so the motivating impact of a high cost of job loss. Similarly, if new machinery is introduced in an attempt by management to extract more effort from their workforce, it was argued that a strong union can reduce this effect by negotiating how that machinery is introduced. HRM policies, on the other hand, were seen to benefit from the presence of a strong union, with the reason put forward being that employees will have more faith in such schemes when backed by a powerful union, as well as union 'voice' effects improving the implementation and effectiveness of the schemes. I feel that future research on trade unions may benefit from a rigorous analysis of these effects, thus concentrating on how unions influence the internal workings of firms, rather than just outcomes such as wage levels.

The final chapter in this thesis attempted to improve the theoretical efficiency wage literature, in particular providing a more formal framework for the sociological 'Gift Exchange' model. This model lacks a rigorous derivation of its central point, that a workgroup will respond to a high wage by forming a group norm for supplying high levels of effort. Economic research rarely considers such 'sociological' phenomena as group norms, and attempts that have been made have struggled to successfully describe group behaviour within the economist's typical individual utility maximisation framework. I think that economics can benefit from consideration of such issues, though, as it tries to explain individual behaviour.

The model of Chapter 5 took as its starting point that individuals in a workgroup are all affected by a single member's shirking, by assuming that the firm implements collective sanction or collective reward schemes. Given this, and without straying from the principle of individual utility maximisation, a group norm for effort was derived as a Nash equilibrium, in which it was in each member's interest to support and adhere to the norm, if all their colleagues did likewise. It was also shown that this norm was more likely to emerge under a system of collective rewards than in a collective sanctions regime. This result added to a recurring theme throughout the thesis that dismissal threats and sanction-based incentive schemes are typical of the secondary sector, while primary sector firms are more likely to adopt more positive, reward-based motivation techniques. Given that the collective reward system mirrors the situation in the 'Gift Exchange' model, it was argued that the presented model provides the theoretical underpinnings for that version of efficiency wage theory, to bring it into line with the other models in that class. It is hoped that this then adds one more piece to our understanding of the determinants of workers' effort.