

EQUILIBRIUM INCENTIVE CONTRACTS AND EFFICIENCY WAGES

Espen R. Moen
Norwegian School of Management

Åsa Rosén
Swedish Institute for Social Research,
Stockholm

Abstract

We analyze the optimal (efficiency) wage contract when output is contractible but firms neither observe the workers' effort nor their match-specific productivity. Firms offer wage contracts that optimally trade off effort and wage costs. As a result, employed workers enjoy rents, which in turn creates unemployment. Nonetheless, the incentive power of the equilibrium wage contract is constrained efficient in the absence of taxes and unemployment benefits. We also show that more high-powered incentive contracts tend to be associated with higher equilibrium unemployment rates. (JEL: E24, J30, J41)

1. Introduction

Efficiency wage theory is a prominent explanation for unemployment. Its core idea is that wages play other roles than clearing the market. In particular, firms may set wages above the market clearing wage in order to motivate workers (Shapiro and Stiglitz 1984; Akerlof 1982), recruit high-quality workers (Weiss 1980), or retain workers (Salop 1979). In all cases there is unemployment in equilibrium.

A key feature of the Shapiro-Stiglitz shirking model is that firms cannot condition wages on output. Although such performance independent remuneration may be an appropriate description for some labour markets, it is less so for

Acknowledgments: We have benefited from comments by seminar participants at the Norwegian School of Management in Oslo, University of Helsinki, Tinbergen Institute (Amsterdam and Rotterdam), the Norwegian School of Economics and Business Administration in Bergen, SED Annual Meeting in Stockholm, Stockholm University (IIES and SOFI), IZA in Bonn, University of Essex, Umeå University, Uppsala University, CEMFI, the CEPR conference Incentives and Beyond, and the Trade Union Institute (Stockholm). Moen would like to thank the Economics Department at the Northwestern University for their hospitality while he was visiting and working with this article. We would also like to thank the editor and an anonymous referee for valuable comments. Financial support from the Norwegian Research Council and the Swedish Research Council is gratefully acknowledged.

E-mail addresses: Moen: espen.moen@bi.no; Rosén: asa.rosen@sofi.su.se

Journal of the European Economic Association December 2006 4(6):1165–1192
© 2006 by the European Economic Association

others. Indeed, empirical studies document that performance pay, broadly interpreted, is common practice.¹ Contractible output eliminates unemployment in the Shapiro-Stiglitz framework, and it is therefore a common perception that rent-based unemployment cannot exist if output is contractible.² The present paper challenges this view and demonstrates that efficiency wages and unemployment may also arise in equilibrium when output is contractible.

Our starting point is to show how the (partial) procurement model of Laffont and Tirole (1993) can be applied in an equilibrium model of the labour market. In the Laffont-Tirole model, the regulator offers a contract to a firm that has private information about both its type and its effort choice. As a result, the optimal contract leaves (information) rents to firms with low production costs. In our model, firms offers wage contracts to workers who have private information about their match-specific productivity and their effort choice. Firms face a trade-off between inducing more effort and conceding larger rents. Because hiring is costly, firms choose a contract such that workers with below maximum match-specific productivity remain employed. The inframarginal workers obtain information rents, and these rents translate into equilibrium unemployment.

Clearly, the equilibrium is inefficient as unemployment is a waste of resources. More interesting is the question whether market contracts differ from those that maximize welfare, given the firms' entry decision and the workers' behavior. We show that the market equilibrium outcome is constrained efficient. This may at first glance seem surprising, as firms do not internalize the rents of their employees when they choose wage contracts. However, worker rents have no social value in equilibrium, as they are offset by a corresponding social cost of unemployment.

Our model allows us to identify three possible factors that lead to more high-powered incentive contracts: greater importance of unobservable effort, lower marginal income taxes, and lesser importance of worker heterogeneity. Typically the increase in incentive power due to such changes coincides with a higher equilibrium unemployment rate. This result suggests that the perceived increase in the use of performance pay (Towers Perrin 1999) may lead to higher equilibrium unemployment.

Efficiency wage theory has also been applied to explain inter-industry wage differentials (Dickens and Katz 1987; Gibbons and Katz 1992). Our efficiency

1. Based on the *National Longitudinal Survey of Youth*, MacLeod and Malcomson (1998) report that 24% of workers in the US in 1990 received performance pay when bonuses and commissions are included. According to Millward et al. (1992), the fraction of workers in the United Kingdom that received some kind of performance-related pay was 34% in 1990. Still, the fraction may be even higher if promotions based on performance and fixed salaries based on past performance are included.

2. See, for example, Cahuc and Zylberberg (2004, ch. 6), Yang (2003), MacLeod and Malcomson (1998), and Weiss (1990, pp. 10–11).

wage model gives the new prediction that industries with a strong relationship between individual performance and wages should pay higher wages.

Dynamic contracting issues are analyzed in different areas of economics. In macroeconomics, there is a literature on the insurance of risk-averse agents when markets are incomplete. In the canonical model, risk-averse agents receive a stochastic income stream. A principal offers partial insurance to the agent, where the coverage of the insurance depends on constraints on the set of feasible contracts. The nature of these constraints varies between papers. In Harris and Holmström (1982) and Krueger and Uhlig (2005), the contract set is limited because the agents cannot commit to stay with the principal after a positive shock, whereas Kocherlakota (1996) assumes two-sided lack of commitment. In Thomas and Worrall (1990), the agents income stream is not observable. In Cole and Kocherlakota (2001), the agents have access to a storing technology, and storage is not observable. An overview of this literature can be found in Ljungqvist and Sargent (2004, chs. 19–20).

Our paper deviates from this literature in several respects. In our model, workers are risk-neutral but have private information at the contracting stage. In the analysis we focus on the interplay between private information, wage contracts, and information rents. In addition, we analyze how information rents translate into equilibrium unemployment.

Another strand of the dynamic contracting literature explores repeated games with random matching between a large number of agents, and includes MacLeod and Malcomson (1989), Ghosh and Ray (1996), and Watson (1999). Common to these papers is the premise that the threat to terminate the relationship serves as a disciplining device. Relationship-specific rents may emerge because it is time-consuming to find a new (high-quality) trading partner. An important issue in several of these papers is how “trust” between the agents, and hence the value of the relationship, builds up over time. Our study differs in a fundamental way from these studies as we allow output to be contractible. Consequently, relationship-specific rents play no role as a disciplining device.

The existing literature on the relationship between performance-based pay and unemployment is rather small. Foster and Wan (1984) study information rents in the labour market and show that with an exogenously given number of firms there may be unemployment. However, their model is not an equilibrium model, as the utility of unemployed workers is exogenous and independent of the number of firms. If entry of firms were introduced in their model, there would be no unemployment in equilibrium. MacLeod and Malcomson (1998) analyze the firms’ choice between performance pay (bonuses) and efficiency wage as means to motivate workers. While performance pay dominates efficiency wages in their symmetric information setting, firms cannot commit to actually paying out the bonus. Thus, if it is easy to replace workers, firms will do so *ex post* rather than paying the bonus. Hence, performance pay can only be used as a motivating

device if the labour market is sufficiently tight, that is, when the unemployment rate is low. In contrast to our model, performance pay is therefore associated with small (no) worker rents in their model.

In a recent paper, Schmitz (2004) analyses a partial model where a single firm extracts worker rents by terminating the employment relationship inefficiently early for low-type workers. He uses this result as a rationale for strict job protection laws. We show that in general equilibrium this may no longer hold. Given our assumptions, worker rents have no social value in labour market equilibrium.

The paper is organized as follows: Section 2 presents the model. We solve for the optimal wage contract in Section 3 and derive the labour market equilibrium outcome in Section 4. Section 5 examines the efficiency and welfare properties of the equilibrium outcome. In Section 6, we derive comparative statics results. In Section 7, we show that the optimal dynamic contract with full commitment coincides with the optimal static contract. Section 8 examines the robustness of the welfare results when workers are risk averse and firm entry is restricted. Section 9 concludes.

2. The Model

The model is set in continuous time. Whereas the measure of jobs is endogenously determined, the measure of workers in the economy is constant and normalized to one. Workers leave the market for exogenous reasons at a rate s and are replaced by new workers who enter the market as unemployed. Unemployed workers search for jobs and firms with vacant jobs search for workers. There is no on-the-job search. Workers and firms discount at a common rate r .

At the hiring stage the expected productivity (given the effort level) is the same for all workers in all firms. We are thus studying a segment of the market in which workers have the same observable characteristics.³ Once employed, the productivity of a given worker also depends on a match-specific productivity term ε . For any worker-firm pair the value of ε is continuously distributed on the interval $[\varepsilon^{\min}, \varepsilon^{\max}]$ with the cumulative distribution function F .⁴ The corresponding density function f has an increasing hazard rate.

3. Observable differences in productivity would not change our results, as the optimal wage contract is contingent on all observable characteristics. The important aspect of the assumption is that workers and firms are symmetrically informed about the worker's productivity at the stage when the worker decides which job to apply for. This is admittedly a strong assumption, as self-selection mechanisms may be empirically important (see for instance Lazear 2000). Self-selection by "informed" workers gives rise to mechanisms that differ substantially from those analyzed in the present paper. Therefore, we abstract from effects of self-selection issues and refer the interested reader to Moen and Rosén (2005).

4. This specification implies that the match-specific productivity of a worker is independent across firms. Thus, a worker's outside option does not depend on his match-specific productivity in that firm. We conjecture that our main results also hold when a worker's productivity term ε is correlated across firms as long as the correlation is not perfect.

The timing is as follows:

- (1) The firm incurs a job creation cost K .
- (2) The firm advertises a wage contract.⁵
- (3) The firm receives job applications from unemployed workers.
- (4) One of the applicants is hired.
- (5) The worker learns his match-specific productivity, ε , and decides whether to stay or not.
- (6) Production starts and continues until the worker leaves the market.

The match-specific term is revealed to the worker after he is hired. Thus, if the worker leaves at this point, the firm incurs the job creation cost K again when a new worker is hired.

The job creation cost K may be given various interpretations. The most direct interpretation is that K denotes the cost of advertising a vacancy. It may also include costs associated with evaluating and testing workers. More generally, K may consist of any costs incurred by the firm (not the worker) before the worker's productivity is revealed, and which are wasted if the worker quits. Thus, K may also include firm-specific training costs during the initial phase of the employment relationship.

As in Laffont and Tirole (1993), we assume that the production function is linear. More specifically, the (flow) value of production of a worker with match-specific productivity ε is

$$y = \bar{y} + \alpha\varepsilon + \gamma e, \quad (1)$$

where e is the worker's effort (unobservable to the firm) and \bar{y} is a constant. The parameter α reflects the importance of the match-specific component for the output and γ the importance of worker effort. Output is contractible, and wage contracts may therefore be made contingent on y . The profit flow of a firm with a worker of match-specific productivity ε is given by

$$\pi = \bar{y} + \alpha\varepsilon + \gamma e - \hat{w}(y), \quad (2)$$

where $\hat{w}(y)$ denotes the wage as a function of output y . A worker's utility flow is given by

$$u = \hat{w}(y) - c(e), \quad (3)$$

where $c(e)$ denotes the effort costs. We assume that $c'(e) > 0$ for $e > 0$, $c'(0) = 0$, $c''(e) \geq 0$, and $c'''(e) \geq 0$. The associated asset values of worker utility and firm profit are $U = u/(r + s)$ and $E[\Pi] = E[\pi]/(r + s)$, respectively.

5. It is possible to show that the equilibrium would be unchanged if the firm proposes the wage contract to the worker after he is hired.

Although we do not impose any restrictions on the shape of $\hat{w}(y)$, we do not allow firms to charge an up-front hiring fee (bonding). This assumption is not innocuous, as such fees would eliminate unemployment.⁶ The absence of bonding may be rationalized in several ways. First, an entrance fee would have to be paid before the worker learns his match-specific productivity. Once the worker knows the match-specific productivity, it is optimal to leave rents to “high-type” workers and bonds would not increase firm profit. Thus, as long as the worker learns ε relatively quickly, implicit bonding, like deferred wage compensation or seniority wages, as in Lazear (1979), do not work. A bond must be interpreted literally as an up-front payment (or at least as a payment that precedes the revelation of ε).

Second, a worker may be reluctant to pay his employer an up-front fee sufficient to eliminate all expected rents. Ritter and Taylor (1994) show that, if firms have private information regarding their bankruptcy probability, a bond can be interpreted as a signal of a high bankruptcy probability. As a result, firms with a low bankruptcy probability leave rents to their employees. More generally, up-front fees may induce firms to fool workers in various ways by hiring and collecting bonds from too many employees, or by prematurely replacing workers (to collect new bonds).⁷ By requiring a low bond or no bond, a firm may signal that it has no such intention.⁸

3. Optimal Contracts

An optimal wage contract maximizes the firm’s expected present discounted profits $E[\Pi]$ subject to the worker’s incentive compatibility (IC) constraint and his individual rationality (IR) constraint. In this section we restrict our attention to the set of static (time-independent) contracts. In Section 7, we show that this contract is also optimal in the wider class of time-dependent contracts, provided that firms can commit not to renegotiate.⁹

The optimal contract is derived using the revelation principle.¹⁰ In order to induce a worker to report his true “type” ε , the following incentive compatibility

6. For a discussion of the so-called bonding critique see, for example, Carmichael (1985, 1990), Dickens et al. (1989), and Akerlof and Katz (1989).

7. In some countries, entrance fees are prohibited. For example, Norwegian legislation does not allow for entrance fees paid to firms. The Contracts Act of 31 May 1918 no 4, §36, in effect deems up-front payments as illegal.

8. Suppose for instance that firms may choose to open a “fake” vacancy at cost $\tilde{K} < K$. A firm with a fake vacancy collects an entrance fee, and then fires the worker. If the workers cannot distinguish between a firm with a fake vacancy and a firm with an ordinary vacancy, the equilibrium entrance fee cannot exceed \tilde{K} , as the market then would be overflowed with fake vacancies. If \tilde{K} is not too high, there would still be (an endogenous amount of) rents in the economy.

9. Once the firm observes the output, it can infer ε and has an incentive to renegotiate. From an ex ante perspective, it is, however, optimal to commit not to renegotiate. We discuss this in some detail in subsequent sections.

10. Because the contract is advertised, and thus constructed before the worker is hired, the revelation principle cannot be interpreted literally.

condition must be satisfied (Appendix A.1):

$$u'(\varepsilon) = c'(e(\varepsilon))\alpha/\gamma. \quad (4)$$

Worker utility is increasing in ε , as long as $e > 0$. This reflects that a high-type worker can produce the same output as a worker of lower type by exerting less effort. More specifically, if a worker's type increases by one unit, the worker can reduce his effort by α/γ units and still obtain the same output, thereby increasing his utility by $c'(e(\varepsilon))\alpha/\gamma$ units. Furthermore, higher effort levels $e(\varepsilon)$ imply larger $u'(\varepsilon)$, and hence larger rents for the workers. Thus, the firm faces a trade-off between incentive provision and rent extraction from the worker.

We denote the worker's flow utility while unemployed by u^0 . Individual rationality requires $u(\varepsilon) \geq u^0$ for any worker who stays with the firm. Let ε_c denote the associated cut-off level, where $u(\varepsilon_c) = u^0$. Inserting $u(\varepsilon) = w(\varepsilon) - c(e(\varepsilon))$ into equation (2) gives $\pi(\varepsilon) = \bar{y} + \alpha\varepsilon + \gamma e(\varepsilon) - c(e(\varepsilon)) - u(\varepsilon)$. The optimal contract thus solves the problem

$$\max_{e(\varepsilon), \varepsilon_c} E[\pi] = \int_{\varepsilon_c}^{\varepsilon^{\max}} [\bar{y} + \alpha\varepsilon + \gamma e(\varepsilon) - c(e(\varepsilon)) - u(\varepsilon)] e^{-rt} dF \quad (5)$$

$$\begin{aligned} \text{subject to: } & u'(\varepsilon) = c'(e(\varepsilon))\alpha/\gamma \\ & u(\varepsilon_c) \geq u^0. \end{aligned}$$

The first-order condition for the optimal e can be written as (Appendix A.1)

$$\gamma - c'(e(\varepsilon)) = \frac{1 - F(\varepsilon)}{f(\varepsilon)} c''(e(\varepsilon))\alpha/\gamma. \quad (6)$$

The intuition for this condition is as follows: Suppose the effort level of a worker with a match-specific productivity ε' increases by one unit. The resulting efficiency gain is $\gamma - c'(e(\varepsilon'))$, and the cost in terms of larger rents for all workers with a match-specific productivity above ε' is $c''(e(\varepsilon'))\alpha/\gamma$ (equation [4]). The likelihood of obtaining a worker of type ε' is reflected in $f(\varepsilon')$, and the measure of workers with a higher match-specific productivity is $1 - F(\varepsilon')$. Equation (6) thus ensures that the gains from effort and rent extraction are balanced at the margin. Given that f has an increasing hazard rate $((1 - F(\varepsilon))/f(\varepsilon))$ decreasing in ε and $c'''(e) \geq 0$, equation (6) implies that $e(\varepsilon)$ is increasing in ε .

The optimal cut-off value solves the equation (Appendix A.1)

$$\bar{y} + \alpha\varepsilon_c + \gamma e(\varepsilon_c) - c(e(\varepsilon_c)) - u^0 = \frac{1 - F(\varepsilon_c)}{f(\varepsilon_c)} c'(e(\varepsilon_c))\alpha/\gamma. \quad (7)$$

This equation uniquely determines ε_c (Appendix A.2). The expected profit of the firm can be written as a function of u^0 , namely, $E[\pi] = E[\pi(u^0)]$, or equivalently $E[\Pi] = E[\Pi(U^0)]$, where U^0 denotes the asset value of an unemployed worker. The function $E[\Pi]$ is strictly decreasing in U^0 .

Let (a, b) denote a linear contract of the form $w = a + by$. It is well known that the optimal nonlinear contract can be represented by a menu $(a(\varepsilon), b(\varepsilon))$ of linear contracts (see for instance Laffont and Tirole 1993). For any b , the worker chooses the effort level such that $c'(e(\varepsilon)) = b\gamma$. Henceforth, we refer to b as the incentive power of the associated linear contract. Using the condition $c'(e(\varepsilon)) = b\gamma$ in equation (6), we obtain

$$b(\varepsilon) = 1 - \frac{1 - F(\varepsilon)}{f(\varepsilon)} c''(e(\varepsilon)) \alpha / \gamma^2. \quad (8)$$

Thus, the optimal wage contract involves distortion for all workers, except the one with maximum match-specific productivity ($b(\varepsilon^{\max}) = 1$).¹¹

In what follows, we are interested in comparing different wage contracts. We call wage contract A *more incentive powered* than wage contract B if $b^A(\varepsilon) \geq b^B(\varepsilon)$ for all ε , with a strict inequality for some ε .

Because $u'(\varepsilon) = \alpha c'(e(\varepsilon)) / \gamma = \alpha b$, the rent for a worker with match-specific productivity ε' is given by

$$\rho(\varepsilon') = \int_{\varepsilon_c}^{\varepsilon'} u'(\varepsilon) d\varepsilon = \int_{\varepsilon_c}^{\varepsilon'} \alpha b(\varepsilon) d\varepsilon. \quad (9)$$

Let $\tilde{F} = F / (1 - F(\varepsilon_c))$ denote the distribution of ε conditional on being above ε_c . The expected rent of a hired worker that remains with the firm is (Appendix A.3)

$$\begin{aligned} E[\rho] &= \int_{\varepsilon_c}^{\varepsilon^{\max}} \int_{\varepsilon_c}^{\varepsilon'} \alpha b(\varepsilon) d\varepsilon d\tilde{F}(\varepsilon') \\ &= \frac{\alpha}{1 - F(\varepsilon_c)} \int_{\varepsilon_c}^{\varepsilon^{\max}} b(\varepsilon) (1 - F(\varepsilon_c) - F(\varepsilon)) d\varepsilon. \end{aligned} \quad (10)$$

The expected income flow of an employed worker can thus be written as $E[u(\varepsilon)] = u^0 + E[\rho]$.

LEMMA 1. *Given that $\varepsilon_c < \varepsilon^{\max}$, employed workers receive a strictly positive expected rent $E[\rho]$.*

11. We have not imposed any restrictions on b . A natural restriction would be that b (or e) are non-negative. This is always the case if $c''(0) = 0$.

Proof. No worker with $\varepsilon \geq \varepsilon_c$ receives a negative rent, as it would violate the individual rationality constraint. From equation (10) it thus follows that $E[\rho]$ is zero if and only if b is zero almost everywhere. However, equation (8) implies that $b(\varepsilon)$ is strictly positive for all ε sufficiently close to ε^{\max} . \square

4. Labour Market Equilibrium

To focus on how efficiency wages lead to unemployment, we assume that there is no time delay in the hiring process.¹²

Free entry of firms ensures that the expected profit $E[\Pi]$ of a firm equals the job creation cost K . Our first equilibrium condition (entry condition) can thus be expressed as

$$E[\Pi(U^0)] = K. \quad (11)$$

Because $E[\Pi]$ is strictly decreasing in U^0 this equation determines the equilibrium value of U^0 uniquely. We denote this equilibrium value by U^{0*} .

Let z denote the utility flow of unemployed workers and p the transition rate from unemployment to employment in steady state. The transition rate p equals the rate at which workers are hired multiplied with $1 - F(\varepsilon_c)$. The relationship between U^0 and p is then given by

$$(r + s)U^0 = z + p(W - U^0), \quad (12)$$

where W is the expected discounted income when employed. Let R denote the asset value of the expected rents ($R = E[\rho]/(r + s)$). As R is by definition equal to $W - U^0$ we can rewrite equation (12) as

$$(r + s)U^0 = z + pR. \quad (13)$$

The equilibrium in the labour market is defined as a pair (p, U^0) satisfying equations (11) and (13).

In the absence of unemployment, workers find a job immediately, which implies that p is infinite. However, this leads to a contradiction, as U^0 defined by equation (13) then goes to infinity and thus exceeds U^{0*} as defined by equation (11).

PROPOSITION 1. *The equilibrium unemployment rate is strictly positive.*

Proof. Given U^{0*} , equation (7) determines ε_c^* . Furthermore, $\varepsilon_c^* < \varepsilon^{\max}$, otherwise the firm would not recoup K . It then follows from Lemma 1 that $E[\rho]$ is

12. In an earlier version, we show how the labour market equilibrium can be derived as the limit equilibrium of an urn-ball model when the frictions go to zero (Moen and Rosén 2003).

strictly positive. But then it follows from equation (13) that U^0 goes to infinity if p does. Thus, p is finite. \square

As being unemployed is the outside option for a worker, rents imply that it is strictly better (in expected terms) to be employed than to be unemployed. However, this is inconsistent with full employment.

The transition rate to employment is such that the rents are dissipated. Inserting $U^0 = U^{0*}$ into equation (13) and rearranging gives

$$p = \frac{(r + s)U^{0*} - z}{R}. \quad (14)$$

Let x denote the unemployment rate in the economy. Using the fact that $(p + s)x = s$ holds in steady state yields

$$x = \frac{s}{r + s} \frac{R}{U^{0*} - Z + \frac{s}{r+s}R}, \quad (15)$$

where $Z = z/(r + s)$ is the asset value of staying unemployed forever.

5. Efficiency

Obviously, the equilibrium outcome is not first-best, as this requires full employment and $c'(e) = \gamma$. (Almost) full employment can be obtained by an arbitrarily high negative unemployment benefit, whereas the efficient level of effort can be approximated by a negative income tax schedule on labour income. (We discuss the impact of taxes on the wage contract in the next section.) For reasons outside our model, these policy recommendations are unlikely to be taken seriously by any government.

In our view a more interesting question is whether the wage contracts chosen by the firms are socially optimal, given the behavior of the workers and the entry decisions of firms. Or putting it differently, what wage contract should a social planner choose given that all other decisions are still taken by the market participants?

We call the equilibrium wage contracts chosen by the firms *constrained efficient* if they maximize welfare subject to the workers' incentive compatibility and individual rationality constraints and to the entry condition (equation (11)).

The social planner maximizes overall production less the job creation and effort costs. Let $V(\Phi)$ denote the expected discounted production value of a worker-firm pair net of the effort costs as a function of the wage contract Φ . This contract also specifies a cut-off level ε_c . For each formed worker-firm pair, the vacancy creation cost K is incurred $1/[1 - F(\varepsilon_c^*(\Phi))]$ times. Finally, assume that the social value of the utility flow of an unemployed worker is equal to z .

The transition rate to employment is determined by equation (14) and can thus be written as a function of Φ . Hence, the planner's objective function is

$$S(\Phi) = \int_0^\infty \left[zx + xp(\Phi) \left[V(\Phi) - \frac{K}{1 - F(\varepsilon_c(\Phi))} \right] \right] e^{-rt} dF. \quad (16)$$

The social planner maximizes S given the constraint that $\dot{x} = s - (p + s)x$.

Within a search context, Acemoglu and Shimer (1999), Moen and Rosén (2004), and Pissarides (2000) show that the social planner chooses the vacancy rate to maximize the welfare of the unemployed workers. An analogous result holds here.

LEMMA 2. *The social planner's problem is equivalent to maximizing the unemployed workers' expected discounted utility.*

See Appendix A.4 for the proof.

We are now in the position to show that the equilibrium wage contract is constrained efficient. The equilibrium contract Φ^* satisfies the following two conditions:

- (1) $\max_{\Phi} E[\Pi(\Phi)]$ subject to: $u'(\varepsilon) = c'(e(\varepsilon))\alpha/\gamma$ and $u(\varepsilon_c) \geq u^0$.
- (2) $E[\Pi(\Phi)] = K$.

The constrained-efficient contract Φ' solves the "dual" maximization problem (because maximizing U^0 is equivalent with maximizing u^0):

- (A) $\max_{\Phi} u^0(\Phi)$ subject to: $u'(\varepsilon) = c'(e(\varepsilon))\alpha/\gamma$ and $u(\varepsilon_c) \geq u^0$.
- (B) $E[\Pi(\Phi)] = K$.

PROPOSITION 2. *The equilibrium wage contract is constrained efficient.*

Proof. Suppose the proposition does not hold. Then $U^0(\Phi') > U^0(\Phi^*)$. However, because $E[\Pi(\Phi')] = K$ when $U = U^0(\Phi')$, it follows that there exists a contract Φ'' such that $E[\Pi(\Phi'')] > K$ when $U = U^0(\Phi^*)$. But then Φ^* cannot be an equilibrium contract, a contradiction. \square

The firms choose the contracts to balance rent extraction and worker effort. Increasing the incentive power of the contract b for some types gives rise to a positive externality, as this tends to increase worker rents. One may therefore expect that the incentive contracts are not high-powered enough (too low values of b). However, this is not correct. The rents of employed workers have no social value in equilibrium, because the unemployment rate is determined so as to dissipate all rents. That is, larger worker rents lead to a higher unemployment rate, leaving the asset value of an unemployed worker constant.

To explore this argument, suppose that firms can choose among production technologies that give rise to different levels of worker rents. For instance, output with different production technologies may depend to a varying extent on unobserved effort. *Ceteris paribus*, technologies that are more sensitive to effort lead to more high-powered incentives and higher worker rents than technologies for which effort is less important. Similarly, production technologies may differ in their sensitivity to worker-firm specific productivity differences.

COROLLARY 1. *Firms choose the constrained efficient production technology.*

See Appendix A.5 for the proof.

When firms choose between different technologies, worker rents do not enter their objective function. However, as worker rents are dissipated in labour market equilibrium they do not enter the objective function of the social planner either. Thus, the firms' choice of technology is constrained efficient. By contrast, Acemoglu and Newman (2002) find within the context of a shirking model that firms' incentives to invest in monitoring are too strong. (Monitoring can be interpreted as a technology choice.) As mentioned in the Introduction, our finding also contrasts with the results in Schmitz (2004). He analyzes a partial equilibrium model in which a single firm extracts worker rents by terminating employment of low-type workers inefficiently early. In our model, firms have the correct incentives to reduce worker rents as worker rents have no social value.

Proposition 2 relies on the assumption that the income z reflects the social value of being unemployed. Thus, z may reflect the value of leisure, home production, or, alternatively, wages in a secondary labour market. An interesting result follows from equation (11) and from the fact that $E[\Pi]$ is independent of z .

COROLLARY 2. *Social welfare is independent of z , regardless of whether z reflects the social value of being unemployed, unemployment benefits, or wages in the secondary sector.*

A higher z makes it more time-consuming to dissipate rents and thus increases unemployment (or employment in the inferior secondary sector) exactly by so much that unemployed workers obtain the same utility level.

Suppose that z (partly) consists of government transfers (unemployment benefits), and hence does not reflect the social value of being unemployed. Corollary 2 implies that the unemployment benefits are a waste of resources, as they do not influence the well-being of unemployed workers. Consequently, the unemployment rate does not influence the equilibrium wage contract or the equilibrium cut-off value either. However, the unemployment rate increases with the unemployment benefit, as it takes more time to dissipate the rents associated with employment.

COROLLARY 3. *With unemployment benefits the equilibrium wage contract is more high-powered than the constrained efficient wage contract.*

See Appendix A.5 for the proof.

A positive unemployment benefit makes the government bear part of the burden associated with being unemployed. This is not taken into account when the incentive contracts are determined. Hence, equilibrium contracts result in an unemployment rate that is too high. By contrast, taxes on labour income tend to reduce the equilibrium incentive power of the contracts below its constrained optimal level. We return to this point shortly.

The fact that the wage contract is constrained efficient does not mean that other policy measures cannot improve efficiency. We have already argued that a negative unemployment benefit may improve welfare. A more interesting policy measure is to subsidize job creation, namely, subsidize K . Suppose the cut-off decision is trivial ($\varepsilon^* = \varepsilon^{\min}$). Then subsidizing K does not influence the wage contract, but increases employment. By assumption, the social value of employment (expected net productivity less value of leisure) exceeds K , otherwise the economy would collapse. Hence, subsidizing K improves welfare.

With an endogenous cut-off level, the argument becomes more involved, as a subsidy of K reduces the cut-off level. Because the cut-off level is optimally set initially, the envelope theorem ensures that a small subsidy still increases welfare. However, for sufficiently large subsidies this will not be the case.¹³

6. Determinants of the Unemployment Rate

In this section we analyze the effects of changes in the wage contract on the equilibrium unemployment rate. These effects may depend on which of the structural parameters in the model triggers the change in the wage contract.

Because we are not able to characterize the effects of a change in the cut-off level ε_c on the worker rents in the general case, we subsequently assume that the cut-off level is below ε^{\min} . Thus, the firms retain all workers, which may in fact be a good approximation of their actual behavior. Given that a firm has spent resources on hiring and possibly training a worker, the likelihood of dismissal may be fairly low by the time the match-specific productivity component is revealed.

The first thing to note is that an increase in the constant term \bar{y} does not influence the optimal contract. However, such a change increases u^0 , and it follows from equation (15) that the unemployment rate falls. More generally, the effects of any parameter shift can be decomposed into two factors: the direct effect of

13. It will never be optimal to set the subsidy equal to K in order to eliminate unemployment. In this case, the cut-off level will be set equal to ε^{\max} , and the hiring cost K will be incurred an infinite number of times.

changes in equilibrium worker rents (keeping u^0 constant), and the income effect associated with changes in u^0 .

We focus on the direct effect of parameter changes. Thus, when changing one parameter, we simultaneously adjust the constant \bar{y} such that u^0 stays constant.¹⁴

As we have seen, $u^0(b)$ is maximized at $b^*(\varepsilon)$. By the envelope theorem u^0 is approximately constant for wage contracts close to $b^*(\varepsilon)$. From equations (10) and (15) it follows that stronger incentives yield, ceteris paribus, more rents to the workers and thereby a higher unemployment rate.

Reduced Relative Importance of Private Information. A reduction in α has two opposing effects on worker rents. On the one hand, a lower α leads to lower expected rents for a given wage contract (equation [10]). On the other hand, a reduction in α leads to an increase in b for all ε types (equation [8]), which tends to increase expected rents. In order to obtain clear-cut results, we assume that $c(e)$ is quadratic. We define the average value of b as $\bar{b} = \int_{\varepsilon^{\min}}^{\varepsilon^{\max}} b(\varepsilon) / (\varepsilon^{\max} - \varepsilon^{\min}) d\varepsilon$.

PROPOSITION 3. *A reduction in the importance of ex post heterogeneities leads to more high-powered wage contracts and more (less) unemployment if $\bar{b} < 1/2$ ($\bar{b} > 1/2$).*

See Appendix A.6 for the proof.

The average value \bar{b} corresponds to the expected value of $b(\varepsilon)$ only when ε is uniformly distributed. Because $b'(\varepsilon) > 0$, \bar{b} may be greater than 1/2 even if $E[b(\varepsilon)]$ is less than 1/2 if most of the probability mass is located in the lower part of the distribution. Finally, if $c'''(e)$ is strictly positive, the responsiveness of the optimal contract to changes in α is reduced, making it more likely that a reduction in α leads to less unemployment than indicated in the proposition.

Increased Importance of Unobservable Effort. An increase in importance of unobservable effort γ has two opposing effects on worker rents. On the one hand, an increase in γ tends to increase the right-hand side of equation (8), and thus $b^*(\varepsilon)$ for a given effort level. On the other hand, an increase in γ tends to increase e for a given b . As $c'''(e) \geq 0$ this tends to reduce incentives, because rent extraction becomes more important. While the net effect is therefore indeterminate, mild assumptions on the cost-function ensure that the first effect dominates. More

14. In a dynamic setting, discrete changes in the technology at a given point in time can be consistent with continuous welfare u^0 . Suppose τ_1 and τ_2 are two alternative technologies. The equilibrium selects the technology that maximizes u^0 . Suppose τ_1 dominates τ_2 initially but that τ_2 dominates from some point in time \tilde{t} . At $t = \tilde{t}$, u^0 is still continuous in time, although technology is discontinuous at \tilde{t} .

specifically, b increases in γ , if $c''(e)/c'(e)$ is non-increasing. All polynomials of the form x^n as well as the exponential function satisfy this restriction.¹⁵

PROPOSITION 4. *An increase in the importance of unobservable effort leads to more high-powered wage contracts and more unemployment.*

See Appendix A.7 for the proof.

If unobserved effort becomes relatively more important, firms provide their workers with stronger incentives, as incentive provision becomes more important relative to rent extraction. As a result, the expected rent associated with employment increases, as does unemployment.

Reduction in marginal taxes. Suppose the income tax T is given by

$$T = tw - A, \quad (17)$$

where A is a constant and t the marginal tax rate. To study the direct effect of an decrease in t , we simultaneously reduce A (here equivalent to reducing \bar{y}) to keep u^0 unchanged. We also maintain the mild restriction that $c''(e)/c'(e)$ is decreasing in e .

PROPOSITION 5. *A reduction in the marginal tax rate leads to more high-powered wage contracts and more unemployment.*

See Appendix A.8 for the proof.

Marginal taxes tend to reduce the incentives to undertake effort, as private agents pay the entire effort cost while the government receives a fraction of the gain. If marginal taxes are reduced, effort becomes more valuable for the worker-firm pair, and firms provide their workers with stronger incentives. As a result, the expected rent associated with employment increases, and thus also unemployment. By contrast, the structure of taxes need not affect unemployment in the Shapiro-Stiglitz model (Pissarides 1998).

7. Time-dependent Contracts

In this section we analyze the optimal contract when we allow for time-dependent contracts. In a contractual setting similar to ours, Baron and Besanko (1984) show that the optimal dynamic contract repeats the optimal static contract, provided that the firm can commit not to renegotiate.¹⁶ In Appendix A.9, we prove that this

15. As will be clear from the proof, it is actually sufficient to assume that $c''(e)/c'(e)^2$ is decreasing.

16. For an instructive proof see Fudenberg and Tirole (1991), p. 299 ff.

result also holds in our model: The optimal time-dependent contract is identical to the contract derived previously. Here, we provide the intuition as to why this is the case.

In the standard shirking model of Shapiro-Stiglitz, the firm may improve the incentives to exert effort by a simple time-dependent contract with upward-sloping wage-tenure profiles (see, e.g., Lazear 1979 and Akerlof and Katz 1989). In our model, the firms do not benefit from such deferred compensation. It is not the threat of being fired that motivates workers in our model. Output is contractible, and a low effort level is immediately “punished” by a lower wage. It is the incentive power of the wage contract that gives the worker an incentive to provide effort. The firm can implement any desired effort level by choosing an appropriate wage contract. For instance, if workers had no private information regarding ε (ε has a degenerate distribution), the firms would simply implement the first best effort level without leaving any rents to the workers.

However, if, as in our model, the workers have private information regarding their productivity, providing incentives is costly to the firm, as it yields information rents to the workers. (Note, though, that the marginal worker receives no rents.) Deferred compensation does not reduce this information rent to high-productivity worker types, because it does not reduce the rent they can obtain by pretending to be of a low type. Furthermore, deferred compensation does not influence the individual rationality constraint at the hiring stage. It may loosen the individual rationality constraint over time, but this has no value to the firm as the worker’s outside option is time independent.

The time-independent contract implies that the effort level stays constant over time. Suppose to the contrary that the firm offered the worker a contract with an effort level $e_1(\varepsilon)$ for the first t periods, and then effort level $e_2(\varepsilon)$, with $e_1(\varepsilon) \neq e_2(\varepsilon)$ for some ε . The firm can always improve by smoothing the effort levels. From the incentive compatibility constraint (4) it follows that the flow value of the information rent is convex in e (because $c'''(e) > 0$). At the same time, the profit flow in (5) is concave in e . Smoothing effort levels therefore increases the profit flow and reduces the information rents, and thereby surely increases profits.

The result that the optimal contract is time-independent stands in contrast to some other equilibrium models with private information. Ghosh and Ray (1996) and Watson (1999) analyze trade between agents with private information about their type. The agents play a prisoner’s dilemma game with private information about their gains from collaborating /defecting and with endogenous chosen inputs (effort levels). In order to obtain cooperation, there must be rents associated with continuing the relationship, for instance as in Shapiro and Stiglitz (1984) and MacLeod and Malcolmson (1989). It is shown that effort levels increase as the agents learn about their opponent’s type. Our model differs from these models in many respects. Most important, we assume that output is contractible,

whereas they do not. Our conjecture is that in both these papers efficiency would be obtained with time invariant contracts if output were contractible.

Albuquerque and Hopenhayn (2004) analyze optimal dynamic lending contracts. They show that an investor is reluctant to lend an entrepreneur too much money in each period, as he may take the money and execute his outside option. The project therefore only gradually approaches its optimal size. Whereas investors provide funds up-front, workers are compensated after effort is exerted. Hence, the incentive problems in Albuquerque and Hopenhayn do not arise.

The result that the optimal dynamic contract repeats the optimal static contract crucially depends on the assumption that the firm can commit not to renegotiate. Once the firm learns ε , both parties benefit from renegotiating the contract and agree on the first-best effort level. Workers realize that the average required output level increases due to renegotiation and hence require a higher information rent to reveal their type truthfully. As shown in Laffont and Tirole (1993), a separating renegotiation-proof contract gives rise to higher expected rents than in the case where the firms can commit not to renegotiate. This indicates that the unemployment rate in our model is higher with renegotiation. However, Laffont and Tirole also prove that the optimal contract with renegotiation always implies pooling among the least productive types, which reduces the expected information rents relative to the separating equilibrium. Thus, the effect of renegotiation on the equilibrium unemployment rate is not obvious.

8. Robustness of the Efficiency Result

In this section we discuss the effects of worker risk aversion and restrictions on firm entry for the constrained efficiency result (Proposition 2).

8.1. Risk Aversion

We now assume that the utility flow of a worker can be written as

$$u = v(d) - c(e), \quad (18)$$

where v is a concave function of the consumption flow d . In line with other papers in this literature (Ljungqvist and Sargent 2004, ch. 19), we do not allow workers to borrow or save. Hence, the consumption flow is equal to w when the worker is employed and z when the worker is unemployed.¹⁷ This assumption ensures that the optimal contract remains time-independent. (The proof presented in Appendix A.9 still holds.)

17. As the worker's income profile is (weakly) increasing over time (an employed worker never returns to unemployment), the constraint on saving is unlikely to bind.

An extension of the standard model of Laffont and Tirole (1993) to include risk averse agents can be found in Laffont and Rochet (1998). The technical apparatus developed for risk-neutral workers still applies. In particular, the incentive compatibility constraint is identical, that is, $u'(\varepsilon) = c'(e(\varepsilon))\alpha/\gamma$ (see Appendix A.10). As before, the gains a high type derives from mimicking a low type are associated with lower effort costs.

Solving for w in equation (18) (with $d = w$) gives $w(\varepsilon) = v^{-1}(u(\varepsilon) + c(e(\varepsilon)))$, compared with $u(\varepsilon) + c(e(\varepsilon))$ with risk neutrality. This complicates the optimization problem. The optimal contract solves the problem

$$\max E[\pi] = \max_{e(\varepsilon), \varepsilon_c} \int_{\varepsilon_c}^{\varepsilon^{\max}} [\bar{y} + \alpha\varepsilon + \gamma e(\varepsilon) - v^{-1}(u(\varepsilon) + c(e(\varepsilon)))] e^{-rt} dF \quad (19)$$

subject to:
$$\begin{aligned} u'(\varepsilon) &= c'(e(\varepsilon))\alpha/\gamma \\ u(\varepsilon_c) &\geq u^0. \end{aligned}$$

Laffont and Rochet (1998) show that the optimal contract is separating provided that the agent is not too risk averse. We do not derive the optimal contract here but assume that it exists and denote it by Φ^r . Note that, as long as $\gamma > 0$, the optimal contract induces at least a positive measure of workers at the top of the distribution to exert effort. Hence, there is an expected rent associated with employment and therefore unemployment in equilibrium.

Furthermore, Φ^r is also constrained socially optimal, as it maximizes the expected discounted utility U^0 of an unemployed worker. The proof follows exactly the same lines as the proof of Proposition 2. Suppose there exists a different contract Φ' which satisfied the workers' IC and IR constraints, gives workers a higher U^0 , and at the same time satisfies the zero-profit condition (11). Then $U^0(\Phi') > U^0(\Phi^r)$. However, because $E[\Pi(\Phi')] = K$ when $U = U^0(\Phi')$, there exists a contract Φ'' such that $E[\Pi(\Phi'')] > K$ when $U = U^0(\Phi^r)$. However, then Φ^r cannot be an equilibrium contract—a contradiction.

8.2. Restrictions on Entry

We now analyze the efficiency of the wage contracts when there are restrictions on firm entry. We assume that there is a fixed number \bar{N} of jobs in the economy.¹⁸ The hiring cost K remains the same, and the cut-off is trivial, namely, $\varepsilon_c = \varepsilon^{\min}$.

Let N^* denote the equilibrium employment rate with free entry, and let R^* and U^{0*} denote the associated expected discounted rent and the expected discounted

18. An alternative interpretation of the fixed number of jobs would be that the demand for the final product is restricted.

utility for an unemployed worker, respectively. Obviously, the fixed number of jobs does not influence the employment rate if $\bar{N} > N^*$. Firms only find it profitable to hire N^* workers anyway. We therefore restrict attention to the case $\bar{N} < N^*$.

We know that the firms' choice of contract, and thus the expected worker rents, are independent of U^0 . Thus, we have¹⁹

$$(r + s)U^0 = z + pR^* = z + s \frac{\bar{N}}{1 - \bar{N}} R^*,$$

which is increasing in \bar{N} .

In this case, the equilibrium contract is not efficient. The planner can increase output net of effort cost in each firm by increasing the incentive power of the contract. As long as the firms make positive expected profits, or, equivalently, as long as $U^0 < U^{0*}$, this does not reduce the number of jobs in the economy. Thus, increasing the incentive power increases welfare. An increase in worker rent above R^* has no social cost, as it does not affect the number of available jobs.

Let \bar{R} denote the highest expected rent the planner can give the worker and still ensure non-negative expected profits for the firms. If \bar{R} is larger than the expected rents associated with first best effort, the constrained efficient wage contract gives all employees full incentives. Otherwise, the constrained efficient contract maximizes the expected output per worker, subject to the constraint that the firms break even (expected profit equals the hiring cost K).

9. Conclusion

We show that efficiency-wage based unemployment may arise in a model with contractible output if workers have private information about their match-specific productivity. Worker heterogeneity at the firm level and the absence of entry fees imply that wage contracts trade off incentive provision and rent extraction. Moreover, the incentive power of the equilibrium wage contract is constrained efficient in the absence of unemployment benefits. At first glance, this may seem surprising because firms do not internalize the value of their employee's rents. Wage contracts are nonetheless constrained efficient because equilibrium rents lead to higher unemployment, which reduces welfare.

The unemployment level is determined by the amount of rents left to the worker through the optimal wage contract. Expected worker rents, and thereby

19. Inflow into employment has to equal outflow. Thus, $p(1 - \bar{N}) = s\bar{N}$, and $p = s\bar{N}/(1 - \bar{N})$.

also the unemployment rate, are large if nonobservable effort is relatively important, the marginal tax rate is lower, or heterogeneity among workers is neither too large nor too small.

Appendix

A.1. Deriving Equations (4), (6), and (7)

When pretending to be type $\tilde{\varepsilon}$, a type- ε worker obtains a (flow) utility

$$\tilde{u}(\varepsilon, \tilde{\varepsilon}) = w(\tilde{\varepsilon}) - c\left(e(\tilde{\varepsilon}) + \frac{\alpha}{\gamma}(\tilde{\varepsilon} - \varepsilon)\right),$$

where $w(\tilde{\varepsilon})$ and $e(\tilde{\varepsilon})$ denote the wage and the effort level as a function of the reported type $\tilde{\varepsilon}$. The indirect flow utility can be written as $u(\varepsilon) = \max_{\tilde{\varepsilon}} \tilde{u}(\varepsilon, \tilde{\varepsilon})$. The incentive compatibility constraint requires that $\varepsilon = \arg \max_{\tilde{\varepsilon}} \tilde{u}(\varepsilon, \tilde{\varepsilon})$, or from the envelope theorem equivalently that

$$u'(\varepsilon) = \left. \frac{\partial \tilde{u}(\varepsilon, \tilde{\varepsilon})}{\partial \varepsilon} \right|_{\tilde{\varepsilon}=\varepsilon},$$

namely, that

$$u'(\varepsilon) = c'(e(\varepsilon))\alpha/\gamma.$$

The current-value Hamiltonian associated with the firm's maximization problem can be written as

$$H = [\bar{y} + \alpha\varepsilon + \gamma e(\varepsilon) - c(e(\varepsilon)) - u]f(\varepsilon) + \lambda c'(e(\varepsilon))\alpha/\gamma,$$

where λ is the adjoint function. For a given cut-off ε_c , the first-order conditions for a maximum are

$$\begin{aligned} (\gamma - c'(e(\varepsilon)))f(\varepsilon) + \lambda c''(e(\varepsilon))\alpha/\gamma &= 0, \\ \dot{\lambda}(\varepsilon) &= f(\varepsilon). \end{aligned}$$

Because there are no terminal conditions at ε^{\max} it follows that $\lambda(\varepsilon^{\max}) = 0$, and $\lambda(\varepsilon) = -(1 - F(\varepsilon))$. The first-order condition for the optimal e can thus be written as

$$\gamma - c'(e(\varepsilon)) = \frac{1 - F(\varepsilon)}{f(\varepsilon)} c''(e(\varepsilon))\alpha/\gamma.$$

The optimal cut-off value solves the equation $H(\varepsilon_s) = 0$, or

$$\bar{y} + \alpha\varepsilon + \gamma e(\varepsilon) - c(e(\varepsilon)) - u^0 = \frac{1 - F(\varepsilon_c)}{f(\varepsilon_c)} c'(e(\varepsilon))\alpha/\gamma.$$

A.2. Unique Cut-off Level

Define

$$\Psi(\varepsilon) = \bar{y} + \alpha\varepsilon + \gamma e(\varepsilon) - c(e(\varepsilon)) - u^0 - \frac{1 - F(\varepsilon)}{f(\varepsilon)} c'(e(\varepsilon))\alpha/\gamma.$$

Equation (7) determines a unique ε_c if and only if $\Psi(\varepsilon) = 0$ is uniquely defined.

$$\begin{aligned} \frac{d\Psi(\varepsilon)}{d\varepsilon} &= \alpha + \gamma \frac{de}{d\varepsilon} - c'(e(\varepsilon)) \frac{de}{d\varepsilon} - c'(e(\varepsilon))\alpha/\gamma \frac{d}{d\varepsilon} \frac{1 - F(\varepsilon)}{f(\varepsilon)} \\ &\quad - \frac{1 - F(\varepsilon)}{f(\varepsilon)} c''(e(\varepsilon))\alpha/\gamma \frac{de(\varepsilon)}{d\varepsilon}. \end{aligned}$$

Inserting

$$-\frac{1 - F(\varepsilon)}{f(\varepsilon)} c''(e(\varepsilon))\alpha/\gamma = c'(e(\varepsilon)) - \gamma$$

(equation [6]) yields

$$\frac{d\Psi(\varepsilon)}{d\varepsilon} = \alpha - c'(e(\varepsilon))\alpha/\gamma \frac{d}{d\varepsilon} \frac{1 - F(\varepsilon)}{f(\varepsilon)} > 0.$$

Hence, $\Psi(\varepsilon) = 0$ is uniquely defined.

A.3. Deriving Equation (10)

The integral

$$E[\rho] = \int_{\varepsilon_c}^{\varepsilon^{\max}} \int_{\varepsilon_c}^{\varepsilon'} \alpha b(\varepsilon) d\varepsilon d\tilde{F}(\varepsilon')$$

can be simplified using integration by parts. We use that

$$\int_a^b u(x)v'(x)dx = \left| u(x)v(x) - \int_a^b u'(x)v(x)dx \right.$$

Let $v = 1 - \tilde{F}$, $v' = -d\tilde{F}$, $u = \int_{\varepsilon_c}^{\varepsilon'} \alpha b(\varepsilon)d\varepsilon$ and $u' = \alpha b(\varepsilon)$. This allows us to rewrite the above integral as

$$\begin{aligned} E[\rho] &= - \left|_{\varepsilon_c}^{\varepsilon^{\max}} (1 - \tilde{F}) \int_{\varepsilon_c}^{\varepsilon'} \alpha b(\varepsilon)d\varepsilon + \int_{\varepsilon_c}^{\varepsilon^{\max}} \alpha b(\varepsilon)(1 - \tilde{F})d\varepsilon \right. \\ &= \int_{\varepsilon_c}^{\varepsilon^{\max}} \alpha b(\varepsilon)(1 - \tilde{F})d\varepsilon \\ &= \frac{\alpha}{1 - F(\varepsilon_c)} \int_{\varepsilon_c}^{\varepsilon^{\max}} b(\varepsilon)(1 - F(\varepsilon_c) - F(\varepsilon)) d\varepsilon. \end{aligned}$$

A.4. Proof of Lemma 2

The current-value Hamiltonian associated with the social planner's maximization problem can be written as

$$H^c = zx + xp(\Phi) \left[V(\Phi) - \frac{K}{1 - F(\varepsilon_c(\Phi))} \right] + \lambda[s - (p(\Phi) + s)x]$$

where the only state variable is x . The first-order conditions are given by

$$r\lambda = \frac{\partial}{\partial x} H^c = z + p(\Phi) \left[V(\Phi) - \frac{K}{1 - F(\varepsilon_c(\Phi))} \right] - \lambda(p(\Phi) + s) \quad (\text{A.1})$$

and

$$\begin{aligned} \Phi &= \arg \max_{\Phi} H^c & (\text{A.2}) \\ &= \max_{\Phi} [zx + xp(\Phi) \left[V(\Phi) - \frac{K}{1 - F(\varepsilon_c(\Phi))} \right] + \lambda[s - (p(\Phi) + s)x]]. \end{aligned}$$

Equation (A.1) implies that

$$(r + s)\lambda = z + p(\Phi) \left[V(\Phi) - \frac{K}{1 - F(\varepsilon_c(\Phi))} - \lambda \right]. \quad (\text{A.3})$$

In the maximization problem (A.2), the state variable x and the adjoint variable λ are regarded as constant, and the maximization problem can equivalently be expressed as

$$\max_{\Phi} p(\Phi) \left[V(\Phi) - \frac{K}{1 - F(\varepsilon_c(\Phi))} - \lambda \right].$$

This problem is equivalent to maximizing λ as defined by equation (A.3).

We want to show that this is equivalent to maximizing U^0 defined by equation (12), given equation (14) and the entry condition $E[\Pi] = K$. The entry condition implies that the expected profit of a firm with a worker who remains in the firm is equal to $K/(1 - F(\varepsilon_c))$. The rest of the surplus accrues to the worker. Thus, $W(\Phi) = V(\Phi) - K/(1 - F(\varepsilon_c))$. Inserted into equation (12), we find that

$$(r + s)U^0 = z + p(\Phi) \left[V(\Phi) - \frac{K}{1 - F(\varepsilon_c(\Phi))} - U^0 \right]. \quad (\text{A.4})$$

The expression for U^0 in equation (A.4) is formally identical to the expression for λ in equation (A.3). We have already shown that the social planner maximizes λ given that $V = V(\Phi)$ and $p = p(\Phi)$. This is equivalent to maximizing U^0 given the same two constraints.

A.5. Proof of Welfare Corollaries

Proof of Corollary 1. Let τ denote the technology, and let $V = V(\tau, \Phi)$ denote the associated expected production value net of effort costs, and let $R(\tau, \Phi)$ denote the associated expected rent that the contract leaves to the worker. The firm chooses the value of τ , τ^* that maximizes $E[\Pi(\tau)] = (1 - F(\varepsilon_c(\tau, \Phi)))[V(\tau, \Phi) - R(\tau, \Phi) - U^0]$. Moreover, $E[\Pi(\tau^*)] = K$ in equilibrium. Let τ' denote the constrained efficient value of τ and $\Phi' = \Phi(\tau')$ the associated optimal contract (which coincides with the equilibrium contract for this technology). Free entry implies that

$$U^{0'} = V(\tau', \Phi') - R(\tau', \Phi') - \frac{K}{(1 - F(\tau', \varepsilon_c(\Phi')))}.$$

It follows that τ^* is constrained efficient. Otherwise, $U^{0'} > U^{0*}$ and the firms in the market could improve by choosing τ' and the wage schedule given by Φ' minus an arbitrarily small constant. \square

Proof of Corollary 3. Whereas worker rents have zero social value with zero unemployment benefits, they have strictly negative values with positive unemployment benefits. The planner's maximization problem is thus identical to the maximization problem (5), with $u(\varepsilon)$ replaced by $ku(\varepsilon)$ in the integrand, where $k > 1$ is a constant. The first-order condition is thus given by

$$b(\varepsilon) = 1 - k \frac{1 - F(\varepsilon)}{f(\varepsilon)} c''(e) \alpha / \gamma^2.$$

As b decreases in k this completes the proof. \square

A.6. Proof of Proposition 3

From equations (8) and (10) it follows that

$$E[\rho] = \int_{\varepsilon^{\min}}^{\varepsilon^{\max}} \alpha \left[1 - \frac{1 - F(\varepsilon)}{f(\varepsilon)} \frac{\alpha c''(e(\varepsilon))}{\gamma^2} \right] (1 - F(\varepsilon)) d\varepsilon.$$

Given $c'''(e) = 0$, the derivative with respect to α is

$$\begin{aligned} \frac{dE[\rho]}{d\alpha} &= \int_{\varepsilon^{\min}}^{\varepsilon^{\max}} \left[1 - \frac{1 - F(\varepsilon)}{f(\varepsilon)} \frac{\alpha c''(e(\varepsilon))}{\gamma^2} - \alpha \frac{1 - F(\varepsilon)}{f(\varepsilon)} \frac{c''(e)}{\gamma^2} \right] (1 - F(\varepsilon)) d\varepsilon \\ &= \int_{\varepsilon^{\min}}^{\varepsilon^{\max}} [2b(\varepsilon) - 1] (1 - F(\varepsilon)) d\varepsilon. \end{aligned}$$

Hence, a reduction in α increases (reduces) worker rents if $\bar{b} \leq (>) 1/2$, and equation (15) implies that the unemployment rate increases (decreases) if $\bar{b} \leq (>) 1/2$.

A.7. Proof of Proposition 4

For any given b , $c'(e) = \gamma b$ and hence $c''(e)/\gamma^2 = b^2 c''(e)/c'(e)^2$. We can thus write equation (8) as

$$b(\varepsilon) = h(\varepsilon, e(\gamma), b(\gamma)),$$

where

$$h(\varepsilon, e(\gamma), b(\gamma)) = 1 - \alpha \frac{1 - F(\varepsilon)}{f(\varepsilon)} \frac{b(\gamma)^2 c''(e(\gamma))}{c'(e(\gamma))^2},$$

which yields

$$\frac{db}{d\gamma} = \frac{\frac{\partial h}{\partial e} \frac{\partial e}{\partial \gamma}}{1 - \frac{\partial h}{\partial b}}.$$

Because $c''(e)/c'(e)^2$ decreases in e and because e increases in γ , $(\partial h/\partial e)(\partial e/\partial \gamma) > 0$. Furthermore, $db/d\gamma > 0$ as $\partial h/\partial b < 0$. From equation (10) it follows that $E[\rho]$ and thereby also R increases in γ . Equation (15) then implies that the unemployment rate increases in γ .

A.8. Proof of Proposition 5

Worker utility is given by $u(\varepsilon) = w(1 - t) + A - c(e)$. Firm profit can thus be written as

$$\pi(\varepsilon) = \bar{y} + \alpha \varepsilon + \gamma e - \frac{u(\varepsilon) - A + c(e)}{1 - t}.$$

The truth-telling condition is unaffected by taxes and remains $u'(\varepsilon) = c'(e)\alpha/\gamma$. Hence, the first-order conditions for the optimal contract are

$$\begin{aligned} \frac{c'(e)}{1 - t} &= \gamma + \frac{1}{f(\varepsilon)} \lambda c''(e)\alpha/\gamma, \\ \dot{\lambda} &= \frac{f(\varepsilon)}{1 - t}. \end{aligned}$$

Integrating λ , inserting, and rearranging gives

$$c'(e) = \gamma(1-t) - \frac{1-F(\varepsilon)}{f(\varepsilon)}c''(e)\alpha/\gamma.$$

In a menu of linear contracts, the corresponding incentive parameter b is such that $c'(e) = \gamma(1-t)b$, or that $b = c'(e)/[\gamma(1-t)]$. It thus follows that

$$b(\varepsilon) = 1 - \frac{1}{1-t} \frac{1-F(\varepsilon)}{f(\varepsilon)}c''(e)\alpha/\gamma^2. \quad (\text{A.5})$$

For any given b , $c'(e) = \gamma b(1-t)$, and hence $\frac{c''(e)}{(1-t)\gamma^2} = \frac{bc''(e)}{\gamma c'(e)}$. We can thus write equation (A.5) as $b(\varepsilon) = h(\varepsilon, e(t), b(t))$, where

$$h(\varepsilon, e(t), b(t)) = 1 - \alpha \frac{1-F(\varepsilon)}{f(\varepsilon)} \frac{b(t)c''(e(t))}{\gamma c'(e(t))}.$$

Differentiating $b(\varepsilon)$ with respect to t yields

$$\frac{db}{dt} = \frac{\frac{\partial h}{\partial e} \frac{\partial e}{\partial t}}{1 - \frac{\partial h}{\partial b}}.$$

Because e decreases in t and $c''(e)/c'(e)$ decreases in e , $(\partial h/\partial e)(\partial e/\partial t) < 0$. Furthermore, $db/dt < 0$, as $\partial h/\partial b < 0$ which proves the first part of the proposition. Because $\rho(\varepsilon') = \int_{\varepsilon_c}^{\varepsilon'} u'(\varepsilon)d\varepsilon$ and $u'(\varepsilon) = c'(e)\alpha/\gamma = \alpha(1-t)b$, a reduction in t increases $\rho(\varepsilon)$ for all ε and thereby also R . However, it then follows from equation (15) that a decrease in t increases the unemployment rate.

A.9. Optimal Time-dependent Contracts

We want to show that the optimal time-independent contract is optimal within the larger class of time-dependent contracts as well. A similar proof can be found in Fudenberg and Tirole (1991, p. 299). To simplify the proof and avoid uninteresting technicalities we assume that time is discrete. We consider first the case where the cut-off level is ε^{\min} . This will be modified at the end.

The revelation principle still holds. Hence, it is sufficient to study the set of contracts that map the worker's (reported) type into a sequence of wages and effort levels $\{w_t(\varepsilon), e_t(\varepsilon)\}_{t=0}^{\infty}$, where t denotes the tenure of the worker in question.

Let $\pi_t(\varepsilon, e_t) = \bar{y} + \alpha\varepsilon + \gamma e_t(\varepsilon) - w_t(\varepsilon)$. The expected discounted profit to the firm is given by

$$\Pi = E^\varepsilon \sum_{t=0}^{\infty} \pi_t(\varepsilon, e_t)\delta^t,$$

where $\delta = (1 - s)/(1 + r)$ is the discount factor, including the exit rate of the worker. The expected discounted utility of a worker of type ε who announce type $\tilde{\varepsilon}$ is given by

$$\bar{W}(\varepsilon, \tilde{\varepsilon}) = \sum_{t=0}^{\infty} \left[w_t(\tilde{\varepsilon}) - c \left(e(\tilde{\varepsilon}) - \alpha \frac{\varepsilon - \tilde{\varepsilon}}{\gamma} \right) \right] \delta^t.$$

Incentive compatibility requires that $\varepsilon = \arg \max_{\tilde{\varepsilon}} \bar{W}(\varepsilon, \tilde{\varepsilon})$. Let $W(\varepsilon) \equiv \bar{W}(\varepsilon, \varepsilon)$. The optimal dynamic contract solves

$$\max_{\{w_t(\varepsilon), e_t(\varepsilon)\}_{t=0}^{\infty}} E^{\varepsilon} \sum_{t=0}^{\infty} \pi_t(\varepsilon, e_t) \delta^t,$$

subject to:

Incentive compatibility. $\varepsilon = \arg \max_{\tilde{\varepsilon}} \bar{W}(\varepsilon, \tilde{\varepsilon})$.

Individual rationality. $W(\varepsilon) \geq U^0$ for all ε . (This constraint binds only for ε^{\min} .)

Note that the participation constraint regards the expected discounted utility of all future periods. It does not require that the employed worker's utility flow is higher than the utility flow of unemployed workers in all periods. Thus, deferred compensation with increasing wage-tenure profile is allowed for.

Let $C^d = \{w_t^d(\varepsilon), e_t^d(\varepsilon)\}_{t=0}^{\infty}$ denote an optimal contract within the larger set of time-dependent contracts, and let $C^* = \{w^*(\varepsilon), e^*(\varepsilon)\}_{t=0}^{\infty}$ denote the time-independent contract. We want to show that C^d is equivalent to C^* , in the sense that it implements the same effort level in each period, the same discounted expected profit to the firm, and the same expected discounted rents to the workers.

Suppose $C^d \neq C^*$. Then C^d cannot implement a time independent effort level, as this contract by definition would be dominated by the optimal static contract C^* . Suppose therefore that C^d does not implement a time independent effort level. We will show that this leads to a contradiction.

To this end, consider the time-independent stochastic mechanism C^{dS} , defined as follows: The contract $(w_t^d(\varepsilon), e_t^d(\varepsilon))$ is implemented with probability $\delta^t/(1 - \delta)$. By definition, this contract is both incentive compatible and satisfies the individual rationality constraint. Furthermore, it yields a higher expected profit to the firm than the static contract $(w^*(\varepsilon), e^*(\varepsilon))$, because C^d dominates C^* , and thus contradicts the optimality of the latter mechanism in the class of time-independent contracts. It thus follows that $C^d = C^*$.

Finally, the same argument holds for any given cut-off value ε_c , and hence the optimal cut-off level with time-dependent contracts, must be equal to the optimal cut-off level with time-independent contracts.

A.10. Incentive Compatibility Constraint with Risk Aversion

When pretending to be type $\tilde{\varepsilon}$ a ε type worker obtains a (flow) utility

$$\tilde{u}(\varepsilon, \tilde{\varepsilon}) = v(w(\tilde{\varepsilon})) - c(e(\tilde{\varepsilon})) + \frac{\alpha}{\gamma}(\tilde{\varepsilon} - \varepsilon)$$

where $w(\tilde{\varepsilon})$ and $e(\tilde{\varepsilon})$ denote the wage and the effort level as a function of the reported type $\tilde{\varepsilon}$. The indirect flow utility can be written as $u(\varepsilon) = \max_{\tilde{\varepsilon}} \tilde{u}(\varepsilon, \tilde{\varepsilon})$. The incentive compatibility constraint requires that $\varepsilon = \arg \max_{\tilde{\varepsilon}} \tilde{u}(\varepsilon, \tilde{\varepsilon})$, or from the envelope theorem equivalently that

$$u'(\varepsilon) = \left. \frac{\partial \tilde{u}(\varepsilon, \tilde{\varepsilon})}{\partial \varepsilon} \right|_{\tilde{\varepsilon}=\varepsilon},$$

namely, that

$$u'(\varepsilon) = c'(e(\varepsilon))\alpha/\gamma.$$

References

- Acemoglu, Daron, and Andrew F. Newman (2002). "The Labor Market and Corporate Structure." *European Economic Review*, 46, 1733–1756.
- Acemoglu, Daron, and Robert Shimer (1999). "Holdups and Efficiency with Search Frictions." *International Economic Review*, 40, 827–849.
- Akerlof, George A. (1982). "Labor Contracts as Partial Gift Exchange." *Quarterly Journal of Economics*, 97, 543–569.
- Akerlof, George A., and Lawrence L. Katz (1989). "Workers' Trust Funds and the Logic of Wage Profiles." *Quarterly Journal of Economics*, 104, 525–536.
- Albuquerque, Rui, and Hugo A. Hopenhayn (2004). "Optimal Lending Contracts and Firm Dynamics." *Review of Economic Studies*, 71, 285–315.
- Baron, David P., and David A. Besanko (1984). "Regulation and Information in a Continuing Relationship." *Information Economics and Policy*, 1, 447–470.
- Cahuc, Pierre, and André Zylberberg (2004). *Labor Economics*. MIT Press.
- Carmichael, Lorne H. (1985). "Can Unemployment be Involuntary? Comment." *American Economic Review*, 75, 1213–1214.
- Carmichael, Lorne H. (1990). "Efficiency Wage Models of Unemployment: One View." *Economic Inquiry*, 28, 269–295.
- Cole, Harold L., and Narayana R. Kocherlakota (2001). "Efficient Allocations with Hidden Income and Hidden Storage." *Review of Economic Studies*, 68, 523–542.
- Dickens, William T., and Lawrence F. Katz (1987). "Inter-Industry Wage Differences and Theories of Wage Differentials." NBER Working Paper No. 2271.
- Dickens William T., Lawrence F. Katz, Kevin Lang, and Lawrence H. Summers (1989). "Employee Crime and the Monitoring Puzzle." *Journal of Labor Economics*, 7, 331–347.
- Foster, James E., and Henry Y. Wan, Jr. (1984). "Involuntary Unemployment as a Principal Agent Equilibrium." *American Economic Review*, 74, 476–484.
- Fudenberg, Drew, and Jean Tirole (1991). *Game Theory*. MIT Press.
- Ghosh, Parikshit, and Debraj Ray (1996). "Cooperation in Community Interaction Without Information Flows." *Review of Economic Studies*, 63, 491–519.

- Gibbons, Robert, and Lawrence F. Katz (1992). "Does Unmeasured Ability Explain Inter-Industry Wage Differentials." *Review of Economic Studies*, 59, 515–535.
- Harris, Milton, and Bengt Holmström (1982). "A Theory of Wage Dynamics." *Review of Economic Studies*, 49, 315–333.
- Kocherlakota, Narayana R. (1996). "Implications of Efficient Risk Sharing without Commitment." *Review of Economic Studies*, 63, 595–609.
- Krueger, Dirk, and Harald Uhlig (2005). "Competitive Risk Sharing Contracts with One-Sided Commitment." Working paper, University of Pennsylvania.
- Laffont, Jean-Jacques, and Jean-Charles Rochet (1998). "Regulation of a Risk Averse Firm." *Games and Economic Behavior*, 25, 149–173.
- Laffont, Jean-Jacques, and Jean Tirole (1993). *A Theory of Incentives in Procurement and Regulation*. MIT Press.
- Lazear, Edward P. (1979). "Why is There Mandatory Retirement?" *Journal of Political Economy*, 87, 1261–1264.
- Lazear, Edward P. (2000). "Performance Pay and Productivity." *American Economic Review*, 90, 1346–1361.
- Ljungqvist, Lars, and Tomas J. Sargent (2004). *Recursive Macroeconomic Theory*. MIT Press.
- MacLeod, W. Bently, and James M. Malcomson (1989). "Implicit Contracts, Incentive Compatibility, and Involuntary Unemployment." *Econometrica*, 57, 447–480.
- MacLeod, W. Bently, and James M. Malcomson (1998). "Motivation and Markets." *American Economic Review*, 88, 388–411.
- Millward, Neil, Mark Stevens, David Smart, and W. R. Hawes (1992). *Workplace Industrial Relations in Transition: The ED/ESRC/PSI/ACAS Surveys*. Dartmouth.
- Moen, Espen R., and Åsa Rosén (2003). "Equilibrium Incentive Contracts and Efficiency Wages." CEPR Working Paper No. 3790.
- Moen, Espen R., and Åsa Rosén (2004). "Does Poaching Distort Training." *Review of Economic Studies*, 71, 1143–1162.
- Moen, Espen R., and Åsa Rosén (2005). "Performance Pay and Adverse Selection." *Scandinavian Journal of Economics*, 107, 279–298.
- Pissarides, Christopher A. (1998). "The Impact of Employment Tax Cuts on Unemployment and Wages: The Role of Unemployment Benefits and Tax Structure." *European Economic Review*, 42, 155–183.
- Pissarides, Christopher A. (2000). *Equilibrium Unemployment Theory*. MIT Press.
- Ritter, Joseph A., and Lowell J. Taylor (1994). "Workers as Creditors: Performance Bonds and Efficiency Wages." *American Economic Review*, 84, 694–704.
- Salop, Steven C. (1979). "A Model of the Natural Rate of Unemployment." *American Economic Review*, 69, 117–125.
- Schmitz, Patrick W. (2004). "Job Protection Laws and Agency Problems under Asymmetric Information." *European Economic Review*, 48, 1027–1046.
- Shapiro, Carl, and Joseph E. Stiglitz (1984). "Equilibrium Unemployment as a Worker Discipline Device." *American Economic Review*, 74, 433–444.
- Thomas, Jonathan, and Tim S. Worrall (1990). "Income Fluctuations and Asymmetric Information: An Example of Repeated Principal-Agent Problem." *Journal of Economic Theory*, 51, 367–390.
- Towers Perrin (1999). "Worldwide Total Remuneration." Report.
- Watson, Joel (1999). "Starting Small and Renegotiation." *Journal of Economic Theory*, 85, 52–90.
- Weiss, Andrew W. (1980). "Job Queues and Layoffs in Labor Markets with Flexible Wages." *Journal of Political Economy*, 88, 525–538.
- Weiss, Andrew W. (1990). *Efficiency Wages: Models of Unemployment, Layoffs and Wage Dispersion*. Clarendon Press, Oxford.
- Yang, Huanxing (2003). "Efficiency Wages and Performance Pay." Working paper, University of Pennsylvania.