Optimal Unemployment Insurance with Monitoring and Sanctions [*]

Jan Boone[†] Peter Fredriksson[‡] Bertil Holmlund[§] Jan van Ours[¶]

April 24, 2006

Abstract

This paper analyzes the design of optimal unemployment insurance in a search equilibrium framework where search effort among the unemployed is not perfectly observable. We examine to what extent the optimal policy involves monitoring of search effort and benefit sanctions if observed search is deemed insufficient. We find that introducing monitoring and sanctions represents a welfare improvement for reasonable estimates of monitoring costs; this conclusion holds both relative to a system featuring indefinite payments of benefits and a system with a time limit on unemployment benefit receipt.

JEL-classification: J64, J65, J68

Keywords: Unemployment insurance, Job search, Benefit sanctions

1. Introduction

It is generally accepted that public provision of unemployment insurance (UI) is socially desirable in a world with risk averse individuals. However, it is also well established that the provision of UI does not come without adverse incentive effects. For example, more generous UI benefits are likely to reduce search effort and raise wage pressure, thus causing some increase

---

[*] We gratefully acknowledge programming assistance from Per Engström as well as useful comments from anonymous referees, David Grubb, Åsa Rosén and seminar participants at IZA, Bonn, Stockholm University, University of Copenhagen, University of Gothenburg, and University of Cyprus. Fredriksson and Holmlund acknowledges the financial support from the Swedish Council for Working Life and Social Research. Boone acknowledges the financial support from NWO (grantnumbers 016.025.024 and 453.03.606).

[†] CentER, TILEC, Tilburg University, ENCORE, UvA, IZA and CEPR. Postal address: Box 90153, 5000 LE Tilburg, The Netherlands. E-mail: j.boone@uvt.nl.

[‡] Department of Economics, Uppsala University and IFAU. Postal address: Box 513, SE-751 20 Uppsala, Sweden. E-mail: peter.fredriksson@nek.uu.se.

[§] Department of Economics, Uppsala University and CESifo. Postal address: Box 513, SE-751 20 Uppsala, Sweden. E-mail: bertil.holmlund@nek.uu.se.

[¶] Department of Economics, Tilburg University, CentER, IZA and CEPR. Postal address: Box 90153, 5000 LE Tilburg, The Netherlands. E-mail: vanours@uvt.nl.

in unemployment. The problem facing policy makers is thus to strike an optimal balance between the insurance benefits on the one hand, and the adverse incentive effects on the other hand. This problem has been the subject of several recent papers. Our paper contributes to this literature by recognizing that the government may condition benefit payments on (imperfectly) observed search effort. This leads us to an analysis of optimal UI design in a search equilibrium framework where the government has several policy instruments at its disposal, including the benefit level, the rate at which search effort is monitored, and the magnitude of the sanction in case search effort is regarded as insufficient. We find that a system with monitoring and sanctions represents a welfare improvement relative to other alternatives for reasonable estimates of the monitoring costs. In particular, the monitoring and sanction system leads to higher welfare than a system with time limits.

Our results on the desirability of monitoring can be contrasted with a well-known result that dates back to Becker's (1968) celebrated paper on optimal crime deterrence. In Becker's analysis (as in ours), monitoring is costly because resources have to be spent on detecting crime (violations of search requirements). Punishment, in the form of a fine (sanction), goes without cost since it involves a transfer of money from one individual to others. To deter crime the expected fine, i.e., the probability of being caught times the fine, should be big enough. By raising the fine, monitoring costs can be reduced without affecting incentives for crime. However, Becker's analysis presupposes risk neutral agents. When agents are risk averse and there are errors in the monitoring technology, Becker's result need not hold. If the monitoring technology is plagued by Type II errors, some complying individuals are sanctioned and these individuals will be subjected to substantial welfare losses when fines are high.[1]

Shavell and Weiss (1979) presented a seminal analysis of optimal sequencing of benefit payments over the spell of unemployment. The key result was that the benefit level should decline monotonically over the unemployment spell, because such a profile involves stronger incentives to search. Recently a number of papers have extended the analysis of Shavell and Weiss. One strand of the literature adds additional policy instruments; Hopenhayn and Nicolini (1997) is a case in point. Another strand of the literature (e.g. Cahuc and Lehmann, 2000, and Fredriksson and Holmlund, 2001) takes account of firm behavior and allows for endogenous wage determination. Endogenous wages is potentially important since a declining benefit profile can raise wage pressure. Wage pressure may rise because it is the value of unemployment upon unemployment entry that enters the worker's outside option. The analysis in Fredriksson and Holmlund (2001), however, suggests that there is still a case for having a declining profile of benefit payments.

The contributions reviewed above, and most of the other literature on optimal UI, do not consider that the government can make the receipt of benefits dependent on the unemployed worker's search effort. As documented by Grubb (2001), existing UI systems condition benefit payments on performance criteria such as "availability for work" and "active job search". These

---

[1] This squares with the conclusion in Polinsky and Shavell (1979). They conclude that risk aversion weakens the case for the Beckerian policy prescription. Furthermore, they note that the possibility of making Type II errors reinforces this conclusion. See Garoupa (1997) and Polinsky and Shavell (2000) for recent surveys of the economic theory of law enforcement.

criteria are enforced by some degree of monitoring of the benefit claimants. The requirements for job search show substantial variations across countries.

Failure to meet search requirements may result in a benefit sanction, i.e., a temporary or permanent cut in benefits. A typical duration of sanctions for a first refusal of a suitable job offer is two to three months. Observed sanction rates – the total number of sanctions over a year relative to the stock of beneficiaries – also vary substantially across countries. For example, sanctions due to insufficient search hovered around 30 % in the US in the late 1990s, whereas other countries (Germany, Denmark, Norway) appear to have undertaken no sanctions related to search inactivity; see Grubb (2001) for further details.

Recent empirical work has shed light on the effects of changes in search requirements and monitoring of job search. The arguably most convincing evidence is based on randomized experiments undertaken in the US. The "treatments" in these experiments involved the number of employer contacts, the required documentation and the frequency of verification.[2] These studies indicate that either more intensive monitoring or more demanding search requirements tend to reduce the length of benefit claims. Recent non-experimental evidence from the Netherlands and Switzerland also suggest that the imposition of sanctions substantially raises the transition rate to employment (Abbring et al., 2005; Lalive et al., 2005; van den Berg et al., 2004). Our reading of the bulk of the evidence is that more intensive monitoring and more stringent search requirements do matter for search activity and transitions out of unemployment.[3]

Very few papers have conducted a theoretical analysis of monitoring and sanctions in the context of UI. The study most closely related to what we do in the present paper is Boone and van Ours (2006).[4] Boone and van Ours used numerical analysis to illustrate that monitoring and sanctions can represent a welfare improvement over a system featuring indefinite payments of UI benefits.[5] We extend their work in three important ways: (i) We contrast a monitoring and sanctions system to more general alternatives (i.e. a system featuring time limits); (ii) we derive analytical results on the optimality of monitoring and sanctions; and (iii) we perform a numerical analysis to illustrate the characteristics of the optimal system with monitoring and sanctions.

The analysis of monitoring and sanctions is clearly related to the analysis of the optimal sequencing of UI benefits. Indeed, one can think of the declining profile of benefit payments as an indirect "sanction" on deficient job search. The defining characteristic of a monitoring and

---

[2] See OECD (2001), Johnson and Klepinger (1994), Benus et al (1997) and Black et al (2003).

[3] There is at least one study, Van den Berg and Van der Klaauw (2001), that fails to confirm that more intensive monitoring affects transitions out of unemployment. The authors conjecture that the result may reflect that more stringent monitoring of formal search induces a substitution away from informal search channels.

[4] Boadway and Cuff (1999) also deal with monitoring of job search, although in a very different framework. In their model, unemployment and the wage rate are not affected by the amount of transfers given to the unemployed.

[5] Boone and van Ours (2006) is mainly written as a complement to the empirical literature on benefit sanctions. Usually only the "ex post" effect – that is how individuals react to the imposition of a benefit sanction – can be estimated. However, the "ex ante" effect – i.e. the increased incentive to search harder to avoid getting a benefit sanction – is difficult to establish empirically. One of the main conclusions is that even if empirical research would indicate that the ex post effect is small, it may still be worthwhile to introduce a system of benefit sanctions because of the ex ante effect.

sanction system, however, is that the risk of being sanctioned depends *directly* on search activity. This feature can have substantial implications for policy prescriptions. Let us illustrate this point by considering a world with risk aversion and a (finite) arrival rate of job offers. In this situation, a system with a time limit on UI benefit receipt can never have “Beckerian properties”. The reason is that some workers will be penalized as time passes. However, a Becker-type solution is a distinct possibility when the risk of being penalized depends directly on search. If the monitoring technology is perfect, the government can implement the optimal search intensity by threatening to impose the maximal sanction.[6] A disadvantage of the monitoring and sanctions system is that resources are needed for monitoring. The argument in favor of monitoring and sanctions depends crucially on the costs associated with such a system. One of the contributions of our paper is to derive an upper bound on the marginal cost of monitoring where the monitoring system dominates the pure time limits for all marginal costs below this upper bound.

In this paper we thus extend the previous literature by offering a normative analysis of a benefit system with costly monitoring and sanctions. The basic model features two benefit levels which can be thought of as unemployment insurance (UI) and unemployment assistance (UA), respectively. Workers who receive UI are monitored at a certain rate and, with some probability, exposed to a benefit sanction. The probability of being sanctioned depends on the worker’s search effort and the precision at which search effort can be observed by the UI provider. Sanctioned workers receive UA, they are not monitored, and they need to become reemployed before they are entitled to UI. We are concerned with the characteristics of the optimal benefit system when there are four available policy instruments: the level of benefits in UI and UA (the difference between the two representing the sanction), the rate at which the unemployed worker entitled to UI is monitored, and the precision of the monitoring technology that determines how the agent’s search effort affect the probability of a sanction.[7]

The next section of the paper presents the basic model. Section 3 derives some analytical results concerning the properties of the optimal UI system. In section 4, we turn to a numerical analysis of the optimal benefit system. Section 5 concludes.

2. The Model

2.1. *The Labour Market*

We consider an economy with a fixed labour force, which is normalized to unity. Workers are either employed or unemployed and have infinite horizons. Time is continuous. An employed worker is separated from his job at an exogenous Poisson rate ϕ. Upon entering unemployment,

---

[6] This a viable strategy with risk aversion since nobody will be sanctioned in equilibrium with a perfect monitoring technology.

[7] Note that because Boone and van Ours (2006) are mainly interested in the difference between “ex ante" and “ex post" effects, they do not derive the optimal (welfare maximizing) sanction system and only investigate the effects of introducing benefit sanctions in an environment where unemployment benefits are fixed and last for ever.

the worker is immediately eligible for UI benefits.

Recipients of UI benefits are monitored with respect to their search behavior. If they fail to meet certain search requirements, they are exposed to a benefit withdrawal (a sanction). We assume that the sanction lasts for the remainder of the unemployment spell. At every instant, there are thus two groups of unemployed workers: *eligible* workers who receive benefits and *sanctioned* workers who have been exposed to a benefit withdrawal.

Let α^j, $j = e, s$, denote the exit rate from unemployment to employment for an *e*ligible and a *s*anctioned worker, respectively. The exit rates differ between the two groups to the extent that their search effort differ. Let s^j, $j = e, s$, denote search effort. The effective number of searchers in the economy is then given as $S = s^e u^e + s^s u^s$, where u^j is the number of unemployed in category j.

The matching function is of the usual constant returns to scale variety: $H = H(S, v)$, where v is the number of vacancies. Let $\theta \equiv v/S$ denote labour market tightness. The probability per unit time that individual i escapes unemployment state j is then obtained as $\alpha_i^j \equiv s_i^j H(S, v)/S = s_i^j \alpha(\theta)$. Also, $\alpha(\theta) = H(S, v)/S = H(1, \theta)$ and hence $\alpha'(\theta) > 0$; the tighter the labour market, the easier to find a job. Firms fill vacancies at the rate $q(\theta) = H(S, v)/v = H(1/\theta, 1)$, and thus $q'(\theta) < 0$; the tighter the labour market, the more difficult to fill a vacancy. By constant returns to scale, we also have $\alpha(\theta) = \theta q(\theta)$.

While unemployed and receiving UI benefits, an unemployed agent is monitored at rate μ. We think of monitoring as random inspections of the worker's search activity. Given monitoring, there is some probability that the observed search effort does not meet the search requirement, in which case the worker is sanctioned. Let $\pi(s^e)$ denote the probability of being sanctioned upon inspection of search effort, implying that UI recipients lose entitlement at the rate $\mu\pi(s^e)$.

Having defined the relevant transition rates, we can formulate the steady state flow equilibrium relationships of the labour market:

$$\phi n = \alpha^e u^e + \alpha^s u^s \tag{1}$$

$$\alpha^s u^s = \mu\pi u^e \tag{2}$$

where $n = 1 - u^e - u^s$ denotes total employment in the economy. The first equation pertains to employment whereas the second equation pertains to the state of unemployment with a sanction. Now we can use (1) and (2) to solve for employment:

$$n = \frac{\lambda\left(\alpha^e + \mu\pi\right)}{\phi + \lambda(\alpha^e + \mu\pi)} \tag{3}$$

where $\lambda \equiv u^e/(u^e + u^s) = \alpha^s/(\alpha^s + \mu\pi))$ is the ratio of eligible unemployment to total unemployment.

2.2. *Monitoring and Sanctions*

Let us make the monitoring and sanctions technology explicit. We choose a reduced form specification which allows us to have as special cases indefinite payments of UI benefits ($\mu = 0$), finite duration of UI benefit receipt ($\mu > 0$ and $\pi(s_i^e) = 1$), and a monitoring and sanctions technology. In particular, we assume that the probability of being sanctioned upon inspection depends linearly on search: $\pi(s_i^e) = 1 - \sigma s_i^e$. Proposition 2 below gives conditions under which $\sigma > 0$ is optimal. Further, we require that $\pi(s_i^e) \geq 0$ for all $s_i^e \in [0,1]$, which, in turn, implies that $\sigma \in [0,1]$.

The parameter σ measures to which extent the sanction probability depends on an agent's own search effort. One way to interpret σ is that it indexes the precision of the inspection technology. For instance, $\sigma = 0$ corresponds to the situation where it is determined by lottery if the agent has searched to rule or not; therefore everyone who is monitored is sanctioned irrespective of search intensity. Alternatively, $\sigma = 0$ can be seen as a UI system with a time limit, as in Fredriksson and Holmlund (2001). If, on the other hand, σ is strictly positive the agent's search effort matters for the sanction probability. The higher is σ, the higher the precision with which an agent's search effort is observed and rewarded.[8]

Whereas $\sigma = 0$ gives little direct incentive to search, it is an inexpensive system to operate. This is due to the fact that there are no inspections of agents' search effort. On the other hand, $\sigma > 0$ gives a direct incentive to search but also implies that more monitoring officials are needed in order to inspect agents' search intensities. So the monitoring cost per monitored agent is increasing in σ.

More precisely, we assume that the cost of running the monitoring and sanctioning system, C, is given by:

$$C = c(\sigma)\,\mu u^e w. \tag{4}$$

The cost of running this system is increasing in the number of monitored individuals (μu^e). The rate of increase is determined by $c(\sigma) \geq 0$. This cost depends on the precision of the inspection technology with $c'(\sigma) \geq 0$ and $c(0) = 0$. We think of the inspection of search as a labour intensive activity and, therefore, the monitoring cost is proportional to the aggregate wage w.

2.3. *Worker Behavior*

The employed worker's (indirect) instantaneous utility is determined by his wage, w. The unemployed worker receives unemployment benefits, B, as long as he is eligible. When sanctioned, he receives Z. We show in proposition 1 below that $B > Z$. We assume that workers do not have access to a capital market, so consumption equals income at each instant.

We take the utility functions to be strictly concave in income and leisure. The unemployed worker's instantaneous utility is decreasing in search effort, since search reduces time available for leisure. The utility function for the eligible unemployed worker is $v(B, s_i^e)$ and for the sanctioned worker it is $v(Z, s_i^s)$. The employed worker's utility is given by $v(w_i, \overline{h})$, where $\overline{h}$ denotes hours of work; we take $\overline{h}$ as exogenously fixed. We assume here that the individuals

---

[8] From a more general point of view, it is possible to derive this technology from "first principles" with the aid of a few assumptions. We present this derivation in Appendix C. The derivation is based on workers being heterogenous, a complication we have not introduced at this stage. We briefly consider worker heterogeneity in section 4.3.4; see also Appendix B.

have identical utility functions. However, as one of the sensitivity checks we will also consider an economy where workers are subject to stochastic shocks to the value of leisure.

Let r denote the subjective rate of time preference and let U^j and E be the expected present values of being unemployed, $j = e, s$, and employed, respectively. The value functions can then be written as:

$$\begin{aligned} rU_i^e &= \max_{s_i^e} \left\{ \upsilon(B, s_i^e) + s_i^e \alpha(\theta) \left(E - U_i^e\right) - \mu\pi\left(s_i^e\right)\left(U_i^e - U^s\right)\right\}, & (5) \\ rU_i^s &= \max_{s_i^s} \left\{ \upsilon(Z, s_i^s) + s_i^s \alpha(\theta)(E - U_i^s)\right\}, & (6) \\ rE_i &= \upsilon(w_i, \overline{h}) - \phi(E_i - U^e). & (7) \end{aligned}$$

The unemployed worker chooses search effort to maximize rU_i^j. The first-order conditions are given by:

$$\upsilon_s(B, s^e) + \alpha(\theta)(E - U^e) - \mu\pi_s\left(s^e\right)\left(U^e - U^s\right) = 0 \qquad (8)$$

$$\upsilon_s(Z, s^s) + \alpha(\theta)(E - U^s) = 0 \qquad (9)$$

where partial derivatives with respect to search effort are indicated by subscript s. In these expressions we have imposed symmetry, i.e., we have made use of the fact that workers are identical and choose the same search effort. The first-order conditions convey the usual message:[9] at the optimum, the marginal cost of search should equal the marginal benefits. The marginal cost is captured by foregone leisure, i.e., $\upsilon_s(B, s^e)$ and $\upsilon_s(Z, s^s)$. The marginal benefit involves the gain in utility associated with a transition to employment, i.e., $\alpha(\theta)(E - U^j)$, $j = e, s$. For the eligible worker, there is an additional benefit of more intensive search, as revealed by the third term on the right-hand side of (8). More intensive search reduces the probability of being sanctioned, thus prolonging the expected duration of benefit payments. This does not imply, however, that eligible workers necessarily search harder than sanctioned workers. The effect pulling in the opposite direction is $B > Z$: sanctioned workers gain more from finding a job than eligible workers since $E - U^s > E - U^e$ holds in equilibrium. Which effect dominates depends on the parameters of the UI system.

In our baseline model we assume that the instantaneous utility functions take the form:[10]

$$\upsilon(m, l) = \ln m + \Gamma(l), \quad m = \{w, B, Z\}, \; l = \{1 - \bar{h}, 1 - s^e, 1 - s^s\}$$

where m denotes (real) income, which depends on the worker's labour market position and $\Gamma(l)$ represents the value of leisure with $\Gamma'(l) > 0$ and $\Gamma''(l) < 0$. The employed worker receives a wage w; the eligible unemployed worker receives unemployment insurance, B; and an unemployed worker who has been exposed to a sanction receives unemployment assistance, Z. Note that, given the monitoring technology, this functional form implies that $Z = 0$ can never be optimal from a social point of view. Hence, unemployment assistance will be strictly positive in any welfare maximizing system.

---

[9] The second-order conditions for a maximum are fulfilled by the concavity of $\upsilon(\cdot)$ and the linearity of π.

[10] In the sensitivity analysis we consider alternative specifications of the utility functions.

2.4. *Firms and Wage Bargaining*

Assume that government expenditure on benefits and monitoring is financed by a proportional payroll tax paid by firms. Labour productivity is constant and denoted y. The cost of holding a vacancy is ky, with $k > 0$. Let V denote the present value of a vacant job and J the present value of an occupied job. The value functions are of the usual form:

$$rV = -ky + q(\theta)(J - V) \tag{10}$$

$$rJ = y - w(1 + t) - \phi(J - V) \tag{11}$$

where t is the proportional payroll tax rate. With free entry of new vacancies, $V = 0$, we obtain the wage cost as proportional to the marginal product of labour, i.e.,

$$w\,(1 + t) = [1 - (r + \phi)k/q(\theta)]y. \tag{12}$$

Defining $\omega \equiv w(1 + t)$ and writing the right-hand side of this equation as $d\,(\theta)\,y$, we refer to $\omega = d(\theta)y$ as the zero profit condition, with $d'(\theta) < 0$.

The outcome of the Nash bargain

$$\max_{w_i} \left[E(w_i) - U^e\right]^{\beta} \left[J(w_i) - V\right]^{1-\beta}, \quad \beta \in (0, 1)$$

is a relationship of the form:

$$\frac{E - U^e}{wv_w} = \frac{\beta}{1 - \beta}\frac{J}{\omega} \tag{13}$$

where $V = 0$ and symmetry have been imposed. The Nash bargain implies a wage-setting relationship, i.e., a relationship between bargained wages and labour market tightness. We assume that the government fixes the replacement rates in this economy. Hence $Z = zw$ and $B = bw$ where z and b are policy parameters.[11] Finally, the relative size of the benefit sanction is denoted by p, i.e. p satisfies $z = (1 - p)b$.

2.5. *Equilibrium*

Our assumptions imply that the model has a convenient recursive structure; the model in Fredriksson and Holmlund (2001) has a similar structure.[12] The zero-profit condition and the wage-setting relationship determine θ and ω. To see this, note that with free entry of vacancies we have $J = ky/q(\theta)$ and $\omega = d(\theta)y$, which imply that the right-hand side of (13) is increasing in θ but independent of s^j. The left-hand side of (13) is a function of θ but independent of w given our chosen utility function and the fact that income during unemployment is proportional to the aggregate wage. Moreover, $E - U^e$ is invariant to derivative changes of s^j in the neighborhood of

---

[11] The replacement rates are defined with respect to the economy-wide average wage which the individual employee perceives to be independent of his wage demands; therefore U^e is perceived by an individual worker to be independent of w_i in the Nash bargain: $\partial U^e/\partial w_i = 0$.

[12] Notice that the recursive structure is obtained only when policy is exogenous. With endogenous policy, the solution is more complex.

the equilbrium. With θ determined, we get s^j from (8) and (9), since the differences in present values are independent of w. With θ and s^j determined, we obtain u^j and n from (1)-(3).

Notice that θ, ω, s^j, u^j and n are independent of the tax rate, t. The latter can be determined residually from the government's budget restriction, noting that the government uses the wage tax to finance benefits and monitoring costs:

$$twn = u^e bw + u^s zw + c\left(\sigma\right)\mu u^e w. \tag{14}$$

With the tax rate determined, the worker's take-home wage is obtained from $w = \omega/(1+t)$.

3. Optimal Unemployment Insurance

The optimal unemployment insurance system involves four instruments: b, p, μ, and σ. We assume that the government has a utilitarian objective function (W). We ignore discounting so it is valid to compare alternative steady states without considering the adjustment process. No discounting also has the convenient implication that the steady state flow of profits is zero.[13] Thus, the welfare objective is given by $W = u^e rU^e + u^s rU^s + nrE$, which simplifies to an employment-weighted average of instantaneous utilities with $r \to 0$:

$$W = nv(w, \overline{h}) + u^e v(B, s^e) + u^s v(Z, s^s). \tag{15}$$

The optimal policy maximizes (15) subject to the market equilibrium conditions, $s^j = s^j(b, p, \mu, \sigma)$ and $\theta = \theta(b, p, \mu, \sigma)$, as well as the balanced budget constraint, $t = t(b, p, \mu, \sigma)$. Let $\rho = \{b, p, \mu, \sigma\}$ denote the vector of policy parameters. Hence the vector of first-order conditions is given by $(dW/d\rho) = 0$.

Before proceeding to the numerical results it is useful to state two analytical results. First of all, the key result in Fredriksson and Holmlund (2001) applies directly. The following proposition reiterates proposition 2 in Fredriksson and Holmlund (2001).

Proposition 1 *The optimal policy involves* $p > 0$.

Proof. The proof is by contradiction.[14] First, note that an optimal replacement rate is strictly positive, $b > 0$. This follows from the utility function above with $\ln B = \ln(bw)$. Consider the trial solution $p = 0$. At $p = 0$, the first-order condition for σ has a solution at $\sigma = 0$ because $c'(\sigma) \geq 0$. Moreover, the condition for μ is irrelevant. So, let us fix μ at some arbitrary, but interior, value: $\mu^0 \in (0, \infty)$. The uniform benefit structure ($p = 0$) cannot be optimal if $(dW/dp) > 0$ at $p = 0$. Some manipulations of the first-order condition for p using the optimality of b (i.e. $dW/db = 0$) yields

$$\left(\frac{dW}{dp}\right)_{p=0} = \frac{\partial W}{\partial s^s}\frac{\partial s^s}{\partial p} > 0 \tag{16}$$

---

[13] Write the aggregate flow of profits as $\pi = n(y - w(1+t)) - kyv$, use equation (12) and the flow equilibrium condition $q(\theta)v = \phi n$ and obtain $\pi = nrky/q(\theta)$.

[14] The structure of the proof follows Fredriksson and Holmlund (2001), where the steps are explained in more detail.

where $\partial W/\partial s^s > 0$ denotes the partial derivative of welfare with respect to s^s holding θ constant, and $\partial s^s/\partial p > 0$ is also defined holding θ constant. ■

There are two key mechanisms that yield the sign of (16): there is a taxation externality associated with search and there is an "entitlement effect". The taxation externality derives from the fact that, given that some insurance is optimal ($b > 0$), taxes are required to finance unemployment expenditure. Individuals, however, do not take into account that taxes can be lowered if search intensity (and hence employment) increases. Therefore, $\partial W/\partial s^s > 0$. Moreover, the so called entitlement effect (c.f. Mortensen, 1977) will operate in this setting. Increasing the penalty will be conducive to search among those who are sanctioned since individuals will be eager to find a new job in order to qualify for (to be *entitled* to) UI benefit receipt, $\partial s^s/\partial p > 0$.[15] As a corollary to proposition 1, the optimal policy will involve an interior μ. In other words, the two tiered benefit structure, $b > 0$, $p > 0$, and $\mu \in (0, \infty)$, dominates the uniform benefit structure in welfare terms.

Another interesting question is whether it will be optimal to have the sanctioning rate depend on search intensity, given an optimal choice of b, p, and μ. Since the inspection of search is the defining characteristic of the monitoring and sanctions system in this setting, we can equally well phrase the question as: Given an optimal choice of a UI system with time limits, is it optimal to introduce a system of monitoring and sanctions? The following proposition gives the condition when the answer turns out to be affirmative:

Proposition 2 *Let $\hat{\rho} = \{b, p, \mu, \sigma = 0\}$ denote the solution to the restricted problem of optimal UI design. Then, the optimal policy will involve $\sigma > 0$ if*

$$\left(\frac{b}{\mu}\frac{\lambda + (1-\lambda)(1-p)}{\lambda s^e + (1-\lambda)s^s}\frac{\partial s^e}{\partial \sigma}\right)_{\rho=\hat{\rho}} > c'(0).$$

Proof. The proof proceeds as follows. Given that the two-tiered benefit structure is optimal, there are interior solutions to the first-order conditions $(dW/db) = 0$, $(dW/dp) = 0$, and $(dW/d\mu) = 0$. An UI system with monitoring and sanctions must be optimal if $(dW/d\sigma) > 0$ at the point where $\sigma = 0$ and the remaining first-order conditions hold. Some manipulations of the first-order condition for σ using $(dW/d\mu) = 0$ yields

$$\frac{dW}{d\sigma} = \frac{\partial W}{\partial s^e}\frac{\partial s^e}{\partial \sigma} - \frac{c'(0)\mu u^e}{(1+t)n} \tag{17}$$

where $\partial W/\partial s^e > 0$ denotes the partial derivative of welfare with respect to s^e holding θ constant, and $\partial s^e/\partial \sigma > 0$ is also defined holding θ constant. Introducing the explicit expression for $\partial W/\partial s^e$ and rewriting slightly, we get

$$\text{sign}\left(\frac{dW}{d\sigma}\right)_{\rho=\hat{\rho}} = \text{sign}\left\{\left(\frac{b}{\mu}\frac{\lambda + (1-\lambda)(1-p)}{\lambda s^e + (1-\lambda)s^s}\frac{\partial s^e}{\partial \sigma}\right)_{\rho=\hat{\rho}} - c'(0)\right\}.$$

■

---

[15] Note that $p = 1 - z/b$, so an increase in b, given z, raises the penalty.

Equation (17) illustrates the basic trade-off in introducing a monitoring and sanctions system. A monitoring and sanctions system restores the search incentives among the eligible, $\partial s^e/\partial \sigma > 0$. Again, this is a good thing since there is a taxation externality which is not taken into account in the private determination of search. However, inspecting search consumes real resources as indicated by the second term in (17). If this cost is sufficiently high, the monitoring and sanctions system will not be introduced.[16]

By inspection of (16) and (17), the extent that search responds to incentives is going to be crucial for the amount of benefit differentiation and the argument for introducing monitoring and sanctions.

4. Numerical Analysis

We have calibrated the model numerically so as to provide some information on plausible numbers. The basic time unit is taken to be a quarter and the matching function is Cobb-Douglas, $H = aS^{1-\eta}v^{\eta}$, where we set $\eta = 0.5$.[17] We fix hours of work exogenously to $\overline{h} = 0.75$ and use the following parametrization of the value of leisure:

$$\Gamma(l) = \chi \frac{l^{\kappa} - 1}{\kappa} \tag{18}$$

where $\kappa < 1$. The marginal product of labour is normalized to unity and we impose $\beta = \eta$. This assumption corresponds to the Hosios (1990) efficiency condition for the case where agents are risk neutral.[18]

We calibrate the model for a uniform benefit system ($p = 0$) in such a way that a number of labour market indicators become comparable with the average across the OECD. A replacement rate of $b = 0.3$ corresponds approximately to the average net replacement rate across the OECD in 2001 (OECD, 2004). The parameters a and χ are chosen with an eye towards vacancy duration and search intensity. We set $a = 1.7$ and $\chi = 0.6$. Remaining parameters (k, κ, and ϕ) are calibrated such that expected unemployment duration is one quarter, the partial equilibrium elasticity of the job hazard with respect to unemployment benefits equals -0.5, which is in the middle range of the available estimates (Layard, Nickell, and Jackman, 1991). The calibrated unemployment rate equals 6.5 %, the average OECD unemployment rate in 2001. The calibrated values further imply that the inflow into unemployment is 7 % per quarter and that the expected vacancy duration, $1/q(\theta)$, is close to half a quarter. The expected vacancy

---

[16] If introducing a sanction system involves a fixed set-up cost (besides C), then clearly the set-up cost should not be too big either. Note also that Proposition 2 relates to the result in Boone and van Ours (2006). Their key result is that a monitoring and sanction system will be more efficient in restoring search incentives than overall benefit reductions. This result is derived by means of numerical solutions, but with $c' = 0$. Proposition 2 shows that their conclusion holds analytically. In addition it extends their result further: given $c' = 0$, a system with monitoring and sanctions will dominate the two-tiered benefit system analyzed by Fredriksson and Holmlund (2001).

[17] Broersma and Van Ours (1999) give an overview of recent empirical studies of the matching function. They find that a value of η of 0.5 is a reasonable approximation.

[18] Note that in our case workers are risk averse and hence $\beta = \eta$ is not sufficient for efficiency.

cost, $ky/q(\theta)$, amounts to almost a quarter of production.[19] In the baseline calibration, search intensity equals $s = 0.7$. These are the main characteristics of our model although for some of these variables there is no empirical equivalent. Vacancy data are scarce but we know the number of vacancies is smaller than the number of unemployed in all countries for which there is available information . Our simulation results are consistent with this stylized fact ($\theta < 1$). Table 1 summarizes the baseline parameter (the parameter δ is discussed in section 4.2) and column 1 of Table 2 reports the outcomes of the calibration for some key variables.

The remainder of this section is concerned with the characteristics of alternative (optimal) UI systems. In section 4.1, we consider UI systems having infinite and finite benefit duration. Section 4.2 analyzes the case for monitoring and sanctions. Section 4.3 reports the results of rather extensive sensitivity analyses. Section 4.4 compares the predictions of our model with sanction rates observed in the data.

4.1. *Infinite vs Finite UI Benefit Duration*

There are two natural focal points in the model. The first is the optimal uniform system (which has infinite UI duration: $\mu = 0$); the second is a system with optimal time limits (finite UI duration: $\mu > 0$ but $\sigma = 0$).

Our primary objective is to characterize alternative UI systems which maximize welfare (i.e. the expected utility of workers). The last line of Table 2 thus presents welfare gains associated with particular policies. The welfare gain has the interpretation of a "consumption tax" (in %) that equalizes welfare across two policy regimes. To be specific, let W^R represent welfare associated with a reference policy regime and W^A welfare associated with an alternative policy. Our measure of the welfare gain of policy A relative to policy R is given by the value of the tax rate τ that solves $W^A[(1-\tau)m;\cdot] = W^R$. With logarithmic utility functions we have $\Delta W \equiv W^A - W^R = -\ln(1-\tau) \approx \tau$. The welfare gains are always reported relative to the optimal uniform system (i.e. the system featuring infinite benefit duration). In order to compare, say, the system with time limits with the base run, one only has to take the difference between the two entries for the welfare gain (ΔW).

The first set of results is contained in column 2 of Table 2, where we report the results of determining the optimal uniform replacement rate. The optimal replacement rate is around 36 %. A higher replacement rate reduces search incentives and incentives for wage restraint, so unemployment increases. With the optimal uniform replacement rate, unemployment rises to 8.1 %. Individuals living in the baseline economy would be willing to pay 0.33 % of consumption to avoid a move from the optimal uniform replacement rate to a replacement rate of 30 %.

The characteristics of the optimal system with time limits are given in column 3. Benefit differentiation is substantial and the duration of UI benefit receipt is fairly short – 65% of the unemployed have run out of unemployment benefits. The UI replacement rate amounts to 55 % of the wage; the penalty associated with the loss of entitlement is around 41 %. The benefit

---

[19] Empirical estimates of vacancy costs are hard to come by. In practice, the values used in calibrated search models are chosen with an eye towards the equilibrium outcomes. The value we have chosen is in the same ballpark as the values used by Marimon and Zillibotti (2000). Note that expenditure on recruitment costs as a fraction of total output in the economy is kv/n, which amounts to almost 10 percent in our baseline economy.

system with limited duration is substantially more generous than the system with infinite duration; with finite duration, unemployment expenditure per non-employed equals 40.5 %.

Because the government has two additional instruments (μ and p) besides b it is not surprising that there is a welfare gain associated with introducting a time limit on benefit duration. Also note that unemployment goes up by moving from the optimal uniform system to exogenous time limits. In other words, unemployment is not a sufficient statistic for welfare in this case.

4.2. *Monitoring and Sanctions*

The argument in favor of monitoring and sanctions hinges crucially on the costs of this system; see proposition 2. Unfortunately, the cost associated with monitoring and sanctions is something of a black box. In this section, we therefore give an upper bound on the marginal cost below which monitoring and sanctions are an ingredient of the optimal system. Since this upper bound turns out to be very high, we go on to characterize the optimal UI system with monitoring and sanctions.

Is it optimal to introduce monitoring and sanctions? In proposition 2 we stated the condition when the introduction of monitoring and sanctions represents a welfare improvement. For the introduction of monitoring and sanctions to be a welfare improvement relative to the case with time limits, $c'(0)$ has to be less than the gain as represented by greater search incentives among UI recipients with monitoring.

To answer the question posed above, we evaluate the condition given in proposition 2 at the solution for the optimal system with time limits. It turns out that if $c'(0) \leq \hat{c} = 0.076$, it is optimal to introduce monitoring and sanctions. This number has to be considered extremely high. Since the marginal product of labour and the labour force are normalized to unity, we can relate the cut-off value to (private sector) GDP by dividing by the employment rate (which is around 92 % in the optimal system with time limits). So, the calculated cut-off values suggest that as long as the marginal cost is no greater than $7.6/0.92 = 8.3$ % of GDP, it is optimal to introduce monitoring and sanctions. Thus, the introduction of monitoring and sanctions is most likely a welfare improvement relative to the case with time limits.

What is the optimal design of a monitoring and sanctions system? This clearly depends on the exact form of the cost function $c(\sigma)$. Assume that $c(\sigma)$ takes the form of $c(\sigma) = \delta\sigma$. To estimate a reasonable value for δ, we used Swedish data on the relative number of employees at the Public Employment Service (PES), since PES officers are responsible for monitoring job search in Sweden. We also used information on how often each PES employee meets a particular unemployed, and the fraction of total time that the PES officer spends in meetings with the unemployed. This calculation, which is presented in greater detail in Appendix A, suggests that the marginal cost of monitoring is in the order of $c(\sigma) = \delta\sigma = 0.00785$. Provided that $\sigma \geq 0.785$ in Sweden, then $\delta = 0.01$ is a conservative estimate.

The fourth column of Table 2 presents the outcome with the optimal system in the baseline economy. The optimal system involves $\sigma = 1$ (given our assumption $\sigma \in [0, 1]$). This particular result should be taken with a due grain of salt given the uncertainty about the costs of monitoring and the properties of the inspection technology. It is nevertheless interesting to note that a system with monitoring and sanctions are associated with non-trivial welfare gains relative to the alternatives characterized in columns (2) and (3). Also note that the optimal UI replacement rate is higher when we introduce monitoring and sanctions. There is a slight fall in unemployment as compared to columns (2) and (3), a result driven by a substantial increase in search effort among the unemployed (particularly those eligible for UI who now face additional incentives to search). Finally, note that the fraction of unemployed with a sanction is considerably lower in column (4) than in the column with exogenous time limits. Less people need to be penalized in a monitoring system in order to get similar (in fact, better) welfare and search incentive effects. Now, only 27% of the unemployed have run out of unemployment benefits.

4.3. *Sensitivity Analyses*

We have performed four types of sensitivity analyses. First, we investigate monitoring and benefit sanctions in a "less flexible" economy. Second, we consider the effects of higher monitoring costs. Third, we examine whether the results are robust to an alternative specification of preferences. Fourth, we consider an economy where workers are heterogenous with respect to the valuation of leisure and thereby the effort devoted to job search.

The motivation for conducting the first sensitivity analysis is that the case for monitoring and sanctions hinges on how responsive search is to incentives; in the less flexible economy search effort is thus less elastic. The reason for doing the second sensitivity analysis is simply that the cost of the monitoring and sanctions system is a black-box. By doing the third sensitivity analysis, we want to investigate whether our results are robust to alternative rates of risk aversion. Finally, the fourth sensitivity analysis is motivated by the wish to add more realism to the model.

The results of these exercises are shown in Table 3. The table only report detailed results for the monitoring and sanctions system. The first column of Table 3 reproduces the baseline results (from column 4 of Table 2). The bottom row of Table 3 reports changes in welfare associated with optimal time limits.[20] In the remainder of this section we will discuss the results in more detail. As a preview to the results, note that the case for monitoring and sanctions remains solid.

4.3.1. *A Less Flexible Economy*

We calibrate an alternative "less flexible" economy which has an identical unemployment rate (with a replacement rate equal to 0.3) but search is less responsive to incentives. We obtain this characterization by lowering the constant a in the matching function and compensating for this by a reduction in χ.[21]

---

[20] Again, all changes in welfare are relative to the optimal uniform case.

[21] A reduction in χ means that individuals place a lower value on leisure. The consequences of this are twofold: first, they are willing to search harder; second, and crucially, search is less responsive to changes in incentives. In particular, the partial equilibrium elasticity of the job hazard with respect to unemployment benefits equals -0.2 in this case. This is at the lower end of the interval given by Layard et al. (1991).

Column 2 reports the results. First, note that the welfare changes – for time limit as well as monitoring and sanctions – are smaller in the less flexible economy. Because unemployed workers react less to incentives, improving these incentives has a smaller effect on welfare. Further, comparing columns 1 and 2 we find the following. The reduction in a implies that the matching is less efficient and hence unemployment is higher in the less flexible economy. Tightness and the wage rate are lower in the less flexible case but the replacement rate is comparable in both economies. Since the unemployed react less to incentives in the less flexible economy, the penalty and monitoring rate are both lower than in the baseline economy.

4.3.2. *Higher Monitoring Costs*

To investigate the effects of introducing monitoring and benefit sanctions in case monitoring costs are higher, we consider the case where $\delta = 0.02$. Column 3 of Table 3 shows that the trade-off between monitoring and sanctions depends on the costs of monitoring: the higher the cost, the lower the monitoring rate and the higher the penalty (compared to column 1). Since the higher monitoring cost does not affect the optimal time limit system, the welfare gain from the monitoring system is smaller than in the baseline economy. The lower monitoring rate reduces search intensity for eligible unemployed and the higher sanction raises the search intensity of the unemployed on UA. Overall, the unemployment rate changes marginally.

We have also examined the trade-off between monitoring and sanctions in greater detail. In particular we are interested in the following question: What kind of monitoring cost would imply that we approach the Beckerian corner solution ($\mu \to 0$, $p \to 1$)? The answer is that the monitoring costs need to be extremely high. For example, it turns out that for $\delta = 0.14$ the optimal system in the baseline economy features $p = 0.929$ and $\mu = 0.234$. Risk aversion in combination with a random monitoring technology implies that it is generally not optimal to impose the maximal sanction ($p = 1$).

4.3.3. *CRRA Preferences*

Our results so far are all based on preferences that are logarithmic in income. To examine whether the results carry over to other utility functions we consider the following specification:

$$v^* = \begin{cases} \left[(ml^{\chi^*})^{\zeta} - 1\right]/\zeta, & \zeta < 1,\ \zeta \neq 0 \\ \ln m + \chi^* \ln l, & \zeta = 0 \end{cases} \tag{19}$$

where the degree of (constant) relative risk aversion (RRA) is $1 - \zeta$. This specification is convenient and commonly used (e.g., Fredriksson and Holmlund, 2001). Relative to the utility function used thus far, the preferences specified in (19) are more flexible in income but also more restrictive in terms of leisure. Clearly, the two sets of preferences are very similar in the log-utility case; when $\zeta = 0$, (19) corresponds to (18) with $\kappa = 0$ instead of $\kappa = 0.24$ as in Table 1.

The fourth column of Table 3 shows results from simulations with log utility (RRA=1); in the fifth column an alternative with stronger risk aversion (RRA=2) is considered.[22] Column 4 shows that the welfare gains relative to an optimal uniform UI system are very similar to the gains reported before. Comparing columns 4 and 5, the message is clear: stronger risk aversion strengthens the case for generous unemployment benefits, but also for monitoring and sanctions. Put differently, the more one wants to redistribute to the unemployed, the stricter the monitoring system becomes.

4.3.4. *Heterogeneous Search*

In the model above, search intensity is the same for all eligible unemployed worker. Although the UI administrators observe search intensity with error, it may be that they eventually figure out what the search effort is for all workers. This would induce them to either sanction all workers or none of them (depending on whether this common search effort is high enough or not). In order to increase the realism of our model we introduce unobserved heterogeneity in search decisions. Consider the possibility that workers differ with respect to the value of leisure. This leads to differences in optimal search effort among workers with identical benefit levels. Now case workers need to monitor each eligible unemployed worker as they cannot know the search intensity of an arbitrary unemployed worker.

We introduce preference heterogeneity in a simple stochastic fashion: workers are ex ante identical but are subject to random shocks to the value of leisure, which can be thought of as shocks affecting motivation for work, health or productivity in household work Specifically, consider our utility function of the form

$$v_k = \ln m + \chi_k \left(\frac{(1-s_k)^{\kappa} - 1}{\kappa} \right) \tag{20}$$

where subscript k indexes preference type. Note that the marginal rate of substitution between consumption and leisure is given by χ_j. Suppose that χ is subject to random shocks and assume for simplicity that χ can take only two values, χ_1 and χ_2 where $\chi_2 >> \chi_1$. Transitions from state 1 to 2 occur at the exogenous rate γ_1 and transitions from state 2 to 1 at the rate γ_2. Assume that χ_2 is so large that search becomes prohibitively costly in state 2. In this state, therefore, there are no transitions to employment.[23]

Let us now take this idea to our model with monitoring and sanctions. We will then get four states of unemployment, viz. eligible type 1 and type 2 unemployed as well as sanctioned

---

[22] In this case, we use the parameter values reported in Fredriksson and Holmlund (2001). The parameter values are slightly different but this is not affecting the results in any way.

[23] There are, of course, alternative assumptions one might consider. For example, it may be that type 2 workers search but with lower intensity than type 1 ones. Suppose further that once type 2 workers become employed, they are activated so as to become type 1 workers. Under these assumptions, there would still be a single equilibrium wage in the economy. The case worker would then face two types of unemployed individuals, where one type searches more intensely than the other. To finalize the model, one has to specify the probability distribution over the estimates of search among the two types of unemployed. Although such a model would be feasible to implement, we feel that our formulation, while a bit extreme, captures the essentials of the heterogeneity aspects.

individuals of type 1 and type 2. Assume that there is only one employment state and that all job losers end up as type 1 eligible workers, *i.e.* workers that pursue active search. Type 2 eligible workers don't search so their benefits expire at the rate μ. Assume as before that monitoring implies that benefits expire at the rate $\mu\pi(s_1^e) = \mu(1 - \sigma s_1^e)$ for eligible workers who search.

From the flow equilibrium equations of this extended model (see Appendix B), we can compute the ratio between the number of eligible type 1 workers and all eligible workers:

$$\frac{u_1^e}{u^e} = \left(\frac{\gamma_2 + \mu}{\gamma_1 + \gamma_2 + \mu}\right) \equiv \gamma. \tag{21}$$

This ratio is a function of the parameters that determine shocks to preferences (γ_1 and γ_2) as well as of the monitoring rate μ: the higher the monitoring rate, the larger the number of type 2 workers who are sanctioned. The policy maker can thus influence search effort on the extensive margin, in addition to the previously considered intensive margin.

The value functions for this case are straightforward extensions of those given before (see Appendix B). The first-order conditions for optimal search only apply to type 1 individuals. Since only type 1 individuals search, the wage bargain is the result of matches between type 1 unemployed and firms and there will be a single equilibrium wage in the economy.[24] The welfare objective is written as:

$$W = nv_1(w, h) + u_1^e v_1(B, s_1^e) + u_1^s v_1(Z, s_1^s) + u_2^e v_2(B, 0) + u_2^s v_2(Z, 0). \tag{22}$$

We have set $\gamma_1 = 1$, which implies that unemployed job searchers experience preference shocks once a quarter on average, comparable in magnitude to the job finding rate. The other key parameter, γ_2, is set to 9 so as to obtain $u_1^e/u^e = 0.9$ when $\mu = 0$, implying that 90 % of unemployed workers eligible for UI is actively searching. Numerical results are given in column (6) of Table 3.[25] As shown, the monitoring rate increases and type 2 workers are more likely to be sanctioned ($u_1^s/u^s < 0.9$). Unemployment is higher compared to the baseline case where all unemployed search for a job. The replacement rate is lower because more unemployed have a bigger impact on the government budget and a high replacement rate is no longer affordable. Notice also that sharper incentives – either through a time limit or through monitoring and sanctions – do not affect the search of the type 2 unemployed. This implies that the overall welfare gain is smaller in the present case than in the baseline case.

This version of the model thus has an explicit rationale for case workers to inspect the unemployed because search intensities differ. Our main conclusion – that monitoring and sanctions do better in terms of welfare than a time limit system – is robust to such heterogeneity.

---

[24] We have ignored preference shocks among employed workers. However, incorporation of such shocks would not change much in the model. A preference shock to employed workers would lead to a higher job separation rate in addition to ϕ (as type 2 workers would quit their jobs) but would not change the fundamentals of wage bargaining.

[25] We also investigated the case in which type 2 individuals are excluded from the welfare objective, *i.e.* policy makers care only about type 1 individuals who seriously look for a job. This changes the results very little.

4.4. *A Brief Look at the Data*

Having calculated the optimal systems with monitoring and sanctions it is tempting to relate the predictions of the model to the data. Some of the parameters of the monitoring and sanctions system are of course unobservable. However, there are observations on the UI replacement rates, the penalties for violating search requirements, and the associated sanctioning rates. Presumably, there is a lot of noise in the data pertaining to sanction rates. Nevertheless, there is great variation in these data as is clear from Grubb (2001). It seems that the US and Switzerland are the extreme cases in terms of having systems with a large number of sanctions. In the US in the late 1990s, around 10 % of beneficiaries were sanctioned each quarter for behavior during the benefit period. In addition, some 25 % of the (stock of) eligible unemployed were "sanctioned" because they exhausted their benefits.[26] Based on these data, the quarterly sanction rate in the US would be in the order of 35 %. With the exception of Switzerland, sanctions during the benefit period are substantially less common in the European countries. In fact, the sanction rates are typically lower than one % per quarter; see Grubb (2001) for further details.

The number of sanctions seems to be inversely related to the severeness of the penalty. In the US, the normal sanction for a job search infringement is a loss of benefits for one week.[27] In Sweden, on the other hand, the penalty until recently was the loss of benefits for twelve weeks.[28]

What does the model have to say about the number of sanctions? Fig. 1 addresses this question by plotting the sanctioning rates against σ and assuming $\delta = 0.01$. In addition to the baseline and the less flexible economy, we also consider an economy with low turnover.[29] Sanctioning rates decline in σ because a rise in σ raises s^e.

When $\sigma = 1$, as is optimal given our assumptions, the quarterly sanction rates hover between 10 and 30 % depending on the exact assumptions; see Table 4. The number of sanctions in the baseline economy best conform to sanctioning data for the US. To get at the numbers for the typical European country, it appears that one would have to apply a combination of less elastic search, lower turnover, and higher monitoring costs.

5. **Concluding Remarks**

In this paper we have analyzed the design of optimal unemployment insurance in a search equilibrium framework where search effort among the unemployed is not perfectly observable.

---

[26] This estimate is a crude average for the period 1995-2000. The number of exhaustions per quarter amounted to some 600 000 individuals, the number of unemployed to 6.5 millions, and the fraction eligible for UI to 35 percent. Source: US Department of Labor (labor force statistics and UI program statistics).

[27] Notice, though, that there is a rather harsh "penalty" associated with the expiration of UI benefits in the US. In 1991, benefits were reduced by more than 60 percent when benefits expired and the individual was forced to claim welfare benefits instead; see Wang and Williamson (1996).

[28] The Swedish system was changed in 2001 in the direction of smaller penalties.

[29] The "low turnover" economy has a lower job destruction rate ϕ (around 22 percent per year) and higher value for a in the matching function to keep unemployment at the baseline value of 6.5 percent.

We have examined to what extent the optimal policy should involve monitoring of search effort and benefit sanctions if observed search is found insufficient. The results suggest that the introduction of a system with monitoring and sanctions represents a welfare improvement for reasonable values of the monitoring costs. Those costs would have to be implausibly high for this conclusion not to hold.

The policy prescription following from our analysis is thus different from Becker's (1968) well known result, where the penalty should be maximal and the probability of getting caught should be close to zero. There are two key assumptions delivering our results. First, individuals are risk averse and, second, monitoring is imperfect. With imperfect monitoring some individuals will be sanctioned even though they search to rule and giving them the maximal penalty is not optimal with risk aversion.

While we are reasonably comfortable in saying that monitoring and sanctions represent a welfare improvement, it is much more difficult to give clear advice on the characteristics of such a system. The reason for this conclusion is that the exact formulation of the monitoring and sanctions system depends on the cost of running such a system. Unfortunately, the cost of running the system is something of a black box.

An issue that we have not addressed is the possibility that formal search requirements may induce individuals to use formal rather than informal search methods and therefore bring little increase in total search intensity. Nevertheless, it is likely that *general* search requirements – such as the number of job applications filed during a week – should minimize the risk of substitution between search channels. Presumably, substitution is going to be more severe in systems where search requirements are linked to formal channels such as referrals by the public employment service. On this account, search requirements specified in terms of independent job search, as used in the US, the Netherlands and Switzerland, seem to be preferable.

Appendix A: Estimating the Marginal Cost of Monitoring

To obtain a reasonable value for the cost of monitoring an additional individual ($c(\sigma)$) we performed the following calculation. We relied on data from Sweden, where PES administrators are responsible for monitoring whether unemployed individuals have searched to rule or not. Three sources of information were used: (*i*) the relative number of employees at the PES; (*ii*) the fraction of time that a PES officer meets with the unemployed; and (*iii*) the number of contacts between the PES officer and a particular unemployed individual. Information pertaining to items (*ii*) and (*iii*) is taken from Lundin (2000).

In the main text the total cost of the monitoring and sanctions system was specified as: $C = c(\sigma)\mu u^e w$. To get an approximate value for C we start by calculating the wage bill paid to individuals involved in monitoring. Since the labour force and the marginal product of labour are normalized to unity, the wage bill is measured relative to these items. The PES service employs approximately 10,000 individuals in Sweden, which translates to around 0.25 % of the labour force. On average PES officers spend 30 % of their time in meetings with the unemployed. Assuming that the unemployed are monitored each time they meet with a PES officer we have $C = 0.0025 \times 0.3w = 0.00075w$. Thus we have $C = c(\sigma)\mu u^e w = 0.00075w$. Turning to the left-hand side of this equation, we set the number of unemployed individuals eligible for UI to 5 %. With this assumption, we only need an estimate of μ to get an estimate of $c(\sigma)$. The information used to estimate μ is derived from a question put to PES officers regarding the number of meetings with individuals searching for a job. When asked about their contact frequency, 35 % of PES officers answered "at most once a month"; 34 % answered "at most once every other month"; and 31 % answered "at most once every quarter". Thus on average a PES officer has $(1 \times 0.35 + 0.5 \times 0.34 + 0.31/3) \times 3 = 1.91$ meetings with a particular unemployed per quarter. Hence we have $c(\sigma) = 0.00075/(\mu u^e) = 0.00075/(1.91 \times 0.05) \approx 0.00785$. There is still one unknown in this equation, however; the estimated value of $c(\sigma)$ pertains to a given value of σ. Assuming that $c(\sigma) = \delta\sigma$, we have $\delta = 0.01$ for $\sigma = 0.785$ and $\delta = 0.02$ for $\sigma = 0.785/2$.

Appendix B: Heterogeneous Preferences

The value functions corresponding to the model with preference heterogeneity among the unemployed take the following form:

$$\begin{aligned}
rU_1^e &= \ln B + \chi_1 \left(\frac{(1 - s_1^e)^\kappa - 1}{\kappa} \right) + s^e \alpha(\theta) \left(E - U_1^e \right) \\
&\quad + \mu\pi \left(s^e \right) \left(U_1^s - U_1^e \right) + \gamma_1 \left(U_2^e - U_1^e \right) \\
rU_2^e &= \ln B + \mu (U_2^s - U_2^e) + \gamma_2 \left(U_1^e - U_2^e \right) \\
rU_1^s &= \ln Z + \chi_1 \left(\frac{(1 - s_1^s)^\kappa - 1}{\kappa} \right) + s^s \alpha(\theta)(E - U_1^s) + \gamma_1 \left(U_2^s - U_1^s \right) \\
rU_2^s &= \ln Z + \gamma_2 \left(U_1^s - U_2^s \right) \\
rE &= \ln w + \chi_1 \left(\frac{(1 - h)^\kappa - 1}{\kappa} \right) + \phi(U_1^e - E)
\end{aligned}$$

The flow equilibrium equations corresponding to this economy are given as follows:

$$\begin{aligned}
\phi \left(1 - u_1^e - u_2^e - u_1^s - u_2^s \right) &= \alpha^e u_1^e + \alpha^s u_1^s \\
\left(\alpha^e + \mu\pi + \gamma_1 \right) u_1^e &= \phi \left(1 - u_1^e - u_2^e - u_1^s - u_2^s \right) + \gamma_2 u_2^e \\
\left(\alpha^s + \gamma_1 \right) u_1^s &= \mu\pi u_1^e + \gamma_2 u_2^s \\
\left(\gamma_2 + \mu \right) u_2^e &= \gamma_1 u_1^e
\end{aligned}$$

where the first equation pertains to outflows from and inflows to employment; note that $1 - u_1^e - u_2^e - u_1^s - u_2^s = n$. The second equation pertains to u_1^e, the third to u_1^s, and the fourth to u_2^e.

Appendix C: The Sanctioning Probability

This appendix addresses the "structural" interpretation of our sanctioning probability: $\pi(s_i^e) = 1 - \sigma s_i^e$. This interpretation builds on the model with heterogenous search; see section 4.3.4.

There are, at each point in time, at most two types of unemployed workers receiving unemployment benefits, viz. those who search with intensity s_1^e and those who don't search. Benefit administrators monitor search intensity randomly, know the distribution of preferences in the population of benefit recipients, but cannot identify the individual type. Inferring types by means of observed search is not possible because search is observed with error.

Let $s_i^e \in \{0, s_1^e\}$ denote individual search intensity and s_o^e observed search intensity. Then

$$s_o^e = s_i^e + \varepsilon$$

where ε is a mean zero observation error with support $\varepsilon \in [\varepsilon_L, \varepsilon_U]$. Since $s_i^e \in [0,1]$ then so should s_o^e. This in turn implies restrictions on $\varepsilon_L, \varepsilon_U$. Let us introduce a parameter that indexes the extent of observation error. In particular, let $\varepsilon \in [-(1-\tilde{\sigma})s_i^e, (1-\tilde{\sigma})(1-s_i^e)]$. If $\tilde{\sigma} = 1$, there is no observation error. If $\tilde{\sigma} = 0$, observed search belongs to the entire admissible range. Suppose also that ε is uniform. Then $\tilde{\sigma} = 0$ is a completely random inspection technology. We think of $\tilde{\sigma}$ as a parameter that the government can invest resources in improving.

An individual is sanctioned whenever $s_o^e \leq R$, where $R \in [0,1]$ denotes the search requirement. The probability of being sanctioned given that the individual supplies s_i^e units of search is then

$$\pi_i = \int_{\varepsilon_L}^{R-s_i^e} \frac{1}{\varepsilon_U - \varepsilon_L} d\varepsilon = \frac{R - s_i^e}{\varepsilon_U - \varepsilon_L} - \frac{\varepsilon_L}{\varepsilon_U - \varepsilon_L}.$$

Since $\varepsilon_U - \varepsilon_L = 1 - \tilde{\sigma}$ and $\varepsilon_L = -(1-\tilde{\sigma})s_i^e$, we get

$$\pi_i = \frac{R}{1-\tilde{\sigma}} - \left(\frac{\tilde{\sigma}}{1-\tilde{\sigma}}\right) s_i^e.$$

Now we want to impose some restrictions on the parameters of the inspections technology $(R, \tilde{\sigma})$ to make sure that $\pi_i \in [0,1]$ for all $s_i^e \in [0,1]$. We impose the following conditions:

1. If $s_i^e = 0$ then $\pi_i = 1$.
2. If $s_i^e = 1$ then $\pi_i \in [0,1]$.

The first condition gives $R = 1-\tilde{\sigma}$. Given $R = 1-\tilde{\sigma}$, the second condition yields $\tilde{\sigma} \in [0, 1/2]$. The conditions we impose on the parameters thus imply that an individual who searches full time is sanctioned with positive probability, i.e., there is a probability of making Type II errors for all values of s_i^e.

In sum, the above assumptions lead to the following formulation for π

$$\pi_i = 1 - \left(\frac{\tilde{\sigma}}{1-\tilde{\sigma}}\right) s_i^e, \; \tilde{\sigma} \in [0, 1/2]$$

or alternatively, defining $\sigma = \tilde{\sigma}/(1-\tilde{\sigma})$

$$\pi_i = 1 - \sigma s_i^e, \; \sigma \in [0,1]$$

which is what we have in the main text.

References

Abbring, J., van den Berg, G. and van Ours, J.C. (2005). 'The effect of unemployment insurance sanctions on the transition rate from unemployment to employment', *ECONOMIC JOURNAL*, vol.115 (June), pp. 602-30.

Becker, G. (1968). 'Crime and punishment: An economic approach', *Journal of Political Economy*, vol. 76(2) (March-April), pp. 169-217.

Benus, J., Joesch, J., Johnson T. and Klepinger, D. (1997). 'Evaluation of the Maryland unemployment insurance work search demonstration: Final report', Batelle Memorial Institute in association with Abt Associates Inc (http://wdr.doleta.gov/owsdrr/98-2/).

Black, D., Smith, J., Berger, M. and Noel B. (2003). 'Is the threat of reemployment services more effective than the services themselves? Experimental evidence from the UI system', *American Economic Review*, 93(4) (September), pp. 1313-27.

Boadway, R. and Cuff K. (1999). 'Monitoring job search as an instrument for targeting transfers, *International Tax and Public Fincance*, vol. 6(3) (August), pp. 317-37.

Boone, J. and van Ours, J.C. (2006). 'Modelling financial incentives to get unemployed back to work', *Journal of Institutional and Theoretical Economics*', forthcoming.

Broersma, L. and van Ours, J.C. (1999). 'Job searchers, job matches and the elasticity of matching', *Labour Economics*, vol.6(1) (March), pp. 77-93.

Cahuc, P. and Lehmann, E. (2000). 'Should unemployment benefits decreases with the unemployment spell?', *Journal of Public Economics*, vol. 77(1) (July), pp. 135-53.

Fredriksson, P. and Holmlund, B. (2001). 'Optimal unemployment insurance in search equilibrium', *Journal of Labour Economics*, vol. 19(2) (April), pp. 370-99.

Garoupa, N. (1997). 'The theory of optimal law enforcement', *Journal of Economic Surveys*, vol. 11(3) (September), pp. 267-95.

Grubb, D. (2001). 'Eligibility criteria for unemployment benefits', in *Labour Market Policies and the Public Employment Service*, pp. 187-216, Paris: OECD.

Hopenhayn, H.A. and Nicolini, J.P. (1997). 'Optimal unemployment insurance', *Journal of Political Economy*', vol. 105(2) (April), 412-38.

Hosios, A.J. (1990). 'On the efficiency of matching and related models of search and unemployment', *Review of Economic Studies*, vol. 57(2) (April), pp. 279-98.

Johnson, T. and Klepinger, D. (1994). 'Experimental evidence on unemployment insurance work-search policies', *Journal or Human Resources*, vol. 29(3) (Summer), pp. 695-717.

Lalive, R., van Ours J.C. and Zweimüller, J. (2005). 'The effect of benefit sanction on the duration of unemployment', *Journal of the European Economic Association*, vol. 3(6) (December), pp. 1386-1417.

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

Lundin, M. (2000). 'Tillämpningen av arbetslöshetsförsäkringens regelverk vid arbetsförmedlingarna', mimeo. 2000:1, Institute for Labour Market Policy Evaluation (IFAU), Uppsala.

Marimon, R. and Zilibotti, F. (2000). 'Employment and distributional effects of restricting working time', European Economic Review, vol. 44(7) (June), pp. 1291-1326.

Mortensen, D.T. (1977). 'Unemployment insurance and job search decisions, *Industrial and labour Relations Review*, vol. 30(4) (July), pp. 505-17.

OECD (2001) *Labour Market Policies and Public Employment Service*, Paris: OECD.

OECD (2004) *Benefits and Wages*, Paris: OECD.

Pissarides, C.A. (1990) *Equilibrium Unemployment Theory*, Oxford: Basil Blackwell.

Polinsky, A.M. and Shavell, S. (1979). The optimal tradeoff between the probability and magnitude of fines, *American Economic Review*, vol. 69(5) (December), pp. 880-91.

Polinsky, A.M. and Shavell, S. (2000). 'The economic theory of public enforcement of law', *Journal of Economic Literature*, vol. 38(1) (March), pp. 45-76.

Shavell, S. and Weiss, L. (1979). 'The optimal payment of unemployment insurance benefits over time', *Journal of Political Economy*, vol. 87(6) (December), pp. 1347-63.

Van den Berg, G. and van der Klaauw, B. (2001). 'Counselling and monitoring of unemployed workers: Theory and evidence from a controlled social experiment', Working Paper 2001:12, Institute for Labour Market Policy Evaluation (IFAU), Uppsala.

Van den Berg, G., van der Klaauw, B. and van Ours J.C. (2004). 'Punitive sanctions and the transition rate from welfare to work', *Journal of Labour Economics*, vol. 22(1) (January), pp. 211-41.

Wang, C. and Williamson, S. (1996). 'Unemployment insurance with moral hazard in a dynamic economy', *Carnegie-Rochester Conference Series on Public Policy*, vol. 44 (June), pp. 1-41.

Table 1: Baseline Parameters

Interest rate (= rate of time preference)	$r = 0$
Job destruction rate	$\phi = 0.069519$
Leisure value	$\kappa = 0.239419$, $\chi = 0.6$
Matching function	$\eta = 0.5$, $a = 1.7$
Wage negotiations	$\beta = \eta = 0.5$
Production	$y = 1$
Vacancy costs	$k = 1.98335$
Monitoring costs	$\delta = 0.01$

Table 2: Numerical Results Baseline Economy

	Base run	Optimal uniform	Optimal time limit	Monitoring & sanctions
	(1)	(2)	(3)	(4)
b	0.300	0.363	0.553	0.626
p	0	0	0.410	0.564
μ	–	–	1.556	1.207
σ	–	–	0	1
s^e	0.700	0.607	0.556	0.755
s^s	–	–	0.670	0.737
θ	0.705	0.578	0.528	0.407
u (%)	6.50	8.13	8.19	7.87
u^e (%)	6.50	8.13	2.84	5.75
u^s (%)	–	–	5.35	2.12
w	0.913	0.909	0.908	0.906
t (%)	2.09	3.21	3.61	4.60
ΔW (%)	–0.33	0	0.21	0.82

Note: See Table 1 for baseline parameters.

Table 3: Sensitivity Analysis; Numerical Results with Monitoring and Sanctions

	Baseline economy	Less flexible economy	Higher costs	RRA=1	RRA=2	Heterogeneous search
	(1)	(2)	(3)	(4)	(5)	(6)
b	0.626	0.619	0.617	0.655	0.781	0.592
p	0.564	0.511	0.584	0.563	0.593	0.501
μ	1.207	1.017	1.039	0.903	0.514	1.276
σ	1	1	1	1	1	1
s^e	0.755	0.856	0.746	0.659	0.614	0.736
s^s	0.737	0.842	0.753	0.645	0.450	0.712
θ	0.407	0.371	0.408	0.157	0.151	0.421
$u(\%)$	7.87	8.47	7.88	7.98	9.19	8.78
$u^e(\%)$	5.75	7.07	5.96	5.87	6.87	5.47
$u^s(\%)$	2.12	1.40	1.92	2.11	2.32	3.32
u_1^e/u^e	–	–	–	–	–	0.910
u_1^s/u^s	–	–	–	–	–	0.883
w	0.906	0.894	0.906	0.897	0.883	0.905
$t(\%)$	4.60	5.33	4.66	4.89	6.76	4.70
$\Delta W(\%)$	0.82	0.55	0.75	0.67	3.53	0.67
Optimal time limit						
$\Delta W(\%)$	0.21	0.11	0.21	0.22	1.17	0.18

Note: Parameter values of the baseline economy: see Table 1; Less flexible economy: $a = 1.445$, $\chi = 0.364165$; Higher costs: $\delta = 0.02$; RRA=1 and RRA = 2: see equation(19); Heterogeneity: see main text.

Table 4: Quarterly Sanction Rates According to the Model

	Baseline economy	Less flexible economy	Low turnover economy
$\delta = 0.01$	0.296	0.146	0.242
$\delta = 0.02$	0.264	0.115	0.221

Figure 1: Quarterly sanction rates, δ=0.01