The joint design of unemployment insurance and employment protection. A first pass.*

Olivier Blanchard† Jean Tirole ‡

September 11, 2006

Abstract

Unemployment insurance and employment protection are typically discussed and studied in isolation. In this paper, we argue that they are tightly linked, and we focus on their joint optimal design in a simple model, with risk averse workers, risk neutral firms, and random shocks to productivity.

We show that, in the "first best", unemployment insurance comes with employment protection—in the form of layoff taxes; indeed, optimality requires that layoff taxes be equal to unemployment benefits. We then explore the implications of four broad categories of deviations from first best: limits on insurance, limits on layoff taxes, ex–post wage bargaining, and ex–ante heterogeneity of firms or workers. We show how the design must be modified in each case.

Finally, we draw out the implications of our analysis for current policy debates and reform proposals, from the financing of unemployment insurance, to the respective roles of severance payments and unemployment benefits.

Keywords: Unemployment insurance, employment protection, unemployment benefits, layoff taxes, layoffs, severance payments, experience rating.

JEL classifications: D60, E62, H21, J30, J32, J38, J65.

---

*We are grateful to Daron Acemoglu, Suman Basu, Larry Katz, Javier Ortega, John Hassler, Aleh Tsyvinski, the editor and the two referees for helpful comments.

†MIT and NBER.

‡IDEI and GREMAQ (UMR 5604 CNRS), Toulouse, and MIT.

Introduction

Unemployment insurance and employment protection are typically discussed and studied in isolation. In this paper, we argue that they are tightly linked, and we focus on their optimal joint design.

To show this, we start our analysis in Section 1 with a simple benchmark. Workers are risk averse; entrepreneurs run firms and are risk neutral. The productivity of any worker-firm match is random. If productivity is low enough, the worker and the firm may separate, in which case the worker becomes unemployed.

In that benchmark, a simple way to achieve the optimum is for the state to pay unemployment benefits so as to insure workers, and to levy layoff taxes so as to lead firms to internalize the cost of unemployment and take an efficient layoff decision. The optimum has two further characteristics: The first is that layoff taxes are equal to unemployment benefits: This common level delivers both full insurance and production efficiency. Thus, the benchmark shows the tight conceptual relation between unemployment insurance and employment protection—defined as layoff taxes. The second is that state intervention is not needed: The same allocation is achieved by having firms voluntarily pay severance payments to their workers; in effect, severance payments act both as unemployment insurance and layoff taxes.

Using this benchmark as a starting point, we then examine, in Sections 2 to 5, how these conclusions are affected by the introduction of four empirically-relevant deviations from the benchmark, namely: limits on unemployment insurance, limits on layoff taxes, ex–post wage bargaining, and ex–ante heterogeneity of either workers or firms. In each case, we ask two questions: The first is how the distortion affects the optimal combination of unemployment insurance and layoff taxes. The second is how the distortion affects the need and the scope for state intervention.

Reforms of both the unemployment insurance and employment protection are high on the policy agendas of many European and Latin American governments. Proposals range from the creation of unemployment accounts, to changes in the financing of the unemployment insurance system, to changes in the form of employment protection. These are complex issues, but we feel that our analysis can help think about the answers. This is what we do in Section 6.

1 A benchmark

In approaching the issue, we make two methodological choices.

First, we use a static, one-period, model. As such, this represents a large step back from recent dynamic models of either unemployment insurance or employment protection. We do so for two reasons. First, we want to focus on the joint design of unemployment insurance and employment protection, which makes things more difficult. Second, we want to explore a number of deviations, starting from as simple a benchmark as feasible. We believe that the basic insights we get from this analysis will extend to more dynamic frameworks; but this obviously remains to be shown.

Second, we use a mechanism design approach to the characterization of the optimum. This approach may appear unnecessarily heavy, especially in the benchmark itself (where the solution is straightforward), but we believe it pays off: It shows most clearly, in each case, first the characteristics of the optimal allocation, and then the role of unemployment benefits, taxes, and severance payments, in achieving this allocation.

1.1 Assumptions

Tastes and technology are as follows:

• The economy is composed of a continuum of mass 1 of workers, a continuum of mass (at least) 1 of entrepreneurs, and the state.

• *Entrepreneurs* are risk neutral. Each entrepreneur can start and run a firm. There is a fixed cost of creating a firm, I, which is the same for all entrepreneurs.

If a firm is created, a worker is hired, and the productivity of the match is then revealed. Productivity is given by y from cdf $G(y)$, with density $g(y)$ on $[0,1]$. The firm can either keep the worker and produce, or lay the worker off, who then becomes unemployed.

Realizations are iid across firms; there is no aggregate risk.

• The *firm*, but not the worker (or for that matter third parties such as an insurance company or the state) observes y.

• *Workers* are risk averse, with utility function $U(.)$. Absent unemployment benefits, utility if unemployed is given by $U(b)$ (so b is the wage equivalent of being unemployed).

1.2 The optimal allocation

Let $\bar{y}$ be the threshold level of productivity below which workers are laid off. Let w be the payment to the workers who remain employed, and μ be the payment to the workers who are laid off.

The optimal allocation maximizes expected worker utility subject to the economy's resource constraint:[1]

$$\max_{\{w,\mu,\bar{y}\}} V_W \equiv \quad G(\bar{y})U(b+\mu) + (1-G(\bar{y}))U(w)$$

subject to:

$$V \equiv -G(\bar{y})\mu + \int_{\bar{y}}^{1} y \, dG(y) - (1-G(\bar{y}))w = I$$

From the first–order conditions, it follows that:

$$w^* = b + \mu^* \tag{1}$$

$$\bar{y}^* = b \tag{2}$$

Given $\bar{y}^*$, the levels of w^* and μ^* are determined by the resource constraint.

Condition (1) is an *insurance condition*: Workers achieve the same level of utility, whether employed or laid-off and unemployed.

Condition (2) is an *efficiency condition*: From the point of view of total output, it is efficient for firms to produce so long as productivity exceeds the wage equivalent of being unemployed (we shall call b the production-efficient threshold level).

1.3 Implementation

Consider now the following implementation of the optimal allocation:

- *Stage 1.* The state chooses a payroll tax rate τ, a layoff tax rate f, and unemployment benefits μ.

- *Stage 2.* Entrepreneurs decide whether to start firms and pay the fixed cost.

[1] We derive the first best allocation ignoring the assumption that y is observed only by the firm. We shall show below that this optimal allocation can indeed be implemented.

They offer contracts to workers. Contracts are characterized, explicitly, by a wage w, and, implicitly (since y is not contractable), a threshold productivity level $\bar{y}$ below which the worker is laid off.

As all firms face the same cost and distribution of productivity, in equilibrium, all workers are initially hired.

• *Stage 3.* The productivity of each job is realized. Firms decide whether to keep or dismiss workers.

To show how the optimal allocation can be implemented, we work backwards in time.

At *Stage 3*, the cutoff $\bar{y}$ is such that the firm is indifferent between keeping the worker and paying $w + \tau$ in wage and payroll tax and dismissing the worker and paying layoff tax f, so:

$$\bar{y} = w + \tau - f. \tag{3}$$

If $y > \bar{y}$, the firm keeps the worker, produces y, pays w to the worker, and τ to the state. If $y < \bar{y}$, the firm lays the worker off, pays f to the state; the state pays μ to the worker.

At *Stage 2*, firms' wage offer w satisfies the free entry condition:

$$V_F \equiv \; -G(\bar{y})f + \int_{\bar{y}}^{1} y \; dG(y) - (1 - G(\bar{y}))(w + \tau) = I. \tag{4}$$

Consider now the *problem faced by the government* in choosing taxes and unemployment benefits at *Stage 1.* Condition (3) implies that to induce firms to take the production-efficient layoff decision $\bar{y}^* = b$, the following condition must hold:

$$w + \tau - f = b. \tag{5}$$

Because optimal insurance further requires that $w = b + \mu$, the state's policy must satisfy:

$$f - \tau = \mu. \tag{6}$$

The net fiscal cost to the firm of laying off a worker must be equal to the unemployment benefits paid to the worker by the state. Note that this condition implies a *positive* relation between the layoff and the payroll tax rates: For given unemployment benefits, the higher the payroll tax, the higher the layoff tax needed to induce the firm to take the production-efficient decision.

The government budget constraint implies a second relation between taxes and benefits:

$$V_G \equiv -G(\bar{y})(\mu - f) + (1 - G(\bar{y}))\tau = 0 \tag{7}$$

This constraint implies a *negative* relation between the layoff and the payroll tax rates: For given unemployment benefits, the higher the payroll tax, the lower the layoff tax required to balance the budget. Combining the two conditions gives:

$$f = \mu, \qquad \tau = 0 \tag{8}$$

The layoff tax must be equal to unemployment benefits, and the payroll tax rate is equal to zero.

We summarize our results in Proposition 1.

Proposition 1. *In the benchmark, the optimal allocation is such that workers are fully insured ($b + \mu^* = w^*$), and the threshold productivity is equal to the production-efficient level ($\bar{y}^* = b$).*

Implementation is achieved through unemployment benefits equal to μ^*, and layoff taxes $f = \mu^*$. Payroll taxes are equal to zero. Put another way, the contribution rate, defined as the ratio of layoff taxes to unemployment benefits, is equal to one.

1.4 Interpretation and discussion

The result that layoff taxes must be equal to unemployment benefits is a classic case of Pigovian internalization: To the extent that the state pays unemployment benefits to laid-off workers, layoff taxes lead firms to internalize these costs. Indeed, it is the rationale behind the experience rating systems in place in the different states in the United States.[2]

[2] See Baicker, Goldin, and Katz (1998) for a description of the politics and the arguments pro- and con-experience rating, presented in the 1920s and 1930s when these systems were put in place.

Indeed, within the assumptions of the benchmark, there is an even simpler way of making sure that firms internalize the costs of unemployment benefits: It is to have them provide unemployment benefits themselves rather than through the state. It is straightforward to see that the optimal allocation can also be implemented by simply letting firms pay severance payments. Firms will then want to offer severance payments equal to μ. There is no need for the state to intervene.

The financing of unemployment benefits through layoff taxes, and the lack of a rationale for state intervention hold in the benchmark. But do they hold more generally? What happens for example if, for moral hazard or other reasons, laid-off workers cannot be fully insured? Is it still optimal to finance unemployment benefits only through layoff taxes? What happens if firms face financial constraints and are sometimes unable to pay the layoff taxes? Is it still optimal to fully insure workers? What happens if wages are renegotiated ex-post, and unemployment insurance increases the reservation wage of workers in negotiations? What happens if some workers or some firms are more exposed to the risk of low productivity than others? Isn't there a risk that higher layoff taxes will affect them adversely? And, in all these cases, is it the case that firms can do it on their own, perhaps pooling resources through a private unemployment agency, or must the state intervene? These are the questions we take up in the next four sections.

2 Limits to insurance

In our benchmark, workers could be and were fully insured. There are various reasons why this may not be feasible. Workers may require incentives not to shirk when employed, or incentives to search when unemployed. Or there may be a non–pecuniary loss associated with becoming unemployed. We explore the implications of this last assumption, and return to a discussion of other potential reasons later.[3]

Assume that the utility of workers is now given by $U(c)$ if employed, and by $U(c) - B$ if unemployed, so $B > 0$ is the utility cost of being unemployed.[4] All other assumptions are the same as in the benchmark.

[3] Empirical evidence suggests that non-pecuniary losses associated with becoming unemployed are indeed large (see for example Winkelmann and Winkelmann (1998)).

[4] The derivation below goes through whatever the sign of B. But the substantive implications are obviously different.

2.1 The optimal allocation

The optimal allocation is the solution to:[5]

$$\max_{\{w,\mu,\bar{y}\}} V_W \equiv \quad G(\bar{y})(U(b+\mu)-B)+(1-G(\bar{y}))U(w),$$

subject to the resource constraint:

$$V \equiv -G(\bar{y})\mu + \int_{\bar{y}}^{1} y \, dG(y) - (1-G(\bar{y}))w = I.$$

From the first–order conditions, it follows that:

$$w^* = b + \mu^*. \tag{9}$$

$$\bar{y}^* = b - \frac{B}{U'(w^*)}. \tag{10}$$

Given $\bar{y}^*$, the levels of w^* and μ^* are determined by condition that the resource constraint holds with equality.

Condition (9) shows that marginal utility is equalized across employment and unemployment. Because $B > 0$ however, this implies that utility is lower when unemployed.

Condition (10) shows that the threshold level of productivity, $\bar{y}^*$, is lower than the production-efficient level b.

2.2 Implementation

As before, assume that the state first chooses taxes and benefits, the firms then enter and offer a wage to workers, and, finally, productivity is realized. Consider the following implementation of the optimal allocation, working backwards in time.

At *Stage 3*, the threshold productivity below which the firm lays a worker off is given by:

$$\bar{y} = w + \tau - f. \tag{3}$$

[5] In deriving the optimal allocation, we again ignore the constraint that y is only observed by the firm. Again, we show below that this allocation can be implemented.

At *Stage 2*, the wage must satisfy the free entry condition:

$$V_F \equiv \quad -G(\bar{y})f + \int_{\bar{y}}^{1} y \, dG(y) - (1 - G(\bar{y}))(w + \tau) = I.$$

Consider thus the *problem faced by the government* in choosing taxes and unemployment benefits at *Stage 1*. From equations (3) and (10), it follows that, to induce firms to take the socially-optimal layoff decision $\bar{y}^* = b - B/U'(w)$, the following condition must hold:

$$f - \tau = \mu + \frac{B}{U'(w)}.$$

The net fiscal cost to the firm of laying off a worker must exceed the unemployment benefits paid to the worker by an amount which depends on the cost of becoming unemployed.

The other condition on taxes and benefits comes from the government budget constraint:

$$-G(\bar{y})(\mu - f) + (1 - G(\bar{y}))\tau = 0.$$

Combining these two conditions gives:

$$f = \mu + \frac{B}{U'(w)}, \qquad \tau < 0 \tag{11}$$

The layoff tax must exceed unemployment benefits, implying, for budget balance, a negative payroll tax.

We summarize our results in Proposition 2.

Proposition 2. *(i) In the presence of limits to insurance, the threshold productivity in the socially efficient allocation is lower than the production-efficient level ($\bar{y}^* < b$), leading to a lower layoff rate than in the benchmark.*
(ii) Unemployment benefits, μ, must be financed by a combination of layoff taxes which exceed these benefits ($f > \mu$) and of negative payroll taxes, ($\tau < 0$). Put another way, the contribution rate must now be greater than one.

2.3 Interpretation and discussion

• The intuition for the two parts of Proposition 2 is straightforward: To the extent that unemployment implies a loss in utility, it is optimal to reduce its incidence, and thus to have a lower productivity threshold than the production-efficient level. This is achieved by increasing the cost of layoffs for firms, thus by having higher layoff taxes. Thus, the higher the utility cost of unemployment, the lower the layoff rate, the larger the layoff taxes: The lower layoff rate serves as a partial substitute for unemployment insurance.[6]

• We have examined the case where unemployment leads to a non-pecuniary loss in utility. The limits to insurance may come instead from incentives.

Consider for example a modification of the benchmark based on shirking. Once hired, but before productivity is revealed, the worker decides whether to shirk or not. Shirking brings private benefits B but results in zero productivity and thus a layoff. Shirking is unobservable. Thus, to prevent shirking, the following condition must hold:

$$(1 - G(\bar{y}))(U(w) - U(b + \mu)) \geq B. \tag{12}$$

The expected utility gain from being employed relative to being unemployed must exceed some value B. In that case, the optimal threshold is given by:

$$\bar{y} = b + [w - (b + \mu) - \frac{U(w) - U(b + \mu)}{U'(w)}]$$

so $\bar{y}$ is lower than the production-efficient level if workers are risk averse. Thus, again, limits to insurance lead to lower layoffs and higher layoff taxes.[7]

An alternative rationalization comes from the need to motivate the unemployed to search. While we cannot formally analyze this case in our one-period model, search incentive constraints are likely to lead however to results similar to those we have derived. The difference in utility between unemployment and employment has to be sufficient to induce search effort. A full treatment would however require a dynamic model, and we cannot provide it here.[8]

• The result that payments by firms in case of layoffs must be larger than payments to the laid–off workers implies that the optimal allocation cannot be achieved just by severance

[6] This "overemployment" result is closely related to the conclusions of the "implicit contract" literature, in particular Baily (1974), Azariadis (1975), Akerlof and Miyazaki (1980).

[7] Under the more general assumption that shirking does not yield zero productivity but instead shifts the distribution of y from $G(\cdot)$ to $H(\cdot)$, with $G(\cdot)$ stochastically dominating $H(\cdot)$, results are however less clear cut. In the absence of further restrictions, it is not necessarily the case that $\bar{y}$ is less than b, equivalently that f is greater than μ.

[8] The challenge here is to extend the research on optimal unemployment insurance—which focuses on the optimal size and timing of benefits (in particular Shavell and Weiss (1979), Hopenhayn and Nicolini (1997), Werning (2002))—to a model where the destruction margin is endogenous.

payments—which imply equal payments by firms and payments to workers. Thus, implementation requires the presence of a third party.[9] We have taken this third party to be the state, collecting layoff taxes and (negative) payroll taxes, and paying unemployment benefits to workers. Formally, what is needed is a pooling or insurance agency, collecting payments from firms that layoff, paying unemployment benefits to workers, and distributing the difference to the remaining firms. In this case, firms who join are better off. The agency may therefore be private and participation voluntary.

The issue arises however of whether the optimal allocation can be implemented through a combination of unemployment benefits—paid either by the state or by a pooling agency—and severance payments from firms. This was indeed the case in the benchmark. It is no longer the case here. We now examine this issue more closely.

2.4 Severance payments versus unemployment insurance.

A recurrent theme of the insurance literature is that the insurer must be wary of the externality imposed by supplemental insurance contracts (Pauly (1974)). For this reason, insurance companies often demand exclusivity and managerial compensation contracts prevent executives from undoing their incentives through insider trading or derivatives contracts with financial institutions. In the context of this paper, this raises the issue of whether insurance can be delivered through a combination of unemployment benefits and severance payments.

In the benchmark, the optimal allocation provided full insurance to workers, so the issue did not arise. Intuition suggests, and analysis confirms, that, in that case, firms would not want to undo the full insurance provided by the state by overinsuring the worker (severance pay) or underinsuring her (asking the worker to return some of the unemployment benefits, assuming this were feasible). But the issue arises here.

To see this, return to Stage 2 and allow firms to offer contracts which specify both a wage w *and* a severance payment, μ_F. The expected utility of workers is given by:

$$V_W \equiv G(\bar{y})(U(b + \mu_F + \mu) - B) + (1 - G(\bar{y}))U(w),$$

[9] This is an example of the general proposition (for example Holmström (1982)) that, when parties in a "team" are subject to incentive problems, there is typically a need for a "budget breaker", such as an insurance company or the state.

And the free entry condition is given by:

$$V_F \equiv \quad -G(\bar{y})(f+\mu_F) + \int_{\bar{y}}^{1} y \; dG(y) - (1-G(\bar{y}))(w+\tau) = I,$$

with $\bar{y} = w + \tau - f - \mu_F$

Now, starting from $\mu_F = 0$, and assuming the economy is at the optimal allocation, so $w = b + \mu$, consider the effects of a small increase in μ_F, $d\mu_F$ together with a decrease in the wage $dw = -(G(\bar{y})/(1-G(\bar{y})))d\mu_F$ so as to satisfy the free entry condition. It follows that

$$dV_W = \frac{g(\bar{y})}{1-G(\bar{y})} \; B \; d\mu_F > 0.$$

Firms therefore have an incentive to offer more insurance than required in the optimal allocation. The reason why is that increasing μ_F has two effects on the expected utility of workers. First, it creates a wedge between marginal utility when employed and unemployed; starting from the optimal allocation, this effect is of second order. The other is that it reduces the probability of a layoff; because the loss in utility from becoming unemployed is equal to $U(w) - U(b+\mu) + B = B$, this effect is of first order and dominates the first. When firms increase μ_F however, they decrease layoffs, and given that layoff taxes exceed unemployment benefits paid by the state, they impose a negative externality on the state. This is why, in the end, letting firms freely choose severance payments is suboptimal.

To summarize, in the presence of limits to insurance, the optimal allocation can be implemented by a state or by a private unemployment agency. But this agency must demand exclusivity, or else, mandate a ceiling for severance payments by firms. Otherwise, there will be overprovision of insurance, and a suboptimal allocation.

3 Shallow pockets

In our benchmark, firms were risk neutral and had deep pockets. These assumptions are again too strong. Even in the absence of aggregate risk, the owners of many firms, especially small ones, are not fully diversified, and thus are likely to act as if they were risk averse. And, even if entrepreneurs are risk neutral, information problems in financial markets are likely to lead to restrictions on the funds available to firms. In this section, we focus on the implications of limited funds.

Perhaps the simplest way of capturing the idea that firms have limited funds is to assume that each entrepreneur starts with assets $I + \bar{f}$, where $\bar{f} \geq 0$ is therefore the free cash flow available to the firm after investment. We explore the implications of this assumption, and discuss a number of extensions later.[10]

3.1 The optimal allocation

The government budget constraint (7), the threshold condition (3), and the condition that payments by the firm in case of layoff cannot exceed free cash flow ($f \leq \bar{f}$), can be combined to give the following constraint on $\bar{y}, w$ and μ:[11] [12]

$$G(\bar{y})\mu - (1 - G(\bar{y}))(\bar{y} - w) \leq \bar{f}.$$

Therefore, the optimal allocation is the solution to:

$$\max_{\{w,\mu,\bar{y}\}} V_W \equiv \quad G(\bar{y})U(b+\mu) + (1 - G(\bar{y}))U(w),$$

subject to the resource constraint:

$$V \equiv -G(\bar{y})\mu + \int_{\bar{y}}^{1} y \, dG(y) - (1 - G(\bar{y}))w = I,$$

and the additional constraint:

$$G(\bar{y})\mu - (1 - G(\bar{y}))(\bar{y} - w) \leq \bar{f}.$$

From the first-order conditions, it follows that the worker still receives full insurance:

$$w^* = b + \mu^*.$$

Furthermore, if the second constraint is binding (that is, if $\bar{f}$ is less than the layoff tax in

[10] In this section, it is important that y be cash (rather than, say, learning experience and so on), so it can be used to pay wages and taxes.

[11] One may wonder whether allowing for job creation subsidies/taxes in addition to payroll and layoff taxes might alleviate the shallow pocket constraint, and improve the allocation. This is not the case. Subsidies, even if allowed in the government budget constraint, would not appear in the equation below.

[12] This constraint is derived as follows. First rewrite the threshold condition as $\tau = \bar{y} - w + f$ and replace τ in the government budget constraint to get $-G(\bar{y})(\mu - f) + (1 - G(\bar{y}))(\bar{y} - w + f) = 0$: For a given $\bar{y} - w$, the lower f, the lower is μ. Reorganize and use $f \leq \bar{f}$ to get the equation in the text.

the optimal allocation derived in Section 1), threshold productivity is given by:

$$\bar{y}^* = b + \frac{(\mu^* - \bar{f})}{(1 - G(\bar{y}^*))} > b. \tag{13}$$

By limiting payments by firms in case of layoff, the shallow pocket constraint prevents the state from achieving the production-efficient threshold, and the layoff rate is now higher than the production-efficient level. The tighter the shallow pocket constraint—i.e. the lower $\bar{f}$—then the larger $(\mu^* - \bar{f})$, the higher $\bar{y}^*$, and so, the larger the layoff rate. The levels of $\bar{y}^*$, w^*, and μ^* are determined by (13), the full insurance condition, and the condition that the resource constraint holds with equality.

3.2 Implementation

If the shallow-pocket constraint is binding, the state chooses the highest feasible layoff tax $f = \bar{f}$. Given unemployment benefits μ^*, the government budget constraint then implies:

$$\tau = \frac{G(\bar{y}^*)}{1 - G(\bar{y}^*)}(\mu^* - \bar{f}) > 0.$$

As unemployment benefits exceed layoff taxes, payroll taxes must be positive.

The threshold productivity chosen by firms is therefore given by:

$$\bar{y}^* = b + \mu^* + \tau - f = b + \mu^* + \frac{G(\bar{y}^*)}{1 - G(\bar{y}^*)}(\mu^* - \bar{f}) - \bar{f} = b + \frac{(\mu^* - f)}{1 - G(\bar{y}^*)}.$$

This is the same expression as in (13), and so, layoff and payroll taxes indeed implement the optimal allocation. The derivation shows that we can think of the shallow pocket constraint as affecting the threshold productivity level directly (through the limit on the layoff tax) and indirectly (through the need for positive payroll taxes); both the lower layoff tax and the higher payroll tax reduce the fiscal cost of layoffs for firms, and thus lead to a layoff rate higher than the production-efficient level.

By the same argument as before, the resource constraint implies that workers receive the optimal w^* and μ^*.

We summarize the results in Proposition 3.

Proposition 3. *In the presence of shallow pockets, workers remain fully insured ($w^* = b + \mu^*$). The threshold productivity is higher than the production-efficient level ($\bar{y} > b$),* *leading to a higher layoff rate than in the benchmark.*

This allocation can be implemented by the government choosing unemployment benefits μ^**, and financing them partly through layoff taxes* $\bar{f}$*, and partly through payroll taxes,* $\tau > 0$*. Put another way, the implementation implies now a contribution rate smaller than one.*

3.3 Interpretation and discussion

• Proposition 3 has two important aspects:

The first is that the presence of limited funds does not prevent full insurance. The reason is that the state can raise the required funds through higher payroll taxes, and by implication a lower equilibrium wage, without violating the shallow pocket constraint.[13]

The second is that the presence of limited funds prevents the state from achieving the production-efficient layoff rate. Limits on layoff taxes affect layoffs in two ways, directly, and indirectly, through the higher payroll taxes required to finance benefits.

• Given that layoff taxes are less than unemployment benefits, it follows that, again, the optimal allocation cannot be achieved by just relying on severance payments by firms. As in Section 2, implementation can be achieved by a pooling agency, receiving contributions $\bar{f}\ G(\bar{y}^*)$ from firms that layoff and contributions $\tau(1 - G(\bar{y}^*)$ from those that do not, and paying unemployment benefits $G(\bar{y}^*)\mu^*$ to laid–off workers. Firms have an incentive to join, and the agency may therefore be private. Also, in this case, because workers are fully insured, the coinsurance problem we looked at in the previous section does not arise: Insurance can be provided by a mix of severance payments by firms, and insurance benefits from the state or the pooling agency.

The assumption of exogenous free cash flow associated with each job is clearly too strong:

• One reason why $\bar{f}$ may be endogenous is that the firms may not want to have deep pockets even if they can. This arises for example, if, in contrast to the maintained assumption of

[13] To see this, consider an allocation where $w > b + \mu$. Now, consider a decrease in the wage of $\Delta w < 0$ and an equal increase in payments by firms to the state, $\Delta\tau$. This change affects neither the threshold condition nor the firm's profit. Use these increased payments to increase unemployment benefits by $-[(1 - G(\bar{y}))/G(\bar{y})]\Delta w$. Together, these changes imply a change in utility of $[-(1-G(\bar{y}))U'(w) + (1-G(\bar{y}))U'(b+\mu)](-\Delta w) > 0$. Thus, welfare can be improved until workers are fully insured.

this paper, the government policy is set after rather than before firms invest. Suppose, for example, that the state cannot commit and sets (τ, f, μ) after firms have invested and hired workers, but before they learn the productivity of the match. In this case, firms will obviously choose to be “judgment proof”, i.e. to have no assets left in case of layoff, so $\bar{f} = 0$. The threshold is then given by: $\bar{y}^* = b + \mu^*/(1 - G(\bar{y}^*))$. The high threshold, and by implication, the high layoff rate, reflects two distortions, one coming from zero layoff taxes and the other from positive payroll taxes.

To the extent that the amount of free cash is endogenous, this point suggests an important and general role for the state, namely to make sure that firms (or, more generally, the pooling agency) cannot become “judgment proof”. This may require asking firms to post collateral when creating a job.

Tirole (2006) explores two other extensions. We summarize his findings here:

• We have assumed that each firm employs only one worker. In reality, firms typically employ many workers. They may have enough free cash flow to pay layoff taxes so long as they layoff a small number of workers, but not if they have to layoff workers on a large scale. The issue is how the layoff tax schedule should look like in this case. Tirole (2006) shows that, even if it is feasible to set layoff tax rates equal to the benchmark case until the number of layoffs is such that the free cash flow constraint binds, this is not optimal. Such a tax schedule would lead firms to face a zero cost of laying off workers when the constraint starts binding. It is better to have lower tax rates from the start, even before the constraint binds. The general lesson is that, to the extent that firms have limited funds, the layoff tax rate should be lower, even if the constraint does not yet bite.

• We have taken free cash flow, $\bar{f}$, as exogenous. But it may in fact respond to policy, in particular to changes in layoff taxes. Tirole (2006) looks at what happens when the shallow pocket constraint arises endogenously in an agency-cost model of entrepreneurs and investors. It finds that, in that context, the divergence of objectives between entrepreneurs and investors distorts the destruction margin, leading to too many layoffs relative to the production-efficient level. The optimal policy in this case implies a contribution rate less than one. The larger the managerial rents, or the lower the wealth of entrepreneurs, the closer is the contribution rate to one. In other words, the worse the capital market imperfections, the lower the contribution rate.

4 Ex-post wage bargaining

Our benchmark embodied the assumption that wages were set ex-ante, i.e. at the time of hiring. This had the implication that, by offering unemployment insurance to risk averse workers, a firm could not only offer a lower wage, but actually lower its expected labor costs.

To some extent however, there is always some room for ex-post bargaining. When this is the case, a firm which has to pay a layoff tax if it lays a worker off is in a weaker bargaining position vis-á-vis that worker; a worker who will receive unemployment benefits if laid off is in a stronger position. The layoff tax, the severance payments, and the provision of unemployment benefits all lead to *higher*, not lower, wages, and thus increase labor costs.

In this section, we therefore modify our earlier assumption about wage setting, assume ex-post wage bargaining instead, and characterize the optimal allocation and its implementation under ex-post wage bargaining. In order to avoid the complexities attached to bargaining under incomplete information, we assume, in this section, that both the firm and the worker observe productivity ex post. Thus "ex-post wage bargaining" includes the assumption of symmetric information between worker and firm ex post.

4.1 A formalization of wage bargaining

A simple way of capturing ex-post wage bargaining is to assume that bargaining takes place after productivity is realized and that workers obtain a proportion β of the private surplus from the match, so:[14]

$$w(y) = b + \mu + \beta(y - \tau + f - b - \mu). \tag{14}$$

The higher the layoff tax, or the lower the payroll tax, or the higher the unemployment benefits, the higher is the wage.

The threshold value for productivity, $\bar{y}$, is given by the condition that $w(\bar{y}) + \tau - \bar{y} = f$.

[14] A simple game that delivers this outcome is a two-stage game. In stage 1, the worker makes a wage offer to the firm. The firm can either accept the offer or turn it down. If it turns it down, the wage is set in stage 2, either by the worker with probability β, or by the firm with probability $1 - \beta$. In stage 2, the highest wage the firm will accept, and therefore the wage offered by the worker, is equal to $y - \tau + f$. The lowest wage the worker will accept, and therefore the wage offered by the firm, is equal to $b + \mu$. Thus, the expected wage in stage 2 is given by $\beta(y - \tau + f) + (1 - \beta)(b + \mu)$. This implies that, in stage 1, the worker will make the highest offer acceptable by the (risk neutral) firm, i.e. an offer of $w(y) = \beta(y - \tau + f) + (1 - \beta)(b + \mu) = w(y) = b + \mu + \beta(y - \tau + f - b - \mu)$.

Using the expression for the wage and rearranging:

$$\bar{y} = b + \mu + \tau - f. \tag{15}$$

Note that the threshold is privately efficient.

Expression (15) allows us to rewrite the wage schedule as:

$$w(y) = (b + \mu) + \beta(y - \bar{y}). \tag{16}$$

The wage paid to the marginal worker—the worker in a job with productivity equal to the threshold level—is equal to $(b + \mu)$, the wage equivalent of being unemployed plus unemployment benefits and severance payments. The wage then increases with β times the difference between productivity and threshold productivity.

4.2 The optimal allocation

The optimal allocation solves the same problem as in the benchmark, subject to the additional constraint that the wage is no longer a decision variable, but is instead given by equation (16):

$$\max_{\{\mu,\bar{y}\}} G(\bar{y})\, U(b + \mu) + \int_{\bar{y}}^{1} U(w(y))\, dG(y),$$

subject to the resource constraint:

$$-G(\bar{y})\mu + \int_{\bar{y}}^{1} (y - w(y))\, dG(y) = I,$$

and the wage relation

$$w(y) = b + \mu + \beta(y - \bar{y}).$$

The solution can be characterized as follows: The optimal threshold level of productivity is implicitly defined by:

$$\bar{y}^* = b + \frac{\beta(1 - G(\bar{y}^*))}{g(\bar{y}^*)} \left[1 - \frac{E[U'|y \geq \bar{y}^*]}{E\, U'}\right], \tag{17}$$

where $E[U'|y \geq \bar{y}^*] \equiv (\int_{\bar{y}^*}^{1} U'(w(y))dG(y))/(1 - G(\bar{y}^*))$ is the expected value of marginal utility if employed, and $E\, U' \equiv G(\bar{y}^*)U'(b + \mu) + \int_{\bar{y}^*}^{1} U'(w(y))\, dG(y)$ is the unconditional expected value of marginal utility.

So long as β is strictly positive, and workers strictly risk averse, then the expected marginal utility if employed is less than the unconditional expected marginal utility, and so $\bar{y}^*$ is greater than b. The layoff rate exceeds the production-efficient level.

Given $\bar{y}$, μ and by implication the wage schedule $w(y)$ are determined by the resource constraint above.

4.3 Implementation

Replacing $\bar{y}$ from (17) in the expression for the threshold decision of firms, (15), gives the first relation between f and τ:

$$f - \tau = \mu - \beta \frac{(1 - G(\bar{y}^*))}{g(\bar{y}^*)} \left[1 - \frac{E[U'|y \geq \bar{y}^*]}{E\, U'} \right]. \tag{18}$$

The other relation is given, as before, by the budget constraint, equation (7). As the second term on the right side of equation (18) is now negative, $f - \tau < \mu$. Together with the government budget constraint, this implies $f < \mu$ and so, a contribution rate below one.

Note that for $\beta = 0$, i.e. if workers have no bargaining power, then we obtain the same characterization as in the benchmark: $f = \mu$ and $\mu = w - b$. As β becomes positive, and the wage schedule is now increasing in productivity, it can be shown that the equations above imply:

$$\frac{df}{d\beta} < \frac{d\mu}{d\beta} < 0.$$

That is, both the unemployment benefit and the layoff tax decrease as the workers acquire more bargaining power, and the layoff tax falls faster, leading to a decreasing contribution rate.

Also, given the unemployment benefits and the layoff and payroll taxes characterized above, it is clear that firms will not want to offer additional severance payments, as these only increase labor costs one for one.[15]

[15] Let μ_F denote severance payments. The new wage function, $\widehat{w}(y)$ is given by $\widehat{w}(y) = b + \mu + \mu_F + \beta(y - \tau + f - b - \mu) = w(y) + \mu_F$. The threshold value for productivity, $\overline{y}$, is given by the condition that $\widehat{w}(\overline{y}) + \tau - \overline{y} = f + \mu_F$; that is, the threshold $\overline{y} = b + \mu + \tau - f$ is independent of severance pay. So severance pay shifts the wage schedule up one for one.

We summarize our results as follows:

Proposition 4. *When wages are set through ex–post bargaining, utility for the marginal worker is the same as for the unemployed workers, and workers with higher productivity receive a higher wage. These outcomes are independent of the state's policy choices.*

The threshold productivity level in the socially efficient allocation is higher than the production-efficient level: ($\bar{y}^* > b$).

This optimal allocation can in turn be implemented through a combination of layoff taxes, payroll taxes, and unemployment benefits. Layoff taxes must be less than unemployment benefits, with payroll taxes used to make up the difference; equivalently, the contribution rate must be less than one.

4.4 Interpretation and discussion

• The way to understand the results in proposition 4 is to start from the wage schedule implied by ex-post wage bargaining, equation (16).

Under this wage schedule, the scope for unemployment insurance is extremely limited: For a given $\bar{y}$, an increase in unemployment benefits increases all wages by an amount equal to unemployment benefits. Put another way, variations in μ affect the level of income for both the employed and the unemployed. For a given $\bar{y}$ and thus given overall production, the resource constraint for the economy and the zero profit condition for firms in turn determine the level of unemployment benefits.

Under these conditions, why not simply achieve production efficiency and set layoff taxes equal to unemployment benefits?[16] This is because, under the wage schedule (16), employment is associated with wage uncertainty. Thus, starting from the production-efficient level, a small increase in threshold productivity has a zero first order effect on efficiency, but decreases the uncertainty faced by the worker and has a first-order effect on expected utility. Thus, the optimal layoff rate is higher than in the production-efficient level; this in turn requires layoff taxes to be lower than unemployment benefits.

• In our model with homogenous firms and free entry, for a given $\bar{y}$, the level of unemployment benefits μ is fully determined by the resource constraint and the free entry

[16] This is indeed the result obtained by Mortensen and Pissarides (2003), in a search model with ex-post wage setting, and with risk neutral workers (and lump sum taxation). In that model, it is optimal for the planner to choose layoff taxes equal to unemployment benefits.

condition. Any lower level would lead to positive net profits for new entrants; any higher level would lead to no entry. There is therefore no need for the state to force firms to provide unemployment insurance.

This is clearly a special feature of the model. If for example, there was heterogeneity either in the distribution of shocks facing firms, or in the fixed cost associated with entry, a higher μ would lead to lower entry by firms, and thus some ex-ante unemployment. Thus, there would be a trade-off between unemployment insurance (and by implication the income of unemployed and employed workers) and unemployment. Higher μ would lead to higher income for workers, lower entry by firms, and some ex-ante unemployment. In this case, it is not clear that firms, either individually or collectively, would choose a level of unemployment insurance identical to that preferred by the state. If the two differed, the level of insurance would have to be mandated by the state. For a given μ however, the choice of the layoff tax would follow the same logic as above.

• The main conclusion we draw from this section is the importance of wage setting for the design of unemployment insurance and employment protection. In contrast to the benchmark, under pure ex-post wage bargaining, unemployment benefits do not provide insurance, but rather determine the general level of income of workers—equivalently the profits of firms. Production efficiency suggests that they should be primarily financed through layoff taxes. Insurance considerations imply some deviation from this rule, in the direction of higher layoffs, thus of lower layoff taxes.

5 Heterogeneity

We have assumed so far that all workers and all firms were ex-ante identical. In reality, they clearly are not. Firms differ in the distribution of productivity shocks (or, more generally, the distribution of productivity and relative demand shocks) they face, and in their initial assets. Workers also differ in the distribution of productivity and the utility associated with being unemployed. We study in this section the implications of heterogeneity in productivity, both on the firm and on the worker side, both observed and unobserved.

5.1 Heterogeneity of firms

A worry often expressed by policy makers is that, if some firms have higher layoff rates than others, a layoff tax will penalize them more, and this may be undesirable.

To explore this idea, suppose there are two types of firms, "strong" and "weak", which differ in their productivity distributions. The productivity of "strong firms" is drawn from cumulative distribution $G_H(\cdot)$ and that of "weak firms" from distribution $G_L(\cdot)$.[17] The distribution function of strong firms stochastically dominates that of weak firms: for all y in $(0,1)$, $G_H(y) < G_L(y)$. The fraction of strong firms is equal to ρ.

Note that given the distribution assumptions and similar taxation of both types, strong firms will have higher expected profits than weak firms. We assume that the number of strong firms is fixed. Thus, at the creation margin, the free entry condition is relevant for weak firms only; the strong firms have pure rents. By contrast, both types may lay workers off, so the destruction margin is relevant for both types of firms. Assume that the state maximizes the welfare of workers (allowing the state to also put some weight on the positive rents earned by the owners of strong firms would not alter the results.)

Assume, first, that heterogeneity is *observable*, so the state can treat the two types of firms differently. The optimal allocation, and the optimal benefit and tax policy, are then easy to characterize: It is for the state to (1) offer job creation subsidies to weak firms. This relaxes the free entry condition for weak firms, allowing them (and the strong firms) to pay higher wages, and thus transferring rents from strong firms to workers in the form of higher wages; (2) offer full insurance to workers; (3) finance unemployment benefits through a net contribution rate—defined as the layoff tax minus the payroll tax, divided by unemployment benefits—equal to one. (Note that both layoff and payroll taxes must be higher than in the benchmark, as extra revenue is needed to finance job creation subsidies to weak firms.)[18] In this way, just as in the benchmark, the state provides full insurance to workers, while maintaining production efficiency. Observable heterogeneity does not, by itself, provide a reason for not having firms fully internalize the costs of unemployment insurance.

Suppose now that heterogeneity is *unobservable*, so the state is unable to tell the two types of firms apart. The state can either set a uniform policy (τ, f, μ), or offer menus. We consider first the case of a uniform policy, and discuss the alternative below.

The fact that the free entry condition is relevant only for weak firms makes it more difficult to adopt the optimal allocation/implementation approach we have followed until

[17] We assume that the weak firms are not too unproductive. Namely we assume that $\int_b^1 (y-b)dG_L(y) \geq I$: Weak firms have positive NPV for the production-efficient threshold.

[18] An alternative way of implementing the same allocation would be to use lump sum taxes rather than payroll taxes. In this case the contribution rate as well as the net contribution rate would be equal to one.

now. We take instead the more "pedestrian " route of solving the optimization problem of the state given firms' behavior.

The state's optimization problem is given by:

$$\max_{\{\tau,f,\mu,w,\bar{y}\}} \quad [\rho G_H(\bar{y}) + (1-\rho)G_L(\bar{y})]\, U(b+\mu) + [\rho(1-G_H(\bar{y})) + (1-\rho)(1-G_L(\bar{y}))]\, U(w),$$

subject to the free-entry condition for weak firms

$$-G_L(\bar{y})f + \int_{\bar{y}}^{1} (y - w - \tau)dG_L(y) = I,$$

the government budget constraint

$$\begin{aligned} -[\rho G_H(\bar{y}) + (1-\rho)G_L(\bar{y})]\, (\mu - f) & \\ +[\rho(1-G_H(\bar{y})) + (1-\rho)(1-G_L(\bar{y}))]\, \tau &= 0, \end{aligned}$$

and the threshold productivity condition, which is the same for weak and strong firms:

$$\bar{y} = w + \tau - f.$$

The solution can then be characterized as follows:

- The state fully insures workers: $w = b + \mu$.
- If the proportion of weak firms is not too low (ρ is not too high), both weak and strong firms operate. The threshold level of productivity is given by:

  $$\bar{y} = b + \frac{\rho(G_L(\bar{y}) - G_H(\bar{y}))}{(\rho g_H(\bar{y}) + (1-\rho)g_L(\bar{y}))}.$$

  By the definition of weak and strong firms, $G_L(\bar{y}) > G_H(\bar{y})$. So, unless $\rho = 0$, the threshold level is higher than the efficient level.

  The solution is implemented by using a contribution rate below unity, with the rest of unemployment benefits being financed by payroll taxes. The contribution rate is less than one.
- If the proportion of weak firms is close to zero (ρ is close to one), then weak firms do

not operate—at the cost of some ex-ante unemployment. Threshold productivity for strong firms is equal to the production-efficient level.

The solution is implemented by using a net contribution rate equal to one for strong firms (payroll taxes must be positive to finance unemployment benefits for the ex-ante unemployed. So, by implication, layoff taxes must be higher as well to make the net fiscal cost of a layoff equal to unemployment benefits)[19].

We summarize our results in the following proposition:

Proposition 5. *Suppose that there are two types of firms: "strong" or "weak". Strong firms have productivity distribution $G_H(\cdot)$, weak firms $G_L(\cdot)$, and $G_L(y) > G_H(y)$ for all y in $(0, 1)$.*

If heterogeneity is observed by the state, the optimal allocation implies transfers from strong to weak firms, full insurance of workers, and production-efficiency. This allocation is implemented through job creation subsidies to weak firms, financed by payroll and layoff taxes, and a unit contribution rate on both weak and strong firms.

If this heterogeneity is unobserved, if the proportion of weak firms is not too low, and if the state relies on a uniform policy, then the optimal policy is still to fully insure workers ($w = b + \mu$), but choose a contribution rate less than one.

• The intuition for why the contribution rate is less than one in the case of unobserved heterogeneity is as follows: Because strong firms have rents, a cross subsidy from strong to weak firms improves workers' welfare. If the state can distinguish between weak and strong firms, it can achieve this redistribution through job creation subsidies to weak firms. If it cannot, it can partly achieve this redistribution by reducing layoff taxes: Weak firms lay workers off more than strong firms and so benefit more from lower layoff taxes. But this now comes at the cost of some production inefficiency.

• Can the state do better by offering *menus* and letting firms self select? Appendix 1 shows that while offering a menu improves efficiency, the solution carries the main characteristics of the uniform policy, namely cross-subsidization of weak firms by strong firms, achieved through a contribution rate for weak firms below one.

[19]This result is related to the result in Cahuc and Jolivet (2003) where the need to finance a public good also leads to higher layoff and payroll tax rates.

• Unobserved heterogeneity (and the assumption that the government maximizes the welfare of workers) provides a clear case where implementation cannot be left to firms or to a private agency. On their own, and in the absence of other distortions, strong and weak firms will want either to offer severance payments, or join a pooling agency with contributions equal to the benefits paid to laid-off workers. To the extent that the state wants to distort the destruction margin to redistribute from strong to weak firms, it will have to impose contribution rates on firms or the private agency.

An important issue in practice is thus how much of the heterogeneity of firms is observable (allowing for the use of job creation subsidies without distorting the destruction margin) and how much is unobservable (requiring distortions at the destruction margin).

5.2 Heterogeneity of workers

Another frequently expressed worry is that some workers are more likely to be laid-off than others, and that a layoff tax will make firms more reluctant to hire them, and thus make these workers worse off.

To explore this idea, we set up a case very similar to that of firms. We assume there are two types of workers, "high-ability" and "low-ability" workers. High-ability workers have a productivity distribution given by $G_H(.)$, low-ability workers a distribution given by $G_L(\cdot)$, with $G_L(y) > G_H(y)$ for all y in $(0,1)$.[20] The fraction of workers with high ability is equal to ρ. The firms know the workers' abilities.

Assume that the state maximizes a utilitarian welfare function, i.e. a weighted average of the expected utility for high- and low-ability workers, with weights ρ and $(1-\rho)$ respectively. And assume first that heterogeneity is observable, so the state can treat the two types of workers differently. The optimal allocation, and the optimal benefit and tax policies, are again easy to characterize: It is for the state to (1) give job creation subsidies to firms that hire low-ability workers, until the wages of high- and low-ability workers are equalized; (2) to offer full insurance to workers; (3) to choose layoff taxes and payroll taxes so the net contribution rate for firms hiring either high- or low-ability workers is equal to one. In this way, just as in the benchmark, the state is able to transfer income to low-ability workers, provide full insurance to all workers, and maintain production efficiency. Observable heterogeneity of workers does not, by itself, provide a reason for not having

[20] We assume that the low productivity workers are not too unproductive. Namely we assume that $\int_b^1 (y-b)dG_L(y) \geq I$. Under observed heterogeneity, it is profitable for firms to hire low-ability workers.

firms fully internalize the costs of unemployment insurance.

Suppose however that heterogeneity is unobservable, and that the state chooses a uniform policy (τ, f, μ).

We can simplify the set-up of the optimization problem, by noting that workers of each type will be fully insured, and that, because high-ability workers are more valuable to firms, they will be paid more, both when employed and when unemployed.

Suppose therefore that low-ability workers receive w when employed, and μ when unemployed, and are fully insured, so $w = b + \mu$. High-ability workers will then receive $w + \Delta$ when employed, and $\mu + \Delta$ when unemployed (so $w + \Delta = b + \mu + \Delta$), where Δ is the additional expected profit brought about to the firm by a high-ability worker. Noting that Δ does not affect the layoff decision, so the threshold productivity is the same for both types of workers, it follows that Δ is given by:

$$\Delta = \left[G_L(\bar{y}) - G_H(\bar{y})\right] f + \int_{\bar{y}}^{1} \left[y - (b + \mu) - \tau\right] \left[dG_H(y) - dG_L(y)\right].$$

Payment of Δ can be achieved by wage-indexed unemployment benefits, so a worker who is paid $w + \Delta$ receives $\mu + \Delta$ (and the firm pays correspondingly higher layoff taxes when laying a high-ability worker off).

With this characterization of wage setting, the optimal policy is the solution to:

$$\max_{\{\tau, f, \mu, \bar{y}\}} \left\{(1 - \rho)\, U(b + \mu) + \rho\, U(b + \mu + \Delta)\right\},$$

subject to the free entry condition

$$-G_L(\bar{y})\, f + \int_{\bar{y}}^{1} \left[y - (b + \mu) - \tau\right] dG_L(y) = I,$$

and the government budget constraint

$$\left[(1 - \rho)\, G_L(\bar{y}) + \rho G_H(\bar{y})\right](f - \mu) + \left[(1 - \rho)\left[1 - G_L(\bar{y})\right] + \rho\left[1 - G_H(\bar{y})\right]\right]\tau = 0,$$

where Δ and $\bar{y}$ have been defined above.

The solution has the following form:

• The state fully insures workers.

• If the proportion of low-ability workers is not too low (ρ is not too high), then both types are hired. The threshold for productivity is given by:

$$\bar{y} = b + \frac{\rho(1-\rho)[G_L(\bar{y}) - G_H(\bar{y})]\,[U'(\mu+b) - U'(\mu+b+\Delta)]}{[(1-\rho)g_L(\bar{y}) + \rho g_H(\bar{y})]\,[(1-\rho)U'(\mu+b) + \rho U'(\mu+b+\Delta)]}.$$

Note that $G_L(\bar{y}) - G_H(\bar{y})$ is positive, and so is $U'(\mu+b) - U'(\mu+b+\Delta)$. So the fraction on the right is positive, and the optimal threshold is higher than the efficient level.

This solution is implemented by using a contribution rate below unity, with the rest of unemployment benefits being financed by payroll taxes.

• If the proportion of low-ability workers is close to zero (ρ is close to one), it is then optimal to keep them unemployed, with unemployment benefits, and to choose threshold productivity for high-ability workers equal to the production efficient level.

This solution is implemented by using a net contribution rate equal to one. (Payroll taxes have to be positive to finance unemployment benefits for the ex-ante unemployed.)

We summarize our results in the following proposition:

Proposition 6. *Suppose that there are two types of workers: "high-ability" and "low-ability". High-ability workers have productivity distribution $G_H(\cdot)$, low-ability workers $G_L(\cdot)$, and $G_L(y) > G_H(y)$ for all y in $(0,1)$.*

If heterogeneity is observed by the state, the optimal allocation implies transfers to firms that hire low-ability workers, full insurance of workers, and production-efficiency. This allocation is achieved through job creation subsidies to firms that hire low-ability workers, financed by payroll and layoff taxes, and a unit contribution rate.

If this heterogeneity is unobserved, if the proportion of low-ability workers is not low, and if the state relies on a uniform policy, then the optimal policy is to fully insure workers ($w = b + \mu$), but rely on a contribution rate less than one.

• The intuition for these results parallels that for firms' heterogeneity. The state wants to transfer income from high-ability to low-ability workers. If it can distinguish between types, it can do so through job creation subsidies. If it cannot, it can partly achieve this redistribution by reducing layoff taxes: Low ability workers are more likely to be laid-off and thus benefit more from the reduction. But this now comes at the cost of some production inefficiency.

The logic of the results under firm and worker heterogeneity is clearly similar: Weak workers (firms) are more likely to be laid off (to lay off), and an incomplete internalization of the externality of the layoff on the UI fund benefits the creation margin.

• What if the state can offer menus and firms then self-select? It is straightforward to show that the state then offers two options, one aimed at firms that announce they have hired a high-ability worker, one aimed at firms that announce they have hired a low-ability worker. The first option has a net contribution rate equal to one. The second option has a net contribution rate below one. Both types of workers are fully insured. Thus, just as in the uniform case, heterogeneity leads to contribution rates below one, but in this case only for low-ability workers.

•. Like unobserved heterogeneity of firms, unobserved heterogeneity of workers (and a utilitarian objective function for the government) provides a clear case where implementation cannot be left to firms or to a private agency. To the extent that the state wants to distort the destruction margin to redistribute from strong to weak firms, it will have to impose tax rates on firms.

An important issue in practice is thus again how much of the heterogeneity of workers is observable, for example correlated with observable characteristics such as age or experience (in which case job creation subsidies can be used, together with a unit net contribution rate and no distortion at the destruction margin) and how much is unobservable.

6 Open issues, and back to policy

The policy implications of the benchmark are straightforward: If the state runs an unemployment insurance system, unemployment benefits should be financed by layoff taxes; but the same result can also be obtained by letting firms voluntarily offer severance payments.

The various deviations from the benchmark show however that the answers are unfortunately more complex. Whether, for example, layoff taxes should be greater or less than unemployment benefits depends on specific distortions, from limits to insurance to financial constraints facing firms. Whether the provision of insurance can be left to firms, either alone or through a pooling agency, depends also on specific distortions, on whether for example whether there is unobserved heterogeneity across firms or workers.

Yet our framework remains much too simple. Many issues just cannot be analyzed in our one-period model, and require a dynamic model, which we have not provided.

This raises two sets of questions. One for researchers, namely the directions in which this framework should be expanded. The other for policy makers, namely whether the analysis, as it stands, can help think about policy reforms. We take both sets of questions in turn. By its nature, this section is obviously more speculative than the previous ones; we hope it can be useful.

6.1 Open issues

Even within our one-period model, there are a number of issues still to be explored. Let us mention two.

- The first is the issue of quits versus layoffs. If we think of layoffs as triggered by productivity shocks (shocks to y), and quits as triggered by reservation wage shocks (shocks to b, or to the disutility of work—which we do not have explicitly in our model), and we think of the layoff tax as applying only in case of layoffs, this raises two sets of issues. The first is actions by firms to induce workers they would like to lay off to quit instead (harassment), and actions by workers to induce firms they would like to quit to lay them off instead (shirking). The second is actions by firms and workers together to mislabel quits and layoffs. The incentives to harass, shirk, or cooperatively misreport, depend very much in each case on the contribution rate. We have informally explored these issues in Blanchard and Tirole (2003b), but a formal treatment remains to be given.
- Another issue is the role of judges, who, in many European countries, play a central role, and are often ultimately in charge of deciding whether layoffs are economically justified or not. Clearly, the logic of our argument is that this is better accomplished through a combination of layoff taxes and severance payments, with the decision then being left to the firm. But our look at the implications of imperfections, from shallow pockets to heterogeneity, also suggests the desirability of adapting layoff taxes to particular situations. This can in principle be done through offering menus, or allowing taxes to be conditional on observable characteristics of firms, or by leaving some discretion to judges. It remains to be shown however if and when judges do in fact have the information, the ability, and the incentives, to take better and more informed decisions.

Then, and obviously so, there are dynamic issues we could not consider at all in our one–period model. Dynamic models of the labor market with risk aversion and imperfections are notoriously hard to solve. The only model we know which derives optimal institutions—defined as the optimal combination of payroll taxes, layoff taxes, job creation subsidies or taxes, and unemployment benefits—was developed by Mortensen and Pissarides (2003). However, it assumes risk neutrality and so cannot deal in a convincing way with the interaction between insurance and efficiency. (A model by Alvarez and Veracierto [2000] has risk averse workers, self-insurance as well as state-provided insurance, payroll and layoff taxes, and severance payments. While it shows (numerically) the effects of changes in some of these instruments, it does not give a characterization of optimal taxes and benefits.) We see three extensions of our model as essential:

- The first is the role and the implications of self insurance by workers (in terms of the model here, the role and implications of the endogeneity of b). (Three papers provide a useful starting point here. All three allow for self-insurance, and look at the role of state-provided insurance in the presence of other imperfections. In Hansen and Imrohoroglu (1992), moral hazard in search limits the scope for state-provided insurance. In Acemoglu and Shimer (1999, 2000), state-provided insurance affects search, which in turn affects match quality.)
- The second, and related issue, is that of the role, if any, of experience rating systems for workers, and of mandatory individual unemployment accounts such as are being considered or introduced in a number of Latin American countries.
- The third is the role of experience rating systems of firms, such as the U.S. system, as ways of implementing the collection of layoff taxes over time. This requires a careful look not only at the dynamic problem of the firm, but at the exact nature of the financial constraints that it faces.

6.2 Policy implications

Can our analysis, as it stands, help think about policy reforms? Our answer is a careful yes. Indeed, the intellectual origin of the paper was a request to define the contours of employment protection reform in France; our conclusions at the time were presented in Blanchard and Tirole [2003a,b].

- *On the general architecture*

The main implication of the benchmark was that unemployment benefits should be financed by layoff taxes. This conclusion was qualified in the different extensions. But, in no case, was the reference point taxation through payroll taxes.

This conclusion stands in striking contrast with the way unemployment insurance is financed in Europe. In all European countries, benefits are financed through payroll taxes, not layoff taxes. As we have seen, payroll taxation goes the "wrong way" for two reasons: The absence of layoff taxes leads firms not to internalize the costs of insurance, and, by increasing labor costs, the presence of payroll taxes gives incentives to firms to lay workers off.

Perhaps because of these perverse financial incentives, a number of countries have put in place a system of employment protection based on heavy judicial intervention. In a number of countries, judges have the authority to decide whether a layoff is justified on economic grounds or not.

This strongly suggests that at least a partial shift from payroll to layoff taxes, accompanied by limits on judicial intervention, would lead to a better allocation. Firms, once forced to internalize the costs of unemployment insurance, are in a much better position than judges to assess whether layoffs are economically justified. We emphasized this conclusion in Blanchard Tirole (2003a).

- *On the optimal contribution rate*

While, in the benchmark, layoff taxes were simply equal to unemployment benefits, the extensions offered arguments for higher or lower layoff taxes: Limits to insurance implied the desirability of higher layoff taxes, so as to decrease the incidence of layoffs. Financial constraints on firms suggested in turn the desirability of lower layoff taxes. So did the presence of ex-post bargaining. So did unobserved firm or worker heterogeneity, to the extent that weak firms or low ability workers benefited more from lower layoff taxes.

Signing the net effect of these distortions with any confidence, and thus, recommending a given contribution rate, is in effect impossible: Our model is much too simple to be calibrated, and the empirical evidence to assess the importance of the relevant distortions is, for the most part, missing. We shall limit ourselves to a few remarks:

Our own instincts (but they are barely more than this) are that the first distortion, limits to insurance, is important. The evidence cited earlier that layoffs come with large non-pecuniary losses, together with the limits to unemployment insurance from moral hazard in search, suggest that the minimal utility loss associated with being laid-off can be substantial; if this is the case, decreasing the incidence of layoffs through a contribution rate above one may well be justified. This also suggests the importance of reforms of unemployment insurance which allow for better insurance while still providing incentives to search. Recent reforms, which make unemployment benefits more explicitly conditional on search and acceptance of jobs if available, go in that direction. If successful, they can bring not only better insurance, but also lower employment protection and lower production inefficiencies.

The quantitative importance of the second distortion, financial constraints, requires having a better sense of the financial situations of the firms that lay workers off. Available evidence suggest that the large majority of layoffs take place in financially healthy firms. If this is the case, lower contribution rates for categories of firms that are known to be more financially fragile, rather than a lower contribution rate across the board, may be the better approach. In practice, this may mean for example lower contribution rates for small firms, which are known to be financially more fragile than larger ones (see for example Gertler and Gilchrist (1994)).

We do not know how much weight to give to the implications of ex-post wage setting for the contribution rate. It is clear that, in that context, the role of unemployment insurance is much more limited. We prefer to take as the main conclusion that, under ex-post wage setting, production efficiency still requires a unit contribution rate and, that, while insurance considerations suggest deviations from this rule, this should remain the reference rate.

In the last case, heterogeneity, the main empirical issue is whether it is mostly observable or unobservable by the state. If observable, job creation subsidies to specific firms or associated with the hiring of specific types of workers are the appropriate solution, together with unit net contribution rates. If unobservable, menus represent an improvement upon uniform taxation.

- *On unemployment accounts*

Many countries, especially in Latin America, have moved in the direction of systems of "unemployment accounts". While they vary in their details, this system requires workers to contribute to savings accounts when employed, and draw from these accounts if they become unemployed—or under other specified circumstances, such as retirement.

The stated purpose of these accounts is to deal with two issues, first insufficient saving by workers, and second, moral hazard in search. Under a strict unemployment account system, lower search effort does not increase expected total unemployment benefits.

Unemployment accounts may well be desirable for dealing with these two problems; our model does not allow us to analyze these issues. Our analysis however shows that they fail in another dimension, the internalization of layoff decisions by firms. In effect, from the point of view of firms, they are equivalent to payroll financing of unemployment benefits, and thus, like the current system based on payroll taxes, lead to excessive destruction.[21]

- *On the role of the state*

In the benchmark, there is no need for state intervention. Severance payments can achieve production efficiency and full insurance, and, in the absence of state provided insurance, will be voluntarily provided by firms.

The extensions show however the limits of the argument. The fact that, in general, contributions from firms in case of layoff need not be equal to unemployment benefits makes direct severance payments unfeasible, and require the use of a pooling agency, financed through both payroll and layoff contributions, and paying out unemployment benefits. The fact that firms have an incentive to be judgment-proof requires prudential regulation, to make sure that either firms, or the pooling agency, are solvent in case of layoffs. Under ex-post wage bargaining, firms would rather not offer unemployment insurance as it simply increases their costs. And unobserved heterogeneity leads the state to want a different structure of financing of unemployment insurance than firms would offer. The last two reasons imply the need for either a state-run unemployment agency, or else a private agency with mandated contribution rates and benefits.

Can there be some role left for firms, in the form of voluntary severance payments if they so want? The answer is, in general, yes, although the issue of excess provision of insurance in the case where there are limits to insurance, is an intriguing one. The logic of the argument leads to prohibiting severance payments, or putting a ceiling on such payments by firms. Such a ceiling can in principle be enforced by mandating courts of

[21] To see this, think of unemployment accounts as requiring workers to pay τ when (if) employed, and allowing them to receive μ when (if) unemployed. Assume that contributions from employed workers equal payments to unemployed workers, so $(1-G(\bar{y}))\mu = G(\bar{y})\tau$. This scheme is exactly equivalent to an unemployment insurance system financed by payroll taxes. The solution can be obtained by going back to the case of shallow pockets and assuming $\bar{f}=0$ in equation (13). Threshold productivity is thus given by $\bar{y} = b + \mu/(1-G(\bar{y})) > b$.

law not to enforce labor contract covenants relative to severance payments. In practice it may be difficult to rule out severance pay for several reasons. For example, severance payment could be disguised in a labor contract as delayed compensation. Also, a large firm may build a reputation vis-a-vis its workers to actually deliver on its severance payments promises even if not forced by courts to actually pay. Thus, we are agnostic about the state's ability to demand exclusivity in the provision of insurance to workers, even though such exclusivity may be desirable.

References

[1] Acemoglu, D., and R. Shimer, (1999), "Efficient Unemployment Insurance", *Journal of Political Economy*, 107-5, October, 893-928.

[2] Acemoglu, D., and R. Shimer, (2000), "Productivity Gains from Unemployment Insurance", *European Economic Review*, 44-7, June, 1195-1224.

[3] Akerlof, G., and H. Miyazaki (1980) "The Implicit Contract Theory of Unemployment Meets the Wage Bill Argument," *Review of Economic Studies*, 48: 321–338.

[4] Alvarez, F., and M. Veracierto (2000), "Labor-Market Policies in an Equilibrium Search Model," NBER Macroeconomics Annual, B. Bernanke and J. Rotemberg eds, 14, 265-304.

[5] Azariadis, C. (1975) "Implicit Contracts and Underemployment Equilibria," *Journal of Political Economy*, 83: 1183–1202.

[6] Baicker, K., C. Goldin, and L. Katz (1998) "A Distinctive System: Origins and Impact of U.S. Unemployment Compensation,“ in *The Defining Moment*, M. Bordo, C. Goldin, and E. White (eds), University of Chicago Press and NBER, 227-263.

[7] Baily, M.(1974) "Wages and Employment under Uncertain Demand,“ *Review of Economic Studies*, 41: 37–50.

[8] Blanchard, O., and J. Tirole (2003a), "Redesigning the Employment Protection System," *De Economist*, 152: 1-20.

[9] —– (2003b), "Licenciements et Institutions du Marché du Travail," Rapport pour le Conseil d'Analyse Economique. La Documentation Française, pp. 7–50.

[10] Cahuc, P. and G. Jolivet, (2003), "Do We Need More Stringent Employment Protection Legislation?", mimeo Paris I.

[11] Cahuc, P. and A. Zylberberg, (2004), "*Le chômage, fatalité ou nécessité?*", Flammarion Paris.

[12] Gertler, M. and S. Gilchrist, (1994), "Monetary Policy, Business Cycle, and the Behavior of Small Business Firms", *Quarterly Journal of Economics*, 109, 309-340.

[13] Hansen, G. and A. Imrohoroglu, (1992) "The Role of Unemployment Insurance in an Economy with Liquidity Constraints and Moral Hazard", *Journal of Political Economy*, 100-1, February, 118-142.

[14] Holmström, B. (1982) "Moral Hazard in Teams," *Bell Journal of Economics*, 13-2: 324-340.

[15] Hopenhayn, H., and J.P. Nicolini (1997) "Optimal Unemployment Insurance, " *Journal of Political Economy*, 105(2): 412-38.

[16] Mortensen, D., and C. Pissarides (2003) "Taxes, Subsidies and Equilibrium Labor Market Outcomes", *Designing Inclusion: Tools to Raise Low-End Pay and Employment in Private Enterprise, Edmund S. Phelps (ed)*, Cambridge: Cambridge University Press.

[17] Pauly, M. (1974) "Overinsurance and Public Provision of Insurance: The Roles of Moral Hazard and Adverse Selection", *Quarterly Journal of Economics*, 88(1): 44-62.

[18] Shavell, S., and L. Weiss (1979) "The Optimal Payment of Unemployment Insurance", *Journal of Political Economy*, 87(6): 1347-62.

[19] Tirole, J., (2006), "Layoff Taxes and Shallow Pocket Firms," mimeo, University of Toulouse.

[20] Werning, I. (2002) "Optimal Unemployment Insurance with Unobservable Savings ", *Mimeo, MIT.*

[21] Winkelmann, L and R. Winkelmann (1998) "Why are the unemployed so unhappy? Evidence from Panel Data", *Economica*, 65(257): 1-15.

Appendix 1: Policy menus under firm heterogeneity

Suppose the state offers an option (τ_L, f_L) targeted at weak firms and another (τ_H, f_H) at strong firms, rather than the single option (τ, f). With obvious notation change (thresholds, payroll and layoff taxes are now indexed by the type of firm), the optimization problem is identical to the problem above, except that the last constraint,

$$\bar{y} - [w + \tau - f] = 0,$$

is replaced by the constraints:

$$\bar{y}_L - [w + \tau_L - f_L] = 0, \quad \bar{y}_H - [w + \tau_H - f_H] = 0,$$

and the incentive compatibility constraint that the strong firms do not want to masquerade as weak ones is given by:

$$-G_H(\bar{y}_H) f_H + \int_{\bar{y}_H}^{1} (y - w - \tau_H)\, dG_H(y) \geq -G_H(\bar{y}_L) f_L + \int_{\bar{y}_L}^{1} (y - w - \tau_L)\, dG_H(y).$$

The solution can then be characterized as follows:

- The state still fully insures workers: $w = b + \mu$.
- Strong firms face a net contribution rate equal to one, and so choose the efficient threshold:
$$\bar{y}_H = b \text{ and } f_H - \tau_H = \mu.$$
- Weak firms face a net contribution rate below one, and so choose a threshold higher than the efficient level:
$$\bar{y}_L = b + \rho \, \frac{G_L(\bar{y}_L) - G_H(\bar{y}_L)}{(1-\rho) g_L(\bar{y}_L)} \text{ and } f_L - \tau_L \leq \mu.$$