

NBER WORKING PAPER SERIES

THE OPTIMAL DESIGN OF UNEMPLOYMENT
INSURANCE AND EMPLOYMENT PROTECTION.
A FIRST PASS

Olivier Blanchard
Jean Tirole

Working Paper 10443
<http://www.nber.org/papers/w10443>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
April 2004

We are grateful to Daron Acemoglu, Suman Basu, Javier Ortega, John Hassler, and to the participants to the third Toulouse Macroeconomics workshop for helpful comments. The views expressed herein are those of the author(s) and not necessarily those of the National Bureau of Economic Research.

©2004 by Olivier Blanchard and Jean Tirole. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

The Optimal Design of Unemployment Insurance and Employment Protection. A First Pass
Olivier Blanchard and Jean Tirole
NBER Working Paper No. 10443
April 2004
JEL No. D60, E62, H21, J30, J32, J38, J65

ABSTRACT

Much of the policy discussion of labor market institutions has been at the margin, with proposals to tighten unemployment benefits, reduce employment protection, and so on. There has been little discussion however of what the ultimate goal and architecture should be. The paper focuses on characterizing this ultimate goal, the optimal architecture of labor market institutions.

We start our analysis with a simple benchmark, with risk averse workers, risk neutral firms and random shocks to productivity. In this benchmark, we show that optimality requires both unemployment insurance and employment protection---in the form of layoff taxes; it also requires that layoff taxes be equal to unemployment benefits.

We then explore the implications of four broad categories of deviations: limits on insurance, limits on layoff taxes, ex-post wage bargaining, and heterogeneity of firms or workers. We show how the architecture must be modified in each case. The scope for insurance may be more limited than in the benchmark; so may the scope for employment protection. The general principle remains however, namely the need to look at unemployment insurance and employment protection together, rather than in isolation.

Olivier Blanchard
Department of Economics
MIT
Cambridge, MA 02139
and NBER
blanchar@mit.edu

Jean Tirole
IDEI
21 Allées de Brienne
31000 Toulouse
France
tirole@cict.fr

Introduction

Much of the policy discussion of labor market institutions has been at the margin, with proposals to tighten unemployment benefits, reduce employment protection, and so on. There has been little discussion however of what the ultimate goal and architecture should be. What should be the level of insurance, how should it be provided, how should it be financed? Should there be some form of employment protection? If so, should firms make payments to the state, or directly to the workers they lay off? What should be the size of these payments?

Our focus in this paper is on characterizing this ultimate goal, the optimal architecture of labor market institutions. We make no attempt to look at transition and political economy issues. These are obviously highly relevant to policy makers, but they should come second, once the ultimate goal is clear.

We start our analysis in Section 1 with a simple benchmark. Workers are risk averse; entrepreneurs run firms and are risk neutral. Productivity of any worker-firm match is random. In this benchmark, the proper institutions can achieve both production efficiency and full insurance. This can be achieved in a number of ways. For example, insurance can be provided through the payment of unemployment benefits by an unemployment agency. And efficiency can be achieved by using layoff taxes, leading firms to internalize the cost of unemployment.

In this benchmark, efficiency requires that layoff taxes be equal to unemployment benefits. This implies that layoff taxes fully finance benefits, so there is no need to use payroll taxes. The contribution rate, defined as the ratio of layoff taxes to unemployment benefits, is equal to one. Unemployment benefits and employment protection—defined as layoff taxes—are essential components of the optimal architecture: The presence of the first requires the presence of the other.

We then consider deviations from the benchmark. There is potentially a thousand ways of doing so, and there is no agreement among labor economists as to which imperfections are most relevant. Thus, our strategy is to explore the implications of four broad categories of deviations:

Limits on insurance. Such limits may come from the need to give incentives either to the unemployed to search for another job, or to the employed workers

not to shirk on the job. We show in Section 2 that, in this case, it is optimal to partly compensate for the limits on insurance by decreasing the incidence of unemployment, so by decreasing layoffs below the efficient level. This is achieved by choosing layoff taxes higher than unemployment benefits, thus by choosing a contribution rate higher than one. As layoff tax revenues now exceed unemployment benefits, the state then balances its budget through a negative payroll tax.

Under this deviation, optimality calls for higher employment protection: The tighter the limits on unemployment benefits, the higher the optimal employment protection, but at an increasing cost in terms of efficiency. Reforms of the unemployment insurance system which allow for higher insurance while maintaining search effort thus have both direct insurance effects, but also indirect ones, as they allow to decrease employment protection and improve efficiency.

Limits on layoff taxes. Shallow pockets may make it difficult or impossible for firms to pay layoff taxes. We show, in Section 3, that, even in this case, it remains feasible and optimal to fully insure workers against unemployment. But the limits on layoff taxes imply that unemployment benefits must now be financed partly through layoff taxes and partly through payroll taxes. This in turn implies that firms only partially internalize the cost of unemployment, and this leads to too many layoffs relative to the efficient level.

We also consider, in this section, the case of multi-activity firms, and the question of whether taxing layoffs may lead to more layoffs—creating “snowball effects”. We show that, under shallow pockets, this may indeed be the case, and derive the optimal layoff tax schedule for a firm composed of two jobs. Intuitively, the layoff tax on the first layoff should not be so high relative to the layoff tax on the second job as to lead the firm to want to close the second job.

Ex-post wage bargaining. Our benchmark assumes that wages are set ex-ante. Thus, better insurance reduces the wage that risk averse workers are willing to accept and reduces expected labor costs. In fact, wages are often partly renegotiated ex-post. In that case, rather than leading to lower wages, more generous social protection, be it through unemployment benefits or layoff taxes, strengthens the bargaining position of workers, and thus is likely to lead to an increase rather

than to a decrease in wages. Ex-post wage determination also implies that the wage is now likely to depend on productivity, introducing wage risk for employed workers. We take up the case of ex-post wage bargaining in Section 4, and show that, in this case, if workers have any bargaining power, the contribution rate should be less than one. The higher the bargaining power of workers, the lower the feasible level of insurance, and the lower the contribution rate.

Heterogeneity. Our benchmark treats all firms and all workers as ex-ante identical. This is clearly not the case in fact. We consider various dimensions of heterogeneity in Section 5.

We first consider heterogeneity of firms. We assume two types of firms, differing in the distribution of productivity. Weak firms are more risky: Their distribution function is stochastically dominated by that of strong firms. We show that if heterogeneity is not observable by the state, and the state relies on a uniform tax policy, the solution is still to provide full insurance to workers, but to finance benefits partly from layoff taxes and partly from payroll taxes—to have a contribution rate below one. This leads to a transfer of resources from strong to weak firms, and, because weak firms are the relevant firms at the creation margin, such a transfer is indeed desirable. If the state can instead offer tax menus and let firms self select, only the contribution rate for weak firms needs to be lower than one; the destruction margin of strong firms is left undistorted. If heterogeneity is observable and the state can condition taxes on observable characteristics, then even better can be done. Job subsidies can be used to transfer resources from strong to weak firms, while leaving destruction margins for both strong and weak firms unaffected. This reasoning suggests the desirability of treating different types of firms, for example new versus mature firms, differently.

We then consider heterogeneity of workers. In parallel to the treatment of firms, we assume two types of workers, “low-ability” and “high-ability” workers, differing in the distribution of their productivity. We show that if heterogeneity is not observable by the state and the state relies on a uniform tax policy, then, to the extent that the state cares about inequality, full insurance and a contribution rate below one are again optimal. Lower layoff tax rates lead to a transfer from high-ability to low-ability workers, a desirable outcome as it reduces income inequality. Again, if

feasible, offering menus or conditioning taxes on observable characteristics lead to a better outcome, allowing for transfers from high-ability to low-ability workers at a reduced efficiency cost. This reasoning again suggests the desirability of treating different workers, for example new entrants versus workers with higher seniority, differently.

Before we start, we want to emphasize the limits of our analysis. Our framework is a static, one-period, model. Our only excuse is that the framework gets complex enough already; but its limitations prevent us from taking up a series of questions, such as the design of experience rating systems as a way of implementing layoff taxes, or the potential role of individual mandatory unemployment accounts. Also, we are aware that our broad categories of deviations from the benchmark may not fit each specific imperfection. Nevertheless, we believe that the conclusions we draw from our model, about the articulation of unemployment insurance and employment protection, about the relation of the optimal contribution rate and the relative use of layoff and payroll taxes to various imperfections, are quite general and can help design better systems. In that context, we draw, in Section 6, what we see as our main conclusions.

1 A benchmark

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 $\text{cd } G(y)$, 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.

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

- *The state* observes only the employment status (whether the worker remains employed or is laid off).

Accordingly, the state has three tools at its disposal, a payroll tax rate τ , a layoff tax f , and unemployment benefits, μ . (Note the absence of job creation subsidies or taxes; these are straightforward combinations of layoff and payroll taxes, and are redundant here).

The state runs a balanced budget.

Decisions are taken in three stages:

- *Stage 1.* The state chooses τ , f and μ , subject to budget balance.
- *Stage 2.* Entrepreneurs decide whether to start firms and pay the fixed cost. They hire workers, and offer them contracts, characterized by w and y^* , a threshold productivity level below which the worker is laid off (the threshold will be assumed to be interior in what follows.)
The assumption of a constant wage w rather than a schedule $w(y)$ follows straightforwardly from workers' risk aversion.
Note the absence of severance payments, i.e. direct payments from the firm to the worker. As will be clear, they would be redundant here.
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 revealed.

If $y > y^*$, then the firm keeps the worker, produces y , pays w to the worker, and τ to the state.

If $y < y^*$, the firm lays the worker off; the firm pays f to the state, and the state pays μ to the unemployed worker.

Given these assumptions, *expected utility* is given by:

$$V_W \equiv G(y^*)U(b + \mu) + (1 - G(y^*))U(w) \quad (1)$$

The *free entry condition* (or participation constraint) for entrepreneurs is given by:

$$V_F \equiv -G(y^*)f + \int_{y^*}^1 y dG(y) - (1 - G(y^*))(w + \tau) \geq I \quad (2)$$

The *government budget constraint* is given by:

$$V_G \equiv -G(y^*)(\mu - f) + (1 - G(y^*))\tau \geq 0 \quad (3)$$

1.2 The optimization problem

The problem of the firm

Before turning to the problem of the state, we need to look at the problem of the firm.

At the time of hiring, a firm chooses the labor contract that maximizes the value of the firm, for a given level of expected utility for the worker:

$$\max V_F \quad \text{subject to } V_W = \bar{V}_W$$

From the first-order conditions, y^* is then given by:

$$y^* = (w + \tau - f) - \left(\frac{U(w) - U(b + \mu)}{U'(w)} \right) \quad (4)$$

and w is determined in turn by the condition that expected utility be equal to \bar{V}_W .

The first term in the expression for the threshold productivity is the net cost of producing. Were this the only term, the firm would simply lay the worker off when output was less than the net cost of producing. But there is in general a second term which says that, unless the worker is fully insured, the threshold productivity level will be lower than the net cost of producing. This insurance-like term may

be familiar from the old line of research on implicit contracts, and will reappear later on; we shall discuss it at more length then.

The threshold defined by equation (4) is not however in general time-consistent. Ex post, i.e. once the firm has hired a worker, given the wage w , the payroll and layoff tax rates τ and f , the threshold productivity below which the firm will want to lay a worker off is given by:

$$y^* = (w + \tau - f) \tag{5}$$

The expression no longer includes the insurance term.

In the benchmark, it will be the case that workers are fully insured so the two expressions for the threshold will coincide. Later on, this will no longer necessarily be the case, and so the question will arise of what assumption to make, the choice of the ex-ante or the ex-post threshold.

If y is observable and contractible, then the threshold will be given by the ex-ante value. If y is not contractible however, the firm may still be able to commit to the ex-ante value. What it needs to do is to put aside funds in amount $(U(w) - U(b + \mu))/U'(w)$, to be paid to a third party (not the worker) in case of layoff. This will clearly lead the firm again to choose the ex-ante value for the threshold.

Our maintained assumption will be that y^* is not contractible, that firms do not put such funds aside, so firms choose the ex-post value of the threshold—the value given by equation (5). We shall discuss the implications of the alternative assumption when it matters in Section 2.

The problem of the state.

The state chooses tax rates and unemployment benefits so as to maximize the expected value of workers' utility, equation (1), (expected profits are equal to zero in this economy, so we can ignore entrepreneurs' utility), subject to the free entry condition, equation (2), the government budget constraint, equation (3), and the threshold chosen by firms, equation (5).

Note that we can consolidate the free entry condition and the government budget

constraint to get:

$$V_{GF} \equiv -G(y^*)\mu + \int_{y^*}^1 y dG(y) - (1 - G(y^*))w \geq I \quad (6)$$

Neither the payroll tax rate nor the layoff tax rate appear in the equation. From the point of view of expected profit, given the balanced budget constraint, it does not matter whether unemployment benefits are financed by one or the other tax. Thus, the consolidated free entry condition depends on μ , y^* and w , not directly on f and τ .

This suggests the following formalization strategy: To set up the maximization problem as a *mechanism design problem*, finding the optimal allocation (w, μ, y^*) subject to condition (6). Then, to find out how the solution can be implemented by the appropriate choice of f and τ . This approach turns out to be useful, not so much in the benchmark but when we introduce deviations later.

The optimization problem

In the first step, we characterize the solution to the following problem.

$$\max_{w, \mu, y^*} V_W \quad \text{subject to } V_{FG} \geq I$$

or, equivalently

$$\max_{w, \mu, y^*} \{G(y^*)U(b + \mu) + (1 - G(y^*))U(w)\}$$

subject to:

$$-G(y^*)\mu + \int_{y^*}^1 y dG(y) - (1 - G(y^*))w \geq I$$

Note that μ, w, y^* fully characterize the real allocation. From the first-order conditions, it follows that:

$$w = b + \mu \quad (7)$$

$$y^* = b \quad (8)$$

and the levels of w and μ are determined by condition (6) holding with equality.

The first condition is a full insurance condition: Workers achieve the same level of utility, whether employed or laid-off and unemployed.

The second condition 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.

Thus, in the optimal allocation, insurance does not stand in the way of efficiency. We now turn to the characterization of taxes which support the optimum.

Implementation through payroll and layoff taxes

Payroll and layoff taxes are determined by the combination of the threshold condition and the government budget constraint.

Return to the *threshold condition*, equation (5). Replace both y^* and w by their expressions from above, to get:

$$f - \tau = \mu \tag{9}$$

The net tax associated with laying off rather than keeping a worker, which is equal to the layoff tax minus the payroll tax, must be made equal to the unemployment benefits paid by the state to the worker. This implies a *positive* relation between the two tax rates: The higher the payroll tax, the higher the layoff tax needed to induce the firm to take the efficient decision.

The second relation between the two tax rates follows from the *budget constraint*, equation (3):

$$G(y^*)f + (1 - G(y^*))\tau = G(y^*)\mu \tag{10}$$

Unemployment benefits must be financed either by layoff or by payroll taxes. This implies a *negative* relation between the two tax rates: The higher the payroll tax, the lower the required layoff tax. Combining equations (9) and (10) implies:

$$f = \mu, \quad \tau = 0$$

The layoff tax should be equal to unemployment benefits. The payroll tax should be equal to zero. Define the *contribution rate* as the ratio of layoff taxes to unemployment benefits: $r \equiv (f/\mu)$. An equivalent way of stating the previous result is that the contribution rate should be equal to one.

We summarize these results in Proposition 1.

Proposition 1. *In the benchmark, workers are fully insured ($b + \mu = w$). The threshold productivity is equal to the wage equivalent of being unemployed ($y^* = b$). The allocation is implemented through layoff taxes equal to unemployment benefits ($f = \mu$), or equivalently, through a unit contribution rate ($r = 1$). Payroll taxes are equal to zero ($\tau = 0$).*

1.3 Implications and discussion

Our benchmark yields a number of conclusions:

- Unemployment benefits and employment protection—defined as the payments by firms to the state at the time of layoff—are *both* central components of the optimal set of labor market institutions. They are complements in that higher unemployment benefits, coming from a higher need for unemployment insurance (a lower b), require higher payments from firms to the state.
- Indeed, in this benchmark, layoff taxes must be equal to unemployment benefits. The reason is simple: This makes firms fully internalize the cost of insurance provided by the state to the unemployed. The outcome is both full insurance of laid off workers, and production efficiency.

We have described the equilibrium as one where the state provides unemployment benefits to workers, and firms pay layoff taxes to the state. But an equivalent allocation can clearly be achieved without the state: All which is needed is for firms to make *severance payments* in amount μ to the workers directly. In this case, payments to workers and payments by firms are automatically equal; there is no need for third party intervention; and, indeed, firms are eager to provide such insurance to the risk averse workers as this allows them to reduce expected labor costs.

In fact, going beyond the confines of this benchmark, there are good reasons to conclude that unemployment insurance should be provided by a separate agency, financed by taxes paid by firms:

- As of the time of the layoff, the loss from unemployment is a random variable: The outcome of search is uncertain, and the worker does not know how long he is going to be unemployed. If the firm were to make a one-time severance payment to offset that loss, this one-time payment would do a poor job of insuring the laid-off worker. If the firm decided instead to pay the laid-off worker over time, contingent on his unemployment status, many other issues would arise: The difficulty for the firm to actually track the worker, and determine whether he is still unemployed or has found another job; the difficulty in monitoring his search effort and making sure that he is indeed looking for another job. Rather obviously, individual firms cannot monitor laid off workers well enough to provide them with adequate insurance. The role of monitoring unemployment status and search intensity must be therefore delegated to an agency, private or public. The state, given its existing administrative structure, is likely to be in the best position to do the monitoring, and to administer the payment of unemployment benefits, either alone or in conjunction with the private sector.
- There are good reasons to dissociate benefits (which provide insurance to workers) from layoff taxes (which provide the right incentives for firms). The two serve different purposes. While the two are equal in the benchmark, this will no longer be the case when we relax some of the benchmark assumptions. This dissociation could not be achieved if one limited oneself to the use of severance payments.

The benchmark model is useful in setting up the stage, and as a way of thinking about the issues. But it clearly ignores a number of important imperfections in the labor market. We now introduce those we see as most important, and derive their implications.

2 Limits on insurance

In our benchmark, workers were fully insured. In practice, incentive reasons command that workers not be perfectly insured. This may be because workers must have incentives to search while unemployed, and because they must have incentives not to shirk when employed.

2.1 A generic constraint

With this in mind, we introduce the following additional constraint in our optimization problem:

$$(1 - G(y^*))(U(w) - U(b + \mu)) \geq B \quad (11)$$

The expected utility gain from being employed relative to being unemployed must exceed some value B . Call this constraint the incentive compatibility condition for the worker, or ICW condition.

A simple shirking interpretation is as follows. Modify the benchmark to allow the worker to shirk or not shirk. Once hired, but before productivity is revealed, the worker has to decide whether to shirk or not shirk. Shirking brings private benefits B but results in zero productivity and thus a layoff. Shirking is unobservable. Thus, to prevent shirking, the condition above must hold.

Under the search effort interpretation, the constraint that the expected utility gain from being employed must be sufficient to justify search effort would be the natural constraint in a dynamic model. Even in our static model, a simple modification of the benchmark leads to a constraint close to equation (11). Suppose that, if laid off, the unemployed can, by exercising search effort, get another job in an alternative sector (home production, job abroad) at wage \bar{w} . Then, the ICW constraint is the same as equation (11), with \bar{w} instead of w . The basic implications, in particular with respect to the contribution rate, are the same as those derived below.

2.2 The optimization problem

The state solves the same problem as in the benchmark, subject to the additional ICW constraint. So:

$$\max_{w, \mu, y^*} \{G(y^*)U(b + \mu) + (1 - G(y^*))U(w)\}$$

subject to:

$$-G(y^*)\mu + \int_{y^*}^1 y dG(y) - (1 - G(y^*))w \geq I$$

$$\text{and } (1 - G(y^*))(U(w) - U(b + \mu)) \geq B$$

The solution to this problem can be characterized as follows.

Unemployment benefits, μ , can be no higher than allowed by the incentive constraint, thus providing incomplete insurance:

$$U(b + \mu) = U(w) - \frac{B}{1 - G(y^*)} \quad (12)$$

The threshold level of productivity, y^* , satisfies:

$$y^* = w - \mu - \frac{U(w) - U(b + \mu)}{U'(w)} \quad (13)$$

Or reorganizing:

$$y^* = b + \left[w - (b + \mu) - \frac{U(w) - U(b + \mu)}{U'(w)} \right] \quad (14)$$

For any concave utility, the term in brackets is negative; the stronger the curvature (the stronger the degree of absolute risk aversion), the more negative. Thus, if workers are risk averse, the optimal threshold is lower than the efficient level. Firms keep some workers although their productivity is lower than the wage equivalent of being unemployed. The lower layoff rate serves as a partial substitute for unemployment insurance.

The level of w itself is given by condition (6), holding with equality.

Implementation through payroll and layoff taxes

Replacing y^* from (13) in the expression for the threshold decision of firms, (5) gives us the first relation between f and τ :

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

or, equivalently:

$$f - \tau = \mu + \frac{B}{(1 - G(y^*)) U'(w)}$$

Together with the budget constraint, this condition yields:

$$f - \mu = \frac{B}{U'(w)} \quad (15)$$

If B is positive, the equilibrium has $f > \mu$ and $\tau < 0$. The higher B , the higher is the layoff tax rate, the more negative the payroll tax rate. In other words, the contribution rate, that is the ratio of the layoff tax paid by the firm to the unemployment benefit received by the worker, must now exceed one. The reason is the need to decrease layoffs, a need coming from the limits on the unemployment insurance that can be provided to the workers. We summarize these results in Proposition 2.

Proposition 2. *In the presence of limits on insurance, the threshold productivity is lower than the wage equivalent of being unemployed ($y^* < b$): The lower incidence of unemployment partly compensates for the limits on insurance when unemployed. The allocation is implemented through layoff taxes that exceed unemployment benefits ($f > \mu$), and negative payroll taxes to balance the budget. In other words, it is implemented through a contribution rate greater than one.*

2.3 Implications and discussion

- Despite our appeal to either shirking or search effort as motivations for equation (11), there is a sense in which the results carry more generality under the search interpretation. In the shirking interpretation, equation (11) was derived under the implicit assumption that, if the worker shirked, productivity was equal to zero. Under the more general assumption that shirking shifts the distribution from $G(\cdot)$ to $H(\cdot)$, where $G(\cdot)$ stochastically dominates $H(\cdot)$, we have built examples in which the optimal contribution rate is less than one. Under a search effort interpretation however, productivity in the firm is unaffected by search effort, and thus the problem does not arise.
- Our results in this extension bear a close relation to the results obtained in the “implicit contract literature” (in particular Baily (1974), Azariadis (1975), Akerlof and Miyazaki (1980)).

That literature looked at the optimal contract between risk neutral firms and risk averse workers. Under the assumption that there were neither severance payments nor unemployment benefits, one of the conclusions was that there would be overdevelopment, that firms would lay too few workers off relative to the efficient outcome. One of the criticisms addressed to those papers was the question of why firms did not offer unemployment insurance or severance payments. In the discussion here, the limits come from monitoring problems, and the solution takes the form of a layoff tax rate imposed by the state. But the logic is very much the same.

Indeed, the connection is even closer than it looks. Recall our discussion of the choice of the threshold by firms. The results above were derived under the assumption that firms choose the threshold according to equation (5), that they choose it ex post and do not feel bound by any earlier commitment to workers at the time of hiring. We can ask what happens if, in fact, firms are bound by that commitment, so that y^* is given instead by equation (4). In this case, it is easy to check that the layoff tax rate is given by: $f = \mu$ and $\tau = 0$. In other words, because firms voluntarily distort the destruction margin, the state does not need to introduce any other distortion, and thus can still set the contribution rate equal to one.

- With respect to variations in B , unemployment benefits and employment protection (defined as layoff taxes) are now substitutes. While our analysis is purely normative, the same trade-off is likely to emerge in a positive model of labor market institutions. To the extent that the unemployment insurance system provides limited insurance (for any reason), there will be a demand for high employment protection. An interesting characteristic of Continental Europe is indeed the presence of such an inverse relation between the generosity of the unemployment insurance system and the degree of employment protection (see for example Boeri [2002]).
- Governments in many countries are exploring ways of providing generous insurance while making sure that the unemployed exert search effort and accept employment (for example by requiring them to accept “reasonable job offers”). In the context of our model, such reforms can be thought of as

decreasing B . As such, they have both favorable direct and indirect effects on welfare. The direct one is better insurance. The indirect one is the decreased need to reduce layoffs below the efficient level, leading to better allocation and an efficiency gain.

3 Shallow pockets

Our benchmark model embodied the assumption that firms were risk neutral and had deep pockets. These assumptions are surely 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. We leave risk aversion aside, and focus on the implications of limited funds.

3.1 A generic constraint

The simplest assumption to make here is 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.

In this context, we need to extend the set of tools available to the state. Until now, job subsidies or taxes were redundant. It is no longer obvious that this is the case. So we give the state four instruments, benefits μ , payroll taxes τ , layoff taxes f , and job creation taxes t (job subsidies if t is negative).

The new constraint on the optimization problem is therefore:

$$f \leq \bar{f} - t \tag{16}$$

Layoff taxes must be less than or equal to free cash flow minus job taxes (equivalently: plus job subsidies). We call it the “shallow pockets” constraint.

This specification raises the issue of the potential role of outside, “deep pockets” investors in alleviating the constraint. We shall return to this point below.

3.2 The optimization problem

With the introduction of job creation taxes, the government budget constraint becomes:

$$G(y^*)(f - \mu) + [1 - G(y^*)]\tau + t \geq 0.$$

Combining this equation with the shallow pocket constraint (16), and the (unchanged) threshold condition (5) gives an equation which no longer depends on τ , f , and t :

$$G(y^*)\mu - [1 - G(y^*)](y^* - w) \leq \bar{f} \quad (17)$$

The *optimization problem* of the state is given therefore by:

$$\max_{w, \mu, y^*} \{G(y^*)U(b + \mu) + [1 - G(y^*)]U(w)\}$$

subject to:

$$\begin{aligned} G(y^*)\mu - [1 - G(y^*)](y^* - w) &\leq \bar{f} \\ \text{and } -G(y^*)\mu + \int_{y^*}^1 y dG(y) - [1 - G(y^*)]w &\geq I \end{aligned}$$

where the second constraint is the consolidated budget constraint, which, despite the introduction of t , turns out to be the same as before, i.e. equation (6).

From the first-order conditions, the solution can be characterized as follows:

Workers are still fully insured:

$$w = b + \mu \quad (18)$$

and the threshold level is given by:

$$y^* = \frac{b}{1 - \left(\frac{\nu}{\nu + \theta}\right) \left(\frac{1}{\eta}\right)} \quad (19)$$

where $\eta \equiv g(y^*)y^*/(1 - G(y^*))$ denotes the elasticity of employment with respect to the threshold (or, equivalently, with respect to the payroll tax, or with respect

to (minus) the layoff tax), and ν and θ are the non-negative shadow prices associated with the first and the second constraints respectively. Provided that the first constraint is binding, $\nu > 0$, then $y^* > b$: The optimal threshold is higher than the efficient level.

Implementation through taxes is straightforward: The government budget constraint and the threshold equation imply that f is smaller than μ . The job creation tax t is irrelevant, so if we set it equal to zero, $f = \bar{f}$. The contribution rate is less than one, and the difference between unemployment benefits and layoff taxes is made up by a positive payroll tax rate.

To understand these results, note first that the shallow pocket constraint does not affect the ability of the state to fully insure workers. To see why, suppose for example that $w > b + \mu$. Consider a decrease in w and an increase in τ by the same unit amount. This affects neither the threshold condition nor the firm's profit. Then, from the government budget constraint, the unemployment benefit can be increased by $[(1 - G(y^*)) / G(y^*)] \Delta\tau$. Together, these changes imply a change in utility of $[-(1 - G(y^*))U'(w) + (1 - G(y^*))U'(b + \mu)] \Delta\tau > 0$. Thus, the state will decrease w and increase μ until workers are fully insured.

Note also that the job creation tax or subsidy still plays no useful role here. To see why, consider the introduction of job subsidies in amount $\Delta t < 0$. From equation (16), this allows the state to increase layoff taxes by $-\Delta t$. But the increase in layoff taxes does not compensate for the increase in job subsidies, and to satisfy the government budget constraint, the state must increase payroll taxes by $-\Delta t$. The effect of similar increases in layoff and payroll taxes leaves the threshold chosen by firms unchanged, and thus has no effect on the allocation. In other words, the state has no way to move cash around so as to soften the shallow pocket constraint.

Given full insurance, and the limit on layoff taxes, the state must therefore finance benefits partly from layoff taxes, and partly from payroll taxes. This is why, if the shallow pocket constraint is binding, the contribution rate is less than one.

We summarize our results in the following proposition:

Proposition 3 *In the presence of shallow pockets, workers are still fully insured*

($w = b + \mu$). Layoff taxes are however less than unemployment benefits, leading to a contribution rate less than one, and a threshold productivity level higher than the wage equivalent of being unemployed ($y^* > b$).

3.3 Multi-activity firms

A concern often expressed about layoff taxes (more generally, about payments imposed on firms in bad times) is that, even if the firm can pay the taxes, it can only do so at the expense of other activities in the firm. For example, in order to pay the layoff tax associated with the closing down of activity 2, a firm may have to close otherwise healthy activity 1.

In the model we have just presented, firms are one-activity (one-job) firms. We extend it as follows. We assume firms to be composed of two activities (two jobs), $i = 1, 2$. The joint distribution of productivity in the two jobs is as follows:

- With probability α , y_1 is drawn from distribution $G(\cdot)$, and $y_2 = y_1$
- With probability $1 - \alpha$, y_1 is drawn from distribution $G(\cdot)$, and $y_2 = 0$.

For convenience, we shall refer to firms for which $y_2 = y_1$ as “strong firms” (although, of course, their productivity may be low), and to firms for which $y_2 = 0$ as “frail firms.”

All firms have initial assets $2(I + \bar{f})$, sink investment $2I$, and hire two workers.

On the government side, the state levies payroll tax τ per job, layoff taxes f_1 in case of a single layoff, $2f_2$ in case of two layoffs, and pays unemployment benefits μ . The assumption that unemployment benefits and payroll taxes are not conditional on the number of jobs preserved is not restrictive. By the same argument as before, the state is in a position to fully insure workers, and so it does. This implies that unemployment benefits are the same for all unemployed, irrespective of the state of the firm that laid them off, and given by $\mu = w - b$. Similarly, it can be shown that there is no gain in allowing for different payroll taxes. Finally, and by the same argument as before, it can be shown that the job creation tax or subsidy plays no useful role here, and so, without loss of generality, we set it equal to zero.

The problem of the firms

Ex-post strong firms face the choice of laying off zero or two workers. If they keep the workers, their profit is given by $y_1 + y_2 - 2(w + \tau) = 2(y_1 - w - \tau)$. If they lay both workers off, their profit is given by $-2f_2$. Thus, their threshold productivity is given by:

$$y^* = (w + \tau) - f_2 \quad (20)$$

Ex-post frail firms face the choice of laying off one or two workers. If they lay one worker off, their profit is given by $y_1 - (w + \tau) - f_1$. If they lay both workers off, their profit is given by $-2f_2$. Thus, their threshold productivity is given by:

$$y^{**} = (w + \tau) - (2f_2 - f_1) \quad (21)$$

This expression embodies the concern described earlier: *An increase in the layoff tax f_1 encourages layoffs by frail firms*. It creates a “spillage” or “snowball effect” on the otherwise healthy activity.

A priori, there appears to be now two constraints on layoff taxes. First, firms—strong or frail—that lay two workers off, have to be able to pay layoff taxes:

$$2f_2 \leq 2\bar{f}. \quad (22)$$

Second, those frail firms that lay only one worker off must also be able to pay the layoff tax. Their profit on the remaining activity is $[y_1 - (w + \tau)]$. As the state does not observe y_1 , it cannot levy more than $2\bar{f} + [y^{**} - (w + \tau)]$, so $f_1 \leq 2\bar{f} + [y^{**} - (w + \tau)]$. Using condition (21), this condition reduces to $2f_2 \leq 2\bar{f}$, so we only need to impose condition (22).

The optimization problem

Using the result (which still holds here) that workers will be fully insured, the optimization problem can then be written as:

$$\max_{\mu, y^*, y^{**}, w} U(b + \mu)$$

subject to two conditions:

The first is obtained by combining the government budget constraint, the full insurance condition ($w = b + \mu$), conditions (20) and (21) for the two thresholds, and the shallow pocket condition (22):

$$2\mu - 2\alpha(1 - G(y^*))(y^* - b) - (1 - \alpha)(1 - G(y^{**}))(y^{**} - b) \leq 2\bar{f} \quad (23)$$

The second is the consolidated budget constraint which, using again the full insurance condition and the two threshold conditions, can be written as:

$$2\alpha \int_{y^*}^1 (y - b)dG(y) + (1 - \alpha) \int_{y^{**}}^1 (y - b)dG(y) \geq 2I + 2\mu \quad (24)$$

From the first-order conditions, the two thresholds are given by:

$$y^* = y^{**} = \frac{b}{1 - \left(\frac{\nu}{\nu + \theta}\right) \left(\frac{1}{\eta}\right)}$$

where, as earlier, η is the elasticity of employment with respect to the threshold and ν and θ are the two non-negative shadow prices associated with the first and the second constraints respectively.

Implementation through taxes is straightforward. The equality of the two thresholds, together with the two threshold conditions, implies that $f_1 = f_2$, and if the firm has shallow pockets so $\nu > 0$, $f_1 = f_2 = \bar{f}$. Whatever difference exists between unemployment benefits and layoff taxes is made up through positive payroll taxes; the contribution rate is less than one. We can summarize our results as follows:

Proposition 4 *Limited assets in a multi-activity firm raise the possibility of “snowball effects”. In that case, the state in general does not want to collect the full collectable amount in case of partial layoffs. In the example we presented, where a two-job firm has assets $2\bar{f}$ after investment, the state levies tax \bar{f} for one layoff, and $2\bar{f}$ for two layoffs (if \bar{f} is not too large).*

The result however that the layoff tax schedule is linear is too strong, and due to the stochastic structure of our example. In our example, f_1 affects only the decision of “frail” firms to lay one or two workers off. By contrast we know that, if the relevant choice were between zero and one layoff as in the previous section, then maximum layoff taxes would be optimal ($f_1 = 2\bar{f}$ whenever $2\bar{f} \leq \mu$).

Unsurprisingly, if we add firms whose relevant trade-off is between zero and one layoff to our example, we obtain a *concave* tax structure: $f_1 \geq f_2$. To see this, suppose that a fraction β of the firms have y_1 and y_2 drawn as above (y_1 drawn from $G(\cdot)$, and $y_2 = y_1$ with probability α and $y_2 = 0$ with probability $1 - \alpha$), and that for the complementary fraction $1 - \beta$ of the firms, $y_1 = 1$ and y_2 is drawn from distribution $H(y)$. (All firms are ex-ante identical.) Then we know that if $\beta = 1$, $f_1 = f_2 = \bar{f}$, and for $\beta = 0$, $f_1 (= 2f_2) = 2\bar{f}$. In general, $f_1 \geq f_2$.

3.4 Implications and discussion

Our benchmark looked at the case in which firms had enough assets to pay the layoff tax and concluded that $f = \mu$. By contrast, the shallow pocket firm we looked at in this section had limited assets and could not pay more than \bar{f} , leading to a lower contribution rate. This raises the question of how \bar{f} is itself determined. We have explored this question in two ways.

The first is to assume, in contrast to the maintained assumption in this paper, that 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 optimal threshold is then given by: $y^* = b/(1 - (1/\eta))$. (This condition directly follows from equation (19): Under ex-post policy setting, the entry condition becomes irrelevant to the decision problem of the state, so $\theta = 0$, and this equation obtains.)

The second is to assume that the firm has limited assets but has access to outside finance. The answer, not surprisingly, depends then on the relation between the firm and the investors. Suppose first that there are deep-pocket, risk-neutral, investors, but that these investors are uninformed: Like the state, they cannot

observe productivity directly, just whether workers are employed or unemployed. Then, it is clear that allowing for the presence of such investors makes no difference to the results we just derived. Whatever the investors do, the state could already do by itself. For example, in the one-activity case, let the funds advanced by the investors to the firm be equal to F , and the payments of the firm to the investors be D_0 if production takes place, and D_1 if it does not. Then, the investors' break even condition is given by:

$$G(y^*)D_1 + (1 - G(y^*))D_0 \geq F$$

But this is exactly the same condition as the condition faced by the state in its choice of t, f, τ (for a given value of μ) earlier. And, for the same reason the state could not use t to relax the shallow pocket constraint, nor can the presence of investors relax this constraint.

Things are clearly different if investors are informed, i.e. can assess the realization of productivity y . In this case, they can ask more of the firm in good states, and give more to the firm in bad states. Indeed, in the absence of other constraints, it is clear that the firm and investors together are just like the deep-pocket firm of the benchmark. In that case f can be set equal to μ , independent of the free cash flow of the entrepreneur.

Things become more interesting if we assume that investors are indeed informed and so can ask for payments from the firm contingent on y , but that the entrepreneur enjoys a rent $R > 0$ in case of continuation (so if y is total value, investors can get at most $y - R$). This assumption implies the standard corporate finance feature that the entrepreneur is more eager than investors to continue since he receives a private benefit even when the firm no longer makes profits. Setting up the model and characterizing the solution would take us too far afield, and we refer the reader to Blanchard and Tirole (2004). In that extension, we find that the threshold level of productivity is higher than the efficient level ($y^* > b$), leading to more layoffs than in the benchmark. But it is still optimal for the state to set a unit contribution rate, and workers are fully insured $f = \mu = w - b$. Thus, in this case, shallow pockets lead to inefficiently high layoffs, but do not justify a lower contribution rate. The reason for these results is that low entrepreneur's

solvency is the culprit and the state, constrained to balance the budget, can do nothing to boost this solvency.

4 Ex-post wage negotiation

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. A firm offered the option of joining the layoff tax cum unemployment benefits system, would have voluntarily joined.

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 a vis that worker; a worker who will receive unemployment benefits if laid off is in a stronger position. The layoff tax and the provision of unemployment benefits both lead to higher, not lower, wages, and thus to increase labor costs.

In this section, we therefore modify our earlier assumption about wage setting, assume ex-post wage bargaining instead, and characterize the shape of optimal labor market institutions under ex-post wage bargaining.

4.1 A formalization of wage bargaining

The following formalization captures the effects of benefits and taxes on ex-post wage determination in a simple way:

Assume wage setting now takes place after productivity is realized, and is the outcome of 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 $1 - \beta$.

Under the assumption that the firm chooses the threshold level of productivity so as to maximize profit ex-post, in stage 2, the highest wage the firm will accept, and therefore the wage offered by the worker, is equal to $y + f - \tau$. The highest

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 $\beta(y + f - \tau) + (1 - \beta)(b + \mu)$. This implies that, in stage 1, the worker will make the highest offer acceptable by the (risk neutral) firm, i.e. an offer of

$$w(y) = \beta(y + f - \tau) + (1 - \beta)(b + \mu) \quad (25)$$

The higher the layoff tax rate, or the higher unemployment benefits, the higher the wage.

The threshold value for productivity, y^* , is given by:

$$y^* = w(y^*) + \tau - f$$

Equivalently, using equation (25) and rearranging:

$$y^* = b + \mu + \tau - f \quad (26)$$

Just as in the benchmark, the combination of $f = \mu$, $\tau = 0$ would deliver the efficient threshold, $y^* = b$. But, as we shall see, other considerations are now relevant.

Expression (26) allows us to rewrite the wage schedule as:

$$w(y) = (b + \mu) + \beta(y - y^*) \quad (27)$$

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. The wage then increases with β times the difference between productivity and threshold productivity.

4.2 The optimization problem

The state 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 (27):

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

subject to:

$$\begin{aligned} -G(y^*)\mu + \int_{y^*}^1 (y - w(y)) dG(y) &\geq I \\ \text{and } w(y) &= b + \mu + \beta(y - y^*) \end{aligned}$$

The solution can be characterized as follows:

The threshold level of productivity is implicitly defined by:

$$y^* = b + \frac{\beta(1 - G(y^*))}{g(y^*)} \left[1 - \frac{E[U'|y \geq y^*]}{E U'} \right] \quad (28)$$

where $E[U'|y \geq y^*] \equiv (\int_{y^*}^1 U'(w(y))dG(y))/(1 - G(y^*))$ is the expected value of marginal utility if employed, and $E U' \equiv G(y^*)U'(b + \mu) + \int_{y^*}^1 U'(w(y)) dG(y)$ 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 y^* is greater than b . The layoff rate exceeds the production efficient level. The more risk averse the workers, or the stronger the workers in bargaining, the higher the threshold level, the higher the layoff rate.

Given y^* , the feasible level of unemployment benefits, μ , is then determined by the first constraint above, the condition that $V_{FG} = I$.

One way of getting more intuition for the optimal threshold is to combine the first and the second constraints, and reorganize to read:

$$G(y^*)(b + \mu) + \int_{y^*}^1 ((b + \mu) + \beta(y - y^*)) dG(y) \leq G(y^*)b + \int_{y^*}^1 y dG(y) - I$$

The left side gives the value of total workers' income (the sum of wages, reservation

wages, and benefits) in the economy. The right side gives the value of total income in the economy. The two are the same, as free entry implies zero profit income.

Suppose for the moment that changes in y^* do not affect total income, and so do not affect total workers' income. Then, changes in y^* change the shape of the income profile without affecting the total workers' income. An increase in y^* implies that more workers receive $(\mu + b)$. The overall workers' income remains the same, but the income schedule is flatter. Indeed, the best value of y^* from the point of insurance is $y^* = 1$, where all workers receive the same income.

The assumption that changes in y^* do not affect total income is however clearly incorrect. Starting from $y^* = b$, small changes in y^* do not affect total income; this implies that, as it decreases risk, at least a small increase in y^* is desirable; this explains why the optimal y^* exceeds b . But, as y^* increases, the efficiency loss increases, and total income decreases. For $y^* = 1$ for example, then total income is just b and so μ must be equal to zero. This implies that the optimal value of y^* is greater than b , but less than one. It is increasing in the degree of concavity of the utility function. It is increasing in β , the share of the surplus going to workers: The higher β , the steeper the wage schedule, the larger the demand for insurance.

4.3 Implementation through payroll and layoff taxes

Replacing y^* from (28) in the expression for the threshold decision of firms, (26), gives us the first relation between f and τ :

$$f - \tau = \mu - \beta \frac{(1 - G(y^*))}{g(y^*)} \left[1 - \frac{E[U'|y \geq y^*]}{E U'} \right] \quad (29)$$

The other relation is given, as before, by the budget constraint, equation (3). As the second term on the right side of equation (29) is now negative, $f - \tau < \mu$. Together with the government budget constraint, this implies $f < \mu$ and so a contribution rate below one. The reason is clear from above: The optimal threshold is higher than the efficient level. This is achieved by lowering the contribution rate from its benchmark value, namely unity. By implication, payroll taxes must be positive, in order to finance the shortfall of the unemployment insurance system.

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.

We summarize our results as follows:

Proposition 5. *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 policy choices. The remaining choice for the state is that of y^* . It is then optimal to choose a threshold higher than the efficient level: ($y^* > b$). This choice decreases the uncertainty faced by workers at some cost in efficiency. This choice is in turn implemented by a contribution rate lower than unity.*

4.4 Implications and discussion

Under pure ex-post wage bargaining, the scope for unemployment benefits to provide insurance to workers is extremely limited: An increase in unemployment benefits does not change the slope of the wage schedule, increasing all wages by an amount equal to unemployment benefits. Put differently, the only degree of freedom is the location of the kink y^* in the worker's income schedule.

Indeed, in our framework, given the threshold level of productivity, the level of unemployment benefits is entirely determined by the wage schedule and the free entry condition—which implies that workers' income, the sum of wages, reservation wages, and unemployment benefits, must be equal to total output.

In that dimension however, our framework is too extreme. The static, one-period, nature of the model implies a fixed unemployment duration. In dynamic models,

such as Mortensen and Pissarides (1994) for example, the state is free to choose the level of unemployment benefits; but the higher the level of unemployment benefits, the longer the equilibrium unemployment duration. And an irrelevance result with the same flavor obtains. In the Pissarides (2000) model for example, an increase in unemployment benefits leads to an increase in expected duration exactly such that the “pain of unemployment”, i.e. the difference in the expected value of being employed or unemployed, remains constant.

5 Heterogeneity

We have assumed so far that all workers and all firms are 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: New entrants have distributions characterized by higher uncertainty than workers with more labor market seniority, and so on.

We study in this section the implications of heterogeneity, both on the firm and on the worker side, both observed and unobserved. We focus on differences in distributions, but discuss some other dimensions of heterogeneity later.

5.1 Heterogeneity of firms

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 $G_L(\cdot)$. 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 .

We start with the assumption that this heterogeneity is unobserved, so the state is unable to tell the two types of firms apart. We also start with the assumption that the state sets a uniform policy (τ, f, μ) ; we shall look later at menus in which strong and weak firms may self select.

Note that given the distribution assumptions and uniform taxation, strong firms have higher expected profits than small firms. Thus, at the creation margin, the free entry condition is relevant for weak firms only. By contrast, both types may lay workers off, so the destruction margin is relevant for both types of firms.

This in turn implies that we can no longer merge the free-entry condition and the government budget balance constraint: The former applies to a subset of firms (the weak firms) and the latter to the whole set of firms. We therefore have to take the more “pedestrian” route of optimizing with respect to all the variables.

We 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. The state’s optimization problem is given by:

$$\begin{aligned} \max_{\tau, f, \mu, w, y^*} \quad & [\rho G_H(y^*) + (1 - \rho)G_L(y^*)] U(b + \mu) \\ & + [\rho(1 - G_H(y^*)) + (1 - \rho)(1 - G_L(y^*))] U(w) \end{aligned}$$

subject to the free entry condition for weak firms

$$-G_L(y^*)f + \int_{y^*}^1 (y - w - \tau)dG_L(y) \geq I$$

the government budget constraint

$$\begin{aligned} & -[\rho G_H(y^*) + (1 - \rho)G_L(y^*)] (\mu - f) \\ & + [\rho(1 - G_H(y^*)) + (1 - \rho)(1 - G_L(y^*))] \tau \geq 0 \end{aligned}$$

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

$$y^* = w + \tau - f$$

The solution can then be characterized as follows:

- The state fully insures workers: $w = b + \mu$.
- The threshold level of productivity is given by:

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

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

This solution is in turn implemented by choosing layoff taxes lower than unemployment benefits, so through a contribution rate lower than unity. The difference is financed through payroll taxes.

The intuition for why the contribution rate is less than one is as follows: Both types of firms may lay workers off and so the destruction margin applies to all firms. By contrast, the creation margin corresponds to the weak firms only. The state can then improve workers' welfare by allowing for some cross-subsidy from strong to weak firms. Because the state is unable to assess the firms' strength, the cross-subsidy operates through the contribution rate: Weak firms lay workers off more than strong firms and so benefit more from a contribution rate smaller than unity.

Can the state do better by offering *menus* and letting firms self select? 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,

$$y^* - [w + \tau - f] = 0$$

is replaced by the constraints:

$$y_L^* - [w + \tau_L - f_L] = 0$$

$$y_H^* - [w + \tau_H - f_H] = 0$$

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

$$\begin{aligned}
& -G_H(y_H^*) f_H + \int_{y_H^*}^1 (y - w - \tau_H) dG_H(y) \\
& \geq -G_H(y_L^*) f_L + \int_{y_L^*}^1 (y - w - \tau_L) dG_H(y)
\end{aligned}$$

It is useful at this point to define the *net contribution rate* as the ratio of the layoff tax minus the payroll tax to unemployment benefits, $(f_i - \tau_i)/\mu_i$, $i = H, L$. Note that, so long as the net contribution rate is equal to one, the layoff decision is efficient: What matters here is not the layoff tax itself, but the difference between the layoff tax and the payroll tax.

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:

$$y_H^* = b \quad \text{and} \quad f_H - \tau_H = \mu$$

- Weak firms face a net contribution rate below one, and so choose a threshold higher than the efficient level:

$$y_L^* = b + \rho \frac{G_L(y_L^*) - G_H(y_L^*)}{(1 - \rho)g_L(y_L^*)} \quad \text{and} \quad f_L - \tau_L \leq \mu$$

In short, 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, with a contribution rate for weak firms below one.

If heterogeneity is observable, i.e. if the state is able to tell weak and strong firms apart, the state can do even better. Characterizing the optimal policy is straightforward: It is still optimal to subsidize weak firms. This can now be done however by offering a job creation subsidy to the weak firms, while setting the net contribution rate equal to 1 for both types of firms. This achieves the desired redistribution from strong to weak firms, while avoiding distortions at the destruction margin (Note that both layoff and payroll taxes must be higher than in the benchmark, as the extra revenue is needed to finance job creation subsidies to weak firms. The

difference between the layoff tax and the payroll tax remains however equal to unemployment benefits, so there is no distortion at the destruction margin).

We have focused above on heterogeneity in productivity distributions. While we shall not present results here, we have also explored heterogeneity in assets, including the model of corporate finance described informally in Section 3 (with informed investors, and some non pledgeable income.) There, the same type of results emerges, for the same reason. The weaker firms, i.e. those firms with the least assets, have smaller profits, and thus are the relevant firms at the creation margin. As, in order to satisfy investors, weaker firms also lay workers off more than other firms, they benefit more from a decrease in the layoff tax than stronger firms. Thus, if heterogeneity is unobservable and the state relies on uniform taxation, it is optimal for the state to decrease the contribution rate below one, transferring profits from stronger to weaker firms, and favoring job creation.

We can summarize our results as follows:

Proposition 6. *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 this heterogeneity is unobserved, and if the state relies on a uniform policy, then the optimal policy is to fully insure workers ($w = b + \mu$), choose a threshold level higher than the efficient level ($y^ > b$), and implement it through a contribution rate less than one.*

If the state offers a menu instead, then the optimal menu separates the firms and has the following properties: Workers are fully insured. Strong firms face a net contribution rate of unity, and choose the efficient threshold. Weak firms face a net contribution rate below one, and the threshold exceeds the efficient level. The underlying mechanism remains the same, cross-subsidization of weak firms through the use of a lower contribution rate.

If heterogeneity is observed by the state, then the optimal policy is to offer a job creation subsidy to weak firms, choose the efficient threshold for productivity, and implement it through a unit net contribution rate for both types of firms.

5.2 Heterogeneity of workers

A similar set of issues arises on the workers' side. Not all workers are alike. In particular, some are more uncertain and thus more likely to be laid off.

To think about the issue, 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 $G_H(\cdot)$, low-ability workers a distribution given by $G_L(\cdot)$, with $G_L(y) > G_H(y)$ for $y \in (0, 1)$. The fraction of workers with high ability is equal to ρ .

We assume that firms know the workers' abilities. Proceeding in parallel with the earlier case of heterogeneity in firms, we start by assuming both that the state does not know the workers' abilities, and that it 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:

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

Payment of Δ can be achieved through higher wages when employed and severance payments from the firm to the high-ability workers when laid-off. Alternatively, it 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, and assuming that the state maximizes

a utilitarian social welfare function, the optimal policy is the solution to:

$$\max_{\tau, f, \mu, y^*} \{(1 - \rho) U(b + \mu) + \rho U(b + \mu + \Delta)\}$$

subject to the free entry condition

$$-G_L(y^*) f + \int_{y^*}^1 [y - (b + \mu) - \tau] dG_L(y) \geq I$$

and the government budget constraint

$$[(1 - \rho) G_L(y^*) + \rho G_H(y^*)] (f - \mu) + [(1 - \rho) [1 - G_L(y^*)] + \rho [1 - G_H(y^*)]] \tau \geq 0$$

where Δ and y^* have been defined above.

The solution has the following form:

- If the proportion of low-ability workers is high enough, then the threshold for productivity is given by:

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

Note that $G_L(y^*) - G_H(y^*)$ 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 in turn implemented by using a contribution rate below unity, with the rest of unemployment benefits being financed by payroll taxes.

- If however, the fraction of low-ability workers is small enough, then it is optimal to not facilitate their employability. No job is created for them and they receive unemployment benefit μ for sure. In this case, threshold productivity is given by:

$$y^* = b$$

This solution is implemented by relying on a gross contribution rate greater

than one: $(f/\mu) > 1$; and a net contribution rate equal to one: $f - \tau = \mu$.
The payroll tax rate is in turn given by $\tau = [(1 - \rho)/\rho] \mu$.

The intuition for these results is as follows. When both types of workers are employed, the use of a contribution rate below one leads to a transfer from high-ability to low-ability workers, and therefore reduces inequality, but at some cost in efficiency. If the proportion of low-ability workers is very low, it is more efficient to have them not work, and then to choose tax rates so as to have efficient separations for high-ability workers. Because low-ability workers receive unemployment benefits but no corresponding layoff taxes are collected, the solution requires having both a positive payroll tax and a correspondingly higher layoff tax on high-ability workers. This increases revenues while maintaining a net contribution rate equal to one (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.)

Note that the logic of the results under firm and worker heterogeneity is 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 menus, one aimed at firms that announce they have hired a high-ability workers, one aimed at firms that announce they have hired a low-ability worker. The first menu has a net contribution rate equal to one. The second menu has a net contribution rate below one. Both types of workers are fully insured. Thus, just as the uniform case, heterogeneity leads to lower layoff taxes, but in this case only for low-ability workers.

What if differences between workers are observable? When the state can tell apart high-ability from low-ability workers—for example if workers are in bad health or disabled—then the optimal policy is to offer a job creation subsidy (or, equivalently, uniformly lower payroll and layoff taxes) to firms if they hire a low-ability worker. The net contribution rate can then be set equal to one for both types, and the destruction margin is undistorted for both types. In general, when differences are only partly observable, the solution combines job creation subsidies and a net

contribution rate below one.

We summarize our results in the following proposition:

Proposition 7. *Suppose that there are two types of workers: “high-ability” or “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 this heterogeneity is unobserved and the state relies on a uniform policy, then the optimal policy is to fully insure workers ($w = b + \mu$), choose a threshold level higher than the efficient level ($y^ > b$), and implement it through a contribution rate less than one. The lower layoff tax leads in effect to a transfer from high-ability workers to low-ability workers. If the proportion of low-ability workers is small enough, then it is optimal to leave them unemployed, and use a unit contribution rate for the remaining, high-ability workers.*

If the state offers a menu instead, then the optimal menu separates workers and has the following properties: Workers are fully insured. Firms hiring high-ability workers face a net contribution rate of unity, and choose the efficient threshold. Firms hiring low-ability workers face a net contribution rate below one, and the threshold exceeds the efficient level. The underlying mechanism remains the same, cross-subsidization of low-ability workers through the use of a lower contribution rate.

If heterogeneity is observed by the state, then the optimal policy is to offer a job creation subsidy to low-ability workers, and choose an efficient threshold for productivity, and implement it through a unit contribution rate.

6 Conclusion

We see our analysis as suggesting a few general principles:

The use of layoff taxes is the natural counterpart to the state provision of unemployment benefits. The second requires the first. Severance payments, direct transfers from firms to workers, can be seen as a crude way of implementing the same principle. In that sense, unemployment insurance and employment protection are basic components of the optimal set of labor market institutions. In the

benchmark, they allow the state to achieve both full insurance and an efficient allocation.

The principle is important. But it must be adjusted to take into account the imperfections present in the labor market.

Some imperfections lead to higher layoff taxes than in the benchmark. This is the case for constraints on the amount of insurance which can be provided without affecting search or work effort. The lower the limit on the provision of insurance, the higher the optimal amount of employment protection, thus the higher the optimal layoff tax. This however comes at some cost in efficiency.

Most others lead to lower layoff taxes than in the benchmark:

Constraints on the assets available to firms in case of layoffs obviously put a limit on layoff taxes. In this case, it is still feasible and optimal to provide full insurance, but it must be financed through both layoff and payroll taxes, again at some efficiency cost. The asset constraint also affects the layoff tax schedule, the relation of the layoff tax to the number of layoffs.

Ex-post wage bargaining reduces the scope for insurance. And it is optimal in this case to have a lower layoff tax rate than in the benchmark. This provides some income insurance, but at some efficiency cost.

Finally, heterogeneity of firms or of workers also typically implies a lower layoff tax rate than in the benchmark. To the extent that weak firms face higher layoff risk, a reduced layoff tax rate, together with a corresponding increase in the payroll tax, transfers resources from strong to weak firms, improving efficiency. A parallel argument holds for workers. A reduced layoff tax rate transfers resources from strong to weak workers, improving aggregate welfare. To the extent that the state can offer menus and let firms self-select, our results suggest offering lower layoff tax rates for weak firms, or weak workers. To the extent that the state can condition on observable characteristics, job subsidies for weak firms or weak workers provide a less distortionary solution than lower layoff tax rates.

We are very much aware however of the limits of our analysis. Even within our one-period model, there is a number of issues still to be explored. Let us mention two.

An important issue not taken up here is that 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 (2003), 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 were 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 that 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 (2001). 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 [1998] 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 two extensions of our model as crucial.

- 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). In this context, an important question is the role, if any, of mandatory individual unemployment accounts such as are being considered or introduced in a number of Latin American countries (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 is the role of experience rating systems, such as the US system, as ways of implementing the collection of layoff taxes. 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.

There is a final set of issues for which our analysis is useful but must also be extended. These issues indeed provided the initial motivation for our project. In most European countries, governments are considering reforms of their labor market institutions. In many emerging countries, governments are considering what labor market institutions to put in place. The challenge is to draw the practical implications of our conclusions for such purposes. We have tried to do this in Blanchard and Tirole (2003), focusing on France, characterizing existing institutions, and sketching a tentative path for reforms. But, here again, much more can and should be done.

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 (1998), “Search, Self-Insurance and Job Security provisions,” working paper WP98-2, Federal Reserve Bank of Chicago
- [5] Azariadis, C. (1975) “Implicit Contracts and Underemployment Equilibria,” *Journal of Political Economy*, 83: 1183–1202.
- [6] Baily, M.(1974) “Wages and Employment under Uncertain Demand,“ *Review of Economic Studies*, 41: 37–50.
- [7] Blanchard, O., and J. Tirole (2003), “Contours of Employment Protection Reform”, MIT working paper 03-35, November
- [8] Blanchard, O. and J. Tirole (2004), “Notes on Corporate Finance and Layoffs”, mimeo, January
- [9] Boeri, T. (2002) “Making Social Europe Compete,” mimeo, Universita Bocconi.
- [10] Cahuc, P. (2003), “Pour une Meilleure Protection de l’Emploi”, mimeo Paris I, France, May
- [11] Cahuc, P. and G. Jolivet, (2003), “Do We Need More Stringent Employment Protection Legislation?”, mimeo Paris I.
- [12] Hansen, G. and A. Imrohorglu, (1992) “The Role of Unemployment Insurance in an Economy with Liquidity Constraints and Moral Hazard”, *Journal of Political Economy*, 100-1, February, 118-142

- [13] Ichino, A., and R. Riphahn (2003) “The Effect of Employment Protection on Effort,” mimeo, EUI and University of Basel.
- [14] Mortensen, D., and C. Pissarides (1994) “Job Creation and Job Destruction in the Theory of Unemployment,” *Review of Economic Studies*, 61-3(208): 397–116.
- [15] Mortensen, D., and C. Pissarides (2001) “Taxes, Subsidies and Equilibrium Labor Market Outcomes”, mimeo, April
- [16] Pissarides, C. (2000) *Equilibrium Unemployment Theory*. Second edition, MIT Press.
- [17] Pissarides, C. (2002) “Consumption and Savings with Unemployment Risk: Implications for Optimal Employment Contracts,” mimeo, March