Efficiency Wage and Efficient Redundancy Pay

Giulio Fella*

Department of Economics, London School of Economics, London, UK

Department of Economics, Queen Mary and Westfield College, London, UK

September 1998

Abstract

A dynamic version of Shapiro and Stiglitz's shirking model features a form of inefficiency which is not captured by the original static model. Since, incentive compatibility *requires* workers to enjoy state-independent rents, any offer by redundant workers to take a wage cut is not credible, as it is not *ex post* incentive compatible. This implies that, if firms cannot commit on future firing *ex ante*, the number of redundancies is inefficiently high, as the externality, in the form of foregone rents, that firms impose on workers on severance cannot be traded. Redundancy payments make firms internalize the externality and fire less. Aggregate employment unambiguously increases and a Pareto improvement can be obtained provided all or part of the cost accrues to workers.

JEL Classification: E24, J32, J41, J65.

Keywords: Efficiency wage, firing costs.

1 Introduction

Dismissal costs in the form of mandatory severance payments, advance notice and other administrative and legal costs have recently received a great deal of attention in the eco-

*E-mail: G.Fella@qmw.ac.uk. Correspondence to: Department of Economics, Queen Mary and Westfield College, Mile End Road, London E1 4NS, UK.

nomic literature as a possible explanatory factor for the persistent differences between US and European unemployment rates over the last twenty years. While there is some empirical support to the thesis that job security provisions are positively correlated with measures of long-term unemployment, as reported in OECD (1994), the empirical evidence on a direct relationship with aggregate unemployment is mixed[1]. Formal modelling of the effect of dismissal costs on employment also yields ambiguous predictions. Dismissal restrictions affect hiring and firing in offsetting ways. In partial equilibrium, their net effect is small and ambiguous in sign: Bentolila and Bertola (1990) calibrate a realistic model of labour demand with non-stationary shocks and find that firing costs have a small positive impact on aggregate labour demand. On the other hand, general equilibrium models such as Hopenhayn and Rogerson (1993), in a Walrasian setup, and Millard and Mortensen (1997), in an individual matching framework, find significant negative effects on both employment and welfare. The two latter papers concentrate on cases in which severance payments result in increased wage payments and lower profits thus reducing aggregate job creation through a lower entry of firms or, equivalently, lower vacancy-posting.

As pointed out in Cabrales and Hopenhayn (1995), these results are very sensitive to apparently innocuous assumptions. It is not obvious that profits have to fall as a consequence of severance payments. First, wages could fall at unchanged revenues, as in Lazear (1990) or Bertola (1998). Second, revenues could increase more than wages. Job security could induce workers to invest more in firm specific human capital or produce higher effort, thus increasing revenues[2].

[1] Lazear (1990) finds that dismissal regulations are negatively related to employment levels in a panel data regression, while Bertola (1990) finds absence of correlation in a cross-country regression.

[2] This view is quite common in the industrial relations literature. See, for example, Piore (1986) and Osterman and Kochan (1990). The often cited examples are DEC, Eli Lilly, IBM which are committed to a policy of job security.

The present work aims to address the effect of dismissal costs on effort[3]. It analyses the general equilibrium effect of severance payments in an economy featuring involuntary labour mobility where firms pay efficiency wages in order to prevent shirking as in Shapiro and Stiglitz (1984).

The model is a simplified version of Saint-Paul (1995) insofar as it assumes, as in Bertola (1990), that the idiosyncratic productivity shock follows a two-state Markov process rather than a Poisson process with a continuum of states. All the results in the original model are preserved. Severance payments allow firms to commit to fire less in downturns. Since mobility has to be costly in order to promote effort, a lower dismissal probability results in a lower wage bill in good times. Also, a pure redundancy pay leaves hiring unaffected but reduces firing and so increases aggregate employment on impact.

The simplified nature of the model, though, provides additional insight. Firing costs, even if they do not accrue to workers, unambiguously increase employment in general equilibrium. More interestingly, and this is the main contribution of the paper, a pure severance payment monotonically increases welfare for any size up to the amount that maximizes the optimized value of hiring firms and equalizes wages across states. As incentive compatibility *requires* workers to enjoy state-independent rents, any offer by redundant workers to take a wage cut is not credible, since it is not *ex post* incentive compatible. This implies that, if firms cannot commit on future firing *ex ante*, the separation rate is inefficiently high, as the externality, in the form of foregone rents, that firms impose on workers on severance cannot be traded. Acting as a Pigovian tax, a pure severance payment increases welfare. It makes firms internalize the externality and fire less, while having no negative effect on hiring, as

[3]Given that the model assumes binary effort choices, job security can only affect the cost of promoting a *given* level of effort. Yet, the intuition is preserved.

workers accept lower wages in good states in exchange for a lower mobility cost in case of economic (but not disciplinary) dismissal. Welfare may still increase even if statutory job security provisions depress hiring by introducing a deadweight loss between the cost to the firm and the indemnity received by workers.

The paper is structured as follows. Section 2 introduces the economic environment and describes the model. Section 3 analyses the equilibrium. Section 4 presents the normative result. Section 5 concludes.

2 Economic environment

We consider a labour market populated by a continuum of firms indexed by n, where $n \in [0,1]$, and a homogeneous (hence anonymous) labour force normalized to one. Firms are price-takers on the product market. As in Shapiro and Stiglitz, workers are risk-neutral and have an instantaneous utility function given by $u(c, e) = c - e$, where c is consumption (assumed to be non-negative) and e is effort, measured in units of consumption. Each individual is endowed with one unit of labour. Effort choices are discrete: workers can decide whether to expend zero effort (shirk) or the positive amount e. We can write the utility associated with employment as $u(w_n, e) = w_n - eI_s$, where w_n is the wage at firm n and I_s is an indicator function equal to 0 if the worker shirks and 1 otherwise. The utility flow associated with unemployment is zero. So is the output of a worker who shirks.

Firms can only imperfectly monitor individual effort and cannot infer it from the total amount of output produced. Firm's revenues are (strictly) concave and are given by

$$R(l_n; \alpha_n), \qquad\qquad R' > 0,\ R'' < 0.$$

where l_n is the employment level at firm n. α_n indexes the revenue function and, as in

Bertola (1990), can take only two values: α_b if firms are hit by a bad productivity shock and α_g if a firm enjoys a favourable supply shock. Suppose $\alpha_g > \alpha_b$ and $\partial R'/\partial\alpha > 0,$ so that labour demand is higher for firms in the good state.

α_n follows a continuous Markov chain with symmetric transition probabilities[4] given by

$$\alpha_n(t+dt) = \begin{cases} \alpha_g \text{ with prob. } pdt \text{ if } \alpha_n(t) = \alpha_b, \text{ with prob. } (1-pdt) \text{ if } \alpha_n(t) = \alpha_g \\ \alpha_b \text{ with prob. } pdt \text{ if } \alpha_n(t) = \alpha_g, \text{ with prob. } (1-pdt) \text{ if } \alpha_n(t) = \alpha_b \end{cases}.$$

This implies that the ergodic probability that a given firm is enjoying good or bad business conditions is 0.5 and, since the number of firms is infinite, this is also the steady state proportion of firms in each state. Given that the total mass of firms is one, steady state aggregate employment equals $(l_g+l_b)/2$, where l_b and l_g are employment at the representative firm in the bad and the good state respectively.

We also assume $R'(l_g;\alpha_g) = R'(2-l_g;\alpha_b) = e$; i.e. the marginal product of labour at full employment equals its social cost. This ensures that full employment is efficient and jobs are rationed[5].

Firms cannot freely fire workers in downturns. They have to pay a redundancy cost F for each worker fired.

2.1 Hiring and firing

In the presence of turnover costs the firms' optimization problem is intertemporal. So the shadow value of the marginal worker at firm n, J_n, must satisfy

[4] Allowing for asymmetric transition probabilities would be straightforward but would just complicate the algebra and the notation with little gain in economic insight.

[5] That the marginal product of labour at full employment equals (rather than being above) its social cost is sufficient, not necessary to have job rationing. Market anonimity and job rationing ensure that it is (constrained) efficient for workers to receive rents and post no bond as argued in MacLeod and Malcomson (1993). This makes it clear that the result in the paper does not rely on any disguised bonding argument.

$$rJ_n = R'\left(l_n; \alpha_n\right) - \varphi_n + \frac{E\left[dJ_n\right]}{dt}, \tag{1}$$

where $r \geq 0$ is the discount rate and φ_n is the marginal cost of employment; i.e. the derivative of the wage bill with respect to firm-level employment. For any firm the flow equivalent of the shadow value of the marginal worker equals the instantaneous marginal profit plus the expected capital gain/loss. Given that the assumed Markov process is time-invariant and firms are identical, n indexes business conditions.

Turnover costs determine an optimal inaction range, so $l_g \geq l_b$ with equality if firms find it profitable not to respond to a shock. We assume that the change in business conditions is such as to generate positive turnover; i.e. $l_g > l_b$. A firm which has just been hit by a negative shock will fire workers up to the point where the shadow value of the marginal job equals minus the dismissal cost; i.e. $J_b = -F$. Replacing in (1) we get

$$-rF = R'\left(l_b; \alpha_b\right) - \varphi_b + p\left(J_g + F\right). \tag{2}$$

The last term is the expected change in the value of the job. The firm turns good with instantaneous probability p and the value of the job changes from $-F$ to J_g.

For a firm in the good state the shadow value of the marginal worker J_g is given by

$$rJ_g = R'\left(l_g; \alpha_g\right) - \varphi_g - p\left(F + J_g\right). \tag{3}$$

As opening a new job is costless, firms in the good state will increase employment up to the point where $J_g = 0$. Replacing J_g in (2) and (3), labour demand at firms in the bad and the good state is then implicitly given by

$$\varphi_b = R'\left(l_b; \alpha_b\right) + (p + r)F, \tag{4}$$

$$\varphi_g = R'\left(l_g; \alpha_g\right) - pF. \tag{5}$$

Equations (4) and (5) show that firing costs introduce a wedge between the marginal cost and the marginal revenue of labour. In "bad" firms marginal cost is higher than marginal revenue while the opposite is true in "good" firms. For a given marginal cost, firing costs increase employment in the former and reduce it in the latter through the well known option-value effect.

2.2 Workers' behaviour

Workers behave exactly as in Shapiro and Stiglitz. If we abstract from turnover costs for a moment, the only difference is that the separation rate is endogenous and depends on firms' transition probabilities. Workers can either provide the effort at a utility cost e or shirk. If a worker shirks she faces a probability q per unit of time of being caught and fired.

Also workers' decisions are intertemporal and must satisfy asset-value equations. Subscripts indicate the firms' state and superscripts the worker characteristics (s if a shirker, n if not).

E_b^i, the value of being employed at a firm in the bad state for a worker of type i ($i = n, s$) is given by

$$rE_b^s = w_b + q\left(U - E_b^s\right) + p\left(\max\left[E_g^s, E_g^n\right] - E_b^s\right), \tag{6}$$

$$rE_b^n = w_b - e + p\left(\max\left[E_g^s, E_g^n\right] - E_b^n\right), \tag{7}$$

where U is the value of being unemployed. According to (6) and (7), the value of being employed at a firm which is in a bad state, for both shirkers and non-shirkers, is given by the current wage plus the expected capital gain if the firm turns good with instantaneous probability p. When the firm's productivity changes, workers adopt the most profitable behaviour between shirking and not shirking. Their values differ insofar as non-shirkers bear the cost e of supplying the effort while shirkers are subject to a capital loss if they are detected shirking, an event which has an instantaneous probability q. We assume workers do not receive any payment in case of disciplinary dismissal[6].

Equations (8) and (9) below are the equivalent of (6) and (7) for the workers employed at a firm in the good state. We assume workers are randomly laid off.

$$rE_g^s = w_g + q\left(U - E_g^s\right) + p\left(1 - \frac{l_b}{l_g}\right)\left(U + Q - E_g^s\right) + p\frac{l_b}{l_g}\left(\max\left[E_b^s, E_b^n\right] - E_g^s\right) \tag{8}$$

and

$$rE_g^n = w_g - e + p\left(1 - \frac{l_b}{l_g}\right)\left(U + Q - E_g^n\right) + p\frac{l_b}{l_g}\left(\max\left[E_b^s, E_b^n\right] - E_g^n\right) \tag{9}$$

are the values of being employed at a firm experiencing good business conditions for a worker of type i. Given that on transiting into the bad state a firm will hold only a proportion l_b/l_g of its original labour force, all workers, shirkers and non-shirkers alike, face an instantaneous probability $p(l_b/l_g)$ of keeping their job but at a firm in the bad state and a probability $p(1-l_b/l_g)$ of becoming unemployed. The latter case entails a loss but also a lump-sum payment Q. As disciplinary dismissals do not entail any cost, a firm whose business conditions turn bad would have an incentive to cut the wages of the redundant workers below the incentive compatible level in order to induce them to shirk and fire them costlessly. To rule this out, we realistically assume that firms have to pay the same wage to homogeneous workers. Cutting wages would cause output to fall to zero. We allow Q, the severance payment the worker receives, to differ from F, the amount the firm pays. We discuss below how this affects the result.

The permanent income of an unemployed worker is

[6] The emphasis of the model is on the effect of redundancy payments. Legislated provisions that reduce the utility loss in case of disciplinary dismissal can just increase the rent that accrues to the worker.

$$rU = h\left(\max\left[E_g^s, E_g^n\right] - U\right), \tag{10}$$

where h is the exit rate from unemployment. As only firms in the good state hire new workers, the value of being unemployed equals the expected capital gain if hired by a firm whose business conditions have improved.

In each time interval the number of firms which switch from the good to the bad state is $p/2$ and each of them will fire $(l_g - l_b)$ workers. In steady state the outflow from unemployment must equal the inflow into it. So the exit rate from unemployment is implicitly given by

$$hu = \frac{p}{2}(l_g - l_b), \tag{11}$$

where u is the stock of unemployed workers which equals

$$u = 1 - \frac{l_g + l_b}{2} \tag{12}$$

given that in steady-state there are 0.5 firms in each state.

So we can rewrite (10) as

$$rU = h\left(\bar{l}_b, \bar{l}_g\right)\left(\max\left[E_g^s, E_g^n\right] - U\right). \tag{13}$$

where bars over variables indicate that the exit rate from unemployment is a function of average firm-level employment that firms take as given.

It is easy to verify that E_j^n and E_j^s $(j = b, g)$ are linear increasing functions of w_j with E_j^n having a lower intercept and higher slope than E_j^s ; i.e. they satisfy the single-crossing property. So, to prevent workers from shirking, firms have an incentive to raise wages up to the point where $E_j^n = E_j^s$. Subtracting respectively (6) from (7) and (8) from (9) and rearranging we get

$$E_g - U = E_b - U = \frac{e}{q}. \tag{14}$$

Equation (14) implies that the punishment on being caught shirking and fired must equal the expected effort a shirker would save before being caught. Keeping in mind that $E_j^n = E_j^s$ we can replace in (7) and (9) using (13) and (14) to get the incentive compatible wages

$$w_b = e\left(1+\frac{r}{q}\right)+h\left(\bar{l}_b,\bar{l}_g\right)\frac{e}{q} \tag{15}$$

and

$$w_g = e\left(1+\frac{r}{q}\right)+h\left(\bar{l}_b,\bar{l}_g\right)\frac{e}{q}+p\left(1-\frac{l_b}{l_g}\right)\left(\frac{e}{q}-Q\right). \tag{16}$$

The intuition behind equations (15) and (16) is the following. The second term in both equations is the permanent income from unemployment. The first one is the compensation for providing the effort. For workers not to shirk they must be compensated not only for the effort put forth but also for the flow opportunity cost of shirking. As long as $Q < e/q$, firms in the good state must pay a higher wage in order to compensate their workers for the mobility cost they bear in case they are made redundant. Mobility *has* to be painful for disciplinary dismissals to provide an effective punishment. In equilibrium no worker shirks, but the fact that mobility has to be costly to prevent shirking imposes a negative externality on workers who are dismissed for economic reasons. The wage differential - the last term in (16) - has to compensate for this cost. Only when p tends to 0 - i.e. when idiosyncratic shocks tend to be permanent - do all firms pay the same wage. In this limit case no worker is dismissed whatsoever in equilibrium and so no compensation is required.

The redundancy pay Q lowers w_g, with respect to the frictionless case, as it reduces the expected loss from being dismissed. Workers are willing to accept a lower wage now in exchange for the severance payment when fired. The transfer lowers the mobility cost for workers who are dismissed for economic reason while leaving unaffected the punishment for potential shirkers. This reduces the wage differential[7].

2.3 Wage setting

Equation (15) shows that firms in the bad state are effectively wage-takers. They cannot affect the unemployed permanent income as the exit rate from unemployment depends on average variables. So their marginal cost of employment is given by

$$\varphi_b = w_b = e\left(1+\frac{r}{q}\right) + h\left(\bar{l}_b, \bar{l}_g\right)\frac{e}{q}. \tag{17}$$

Firms in the good state, instead, control the premium they pay over wages in the bad state. Their employment choice affects the probability that a worker enters unemployment when business conditions turn bad, as can be seen from equation (16). Their wage bill is

$$w_g l_g = \left[e\left(1+\frac{r}{q}\right) + h\left(\bar{l}_b, \bar{l}_g\right)\frac{e}{q}\right] l_g + p\left(\frac{e}{q} - Q\right)\left(l_g - l_b\right). \tag{18}$$

The last term looks exactly like a cost of turnover and is conceptually equivalent. Firms in the good state have to compensate their workers for the risk of being dismissed. So turnover imposes a cost on firms in the form of a higher wage bill. Their marginal cost of employment, $\partial\left(w_g l_g\right)/\partial l_g$, is then

$$\varphi_g = e\left(1+\frac{r}{q}\right) + h\left(\bar{l}_b, \bar{l}_g\right)\frac{e}{q} + p\left(\frac{e}{q} - Q\right). \tag{19}$$

As in Lazear (1990), the marginal employment cost for hiring firms falls by the full size of the payment.

Note also that, for given l_g, the risk of becoming unemployed is lower for higher l_b. So employment at firms in the bad state has a positive externality on firms in the good state. The marginal benefit is $p\left(e/q - Q\right)$ as can be seen from equation (18). This important point

[7] Equations (15) and (16) show that w_g could fall below w_b for large enough values of Q. For a high enough payment mobility would actually be subsidized and workers would like to be dismissed. Anyway no arbitrage would be possible as workers have no control over their dismissal and they do not receive any payment if they quit.

was made by Saint-Paul (1995). Higher employment in downturns reduces the current wage bill in the good state by decreasing the dismissal probability $p\,(1 - l_b/l_g)$. The value of the firm is improved up to the point where the marginal current benefit from firing less in a downturn equals the marginal expected discounted loss from having to hoard labour in the bad state. As for the optimized value of firing firms, it is reduced by labour hoarding for a positive discount rate. At the moment of transiting into the bad state, any reduction in the wage bill in the good state is already sunk. The marginal expected gain accrues in the future and is discounted while the current marginal loss is not.

A time-inconsistency problem exists. Firms in the good state would like to commit *ex ante* to fire less, but this commitment is not credible as it would not be optimal from the firm's point of view *ex post.*

Severance payments allow a hiring firm to credibly commit to fire less in the event that business conditions turn bad[8].

Saint-Paul (1995) demonstrates that the level of redundancy pay which maximizes the optimized value of a firm in the good state is $Q = F = e/q$. Equations (15) and (16) show that this is also the amount that equalizes wages across states by eliminating the mobility cost involuntarily born by dismissed workers. Efficiency wage considerations do not require workers to actually bear any mobility cost, but only the *threat* of facing one in the event of underperformance.

[8]Obviously, firms and workers could achieve the same result by writing an explicit state-contingent employment contract. However, in a more realistic set up in which shocks to productivity can take more than just two values, writing an explicit contract that fixes a unique, state-independent, level of the severance payment is much easier than writing a contract that specifies the number of workers to fire in each possible state of the world.

3 Market equilibrium

3.1 Definition

In equilibrium firms choose employment so that its (shadow) marginal revenue, the right-hand-side of equations (4) and (5), equals its marginal cost, the right-hand-side of (17) and (19). So replacing for φ_b and φ_g using (17) and (19) in (4) and (5) respectively, results in the two equilibrium conditions

$$R'\left(l_b;\alpha_b\right) = e\left(1+\frac{r}{q}\right) + h\left(l_b,l_g\right)\frac{e}{q} - \left(p+r\right)F \tag{20}$$

and

$$R'\left(l_g;\alpha_g\right) = e\left(1+\frac{r}{q}\right) + h\left(l_b,l_g\right)\frac{e}{q} + p\left(\frac{e}{q} - Q + F\right), \tag{21}$$

where h is given by equations (11) and (12) and bars over l_b and l_g have been dropped as all firms in the same state are identical.

3.2 The effect of dismissal costs

Equations (20) and (21) show that, provided $F = Q$, severance payments do not alter the shadow cost of labour - the right-hand side of (20)-(21) - at firms in the good state but they reduce it at firms in the bad state. This result was already present in Saint-Paul (1995), yet its implications for the aggregate unemployment rate were not fully spelled out. In addition, that article did not consider the case in which the payment involves a deadweight loss (i.e. $F > Q$).

The general equilibrium effect of firing costs on employment is easier to derive in this simpler setup. Equations (20)-(21) show that the firing cost F unambiguously reduces turnover and so the inflow into unemployment - the right-hand-side of equation (11). For equilibrium to be reestablished, the outflow from unemployment, the left-hand-side of (11), has to fall. So either the unemployment rate u or the exit rate h or both must fall. Suppose first that $F > Q = 0$ (i.e. F is a pure red-tape administrative cost). Firing costs have only small and ambiguous effects on average labour demand as proved by Bertola (1990, 1992). Since the effect on turnover is larger than the impact effect on the unemployment rate, the exit rate from unemployment has to fall to reestablish equilibrium. This further boosts employment as it reduces the permanent income from unemployment and wages in both states. If $F = Q$ instead, it is clear from equations (20)-(21) that severance payments not only reduce turnover, but have a positive impact effect on aggregate employment. The direction of the general equilibrium change in the exit rate from unemployment is, then, ambiguous and so is the effect on wages. It is possible to prove, though, that the positive impact effect always prevails and aggregate employment unambiguously rises [9].

This finding that severance payments increase aggregate employment appears in contrast with Lazear's (1990) point according to which legislated severance payments can be undone if markets are perfect and complete. Unlike Lazear (1990), severance payments have a positive impact effect on employment in firing firms as they leave the wage they pay unaffected. The intuition is the following. In models featuring voluntary separation, such as Lazear's, redundancy pay raises wages in downturns as it tightens the employed workers' participation constraint by increasing their fallback utility. The employment effect is null as the increase in wages completely offsets the partial equilibrium fall in the shadow cost of labour induced by firing costs. In the present model, wages are determined by an incentive compatibility constraint. As unemployment is involuntary, the workers' participation constraint is never binding. Severance payments increase the expected utility of the workers made redundant

[9] The proof can be obtained on request to the author.

but, given that the workers who quit or are caught shirking are not entitled to them, they do not affect the incentive compatibility constraint. So they cannot push up wages in downturns[10].

So firing costs, whether totally ($F = Q$) or partially ($F > Q$) received by workers, increase aggregate employment[11].

Yet, higher employment does not necessarily result in higher welfare for workers. In a first-best world severance payments cannot improve on the decentralized outcome. In efficiency wage models in which profits are entirely distributed to workers, though, the *laissez faire* equilibrium is not, in general, Pareto optimal - unemployment is inefficiently high. This is the case, for example, in Shapiro and Stiglitz's model and has been proved for the general case in Greenwald and Stiglitz (1988). In our model aggregate welfare depends also on the cross-sectional distribution of employment. An employment-increasing intervention may or may not improve welfare depending on how it affects the distribution of employment across firms.

4 Efficient redundancy pay

For simplicity we discuss efficiency in terms of steady state undiscounted optimization. By continuity the qualitative result is unchanged for a small enough positive r.

The social planner maximizes steady state consumer utility. Given linear preferences this is tantamount to maximizing the utility of working plus any income transfer - namely

[10] Even with a positive probability that currently firing firms carry out further redundancies in the future the severance payment Q would reduce, not increase, the marginal labour cost in downturns.

[11] The result does not depend on the simple stochastic structure of the model. It does not rely on the fact that firms in the bad state do not carry out redundancies, but on the fact that severance payments do not push up wages. Also, allowing for entry/exit of firms is likely to strengthen the result, as the value of hiring firms is increased while the fall in the value of firing firms is small if r is, realistically, close to zero.

profits and the severance payment Q. This is equivalent to maximize the difference between total output and its social cost of production minus any waste of resources associated with redundancy pay; i.e.

$$W = \frac{1}{2}\left(R\left(l_b;\alpha_b\right) - el_b\right) + \frac{1}{2}\left(R\left(l_g;\alpha_g\right) - el_g\right) - \frac{p}{2}(l_g - l_b)(F - Q). \tag{22}$$

Provided they are fully received by workers ($F = Q$), redundancy payments have no direct effect on aggregate welfare because they are just a redistribution[12]. Obviously they do affect welfare through their effect on the allocation of labour. Since severance payments cancel out in (22) and leave the value of firms unaffected[13], they do not alter the social planner problem and we can harmlessly set them to zero in the welfare analysis.

The social planner has no monitoring advantage over private agents. She has to pay incentive compatible wages. Furthermore, she cannot enforce allocations which are not profitable from the private point of view. So she maximizes (22) subject, not only to the wage constraints (15) and (16), but also to

$$\frac{1}{2}\left(R\left(l_b;\alpha_b\right) - w_b l_b\right) + \frac{1}{2}\left(R\left(l_g;\alpha_g\right) - w_g l_g\right) = 0. \tag{23}$$

The intuition behind the expression in (23) is the following. The social planner can at best tax and redistribute pure profits. So total aggregate profits (or equivalently the value of firms) cannot be negative.

It can be easily proved that the three constraints are all binding. So it is possible to replace w_b and w_g in equation (23) using (15) and (16). The first best equilibrium can then be fully characterized by tangency between the resulting isoprofit locus in the $(l_{b,}l_g)$ space

[12]The fact that the gains may accrue to individuals who are different from the ones who bear the losses does not matter *ex ante* and, given no discounting, not even *ex post* since, as the individual distribution coincides with the cross-sectional one, all individuals have the same expected utility, .

[13]The optimized value of the firms in the two states satisfies the Bellman equations $rV_g = R\left(l_g;\alpha_g\right) - w_g l_g + p\left(V_b - V_g\right) - pF\left(l_g - l_b\right), rV_b = R\left(l_b;\alpha_b\right) - w_b l_b + p\left(V_g - V_b\right)$. If $F = Q$, replacing for w_g confirms that the value of firms is unaffected by dismissal costs at unchanged employment levels.

and a social indifference curve.

We know from Shapiro and Stiglitz (1984) that the "invisible hand" of the market would not achieve the social optimum in their model. Employment is inefficiently low in the decentralized equilibrium as the private cost of labour is above its social cost. Given that they assume an exogenous separation rate, this means that hiring and, hence, the exit rate from unemployment are inefficiently low in the decentralized equilibrium.

The endogeneity of the separation rate in the present model implies an additional source of inefficiency. As incentive compatibility *requires* workers to enjoy state-independent rents, any offer by redundant workers to take a wage cut is not credible, since it is not *ex post* incentive compatible. This implies that, if firms cannot commit on firing, the separation rate is inefficiently high, as the externality, in the form of foregone rents, that firms impose on workers on severance cannot be traded. We have shown in section 3.2 that the externality is pecuniary, as it is reflected in higher wages in good states. Firms enjoying good business conditions would *ex ante* internalize the externality and commit to fire less in downturns in order to reduce their current wage bill. Yet, in the absence of reputational or other commitment devices, their pledge is not time-consistent. Greenwald and Stiglitz (1986) proved that pecuniary externalities may not disappear - i.e. may become real externalities - if markets are not complete (or other real distortions are present). This is a case in point. Moral hazard on the part of workers and the time-inconsistency of firing decisions prevent the externality from being traded either in spot or future markets.

Given that in the decentralized equilibrium there is too little job creation and too much job destruction, a first-best intervention would involve an increase in employment in both states, namely an increase in hiring and a reduction in firing, until aggregate profits have been taxed away. So it would consist of a hiring subsidy and tax on redundancies financed through a lump-sum tax on firms' value. Dismissal payments do not address the inefficiency in job creation, but, by reducing job destruction, they are second-best Pareto improving. In fact, the following proposition can be proved.

Proposition 1. *If $Q = F$, the change in the frictionless allocation induced by severance payments is Pareto improving for a size of the severance payment lower or equal to the privately optimal size e/q.*

Proof. See appendix.

The intuition behind the latter result is the following. As a Pigovian tax, dismissal costs make firms internalize the externality, in the form of foregone rents, that they impose on dismissed workers. If firing costs did not accrue to workers ($Q = 0$), they would increase the shadow cost of labour, the right hand side of equation (21), and depress hiring in the good state. This second effect is clearly welfare reducing. In so far as it involves a payment to workers, redundancy pay reduces the latter's mobility cost in case of economic dismissal, hence resulting in lower wages in good states. In the case of a pure severance payment, the fall in wages fully offsets the partial equilibrium effect and the firing cost works as a tax on separation, but does not depress hiring. The privately optimal size is also Pareto improving[14] because it is the size which determines the same allocation of labour which would prevail if future markets for state-contingent employment existed. If this were the case, the externality would be internalized by the price mechanism in the *laissez faire* equilibrium.

Outside from the case of a pure severance payment, the effect of dismissal costs on aggregate welfare is ambiguous. Government intervention in the form of mandatory redundancy

[14] Lack of reputational mechanims and costs of writing private contracts may partly explain the existence of legislated dismissal regulations. The time inconsistency problem may explain how political pressures to scrap job security provisions may arise in response to changes in the balance between the number of firms in the good and bad state.

pay is likely to entail significant deadweight losses insofar as it differs from the privately optimal size or it introduces red-tape and enforcement costs. In this case $Q < F$ and the net outcome is the result of three different effects. First, the reduction in firing enhances welfare. Second, hiring falls as the net implicit turnover cost increases with $(F - Q)$. Despite that aggregate employment unambiguously increases, as we have argued in section 3.2, this reallocation of labour from high to low productivity firms reduces output at unchanged aggregate employment. Finally, as part of the cost born by firms is just a waste of resources, firing costs entail a deadweight loss given by the term $p(l_g - l_b)(F - Q)/2$ in equation (22). These two latter effects are obviously welfare reducing.

The net result is more likely to be negative the higher the variability of marginal productivity and employment across states; i.e. the higher is the fall in output at unchanged employment and the higher is the deadweight loss associated with turnover.

To illustrate the effect of firing costs on aggregate welfare we simulate the model for different values of the deadweight loss. The production function is assumed to be Cobb-Douglas with labour share equal to 0.64, which is standard in the Real Business Cycle literature. The Cobb-Douglas specification allows to normalize e to 1. In fact, if shocks are multiplicative and severance payments are realistically modelled as a linear function of the wage, the system (20)-(21) is homogeneous of degree zero in e. The multiplicative productivity shocks are $\alpha_g = 2.7$ and $\alpha_b = 1.9$, which imply that revenues increase by 35 percent at unchanged employment as a consequence of a positive productivity shocks. We also assume $r = 0$, $q = 1$ and $p = 0.15$. On average, a shirker can expect not to be caught for one year and firms experience shocks of the size we consider every six and a half years. The chosen parameter values imply a significant fluctuation in employment between states. In the decentralized equilibrium, firm-level employment falls by fifty per cent in response

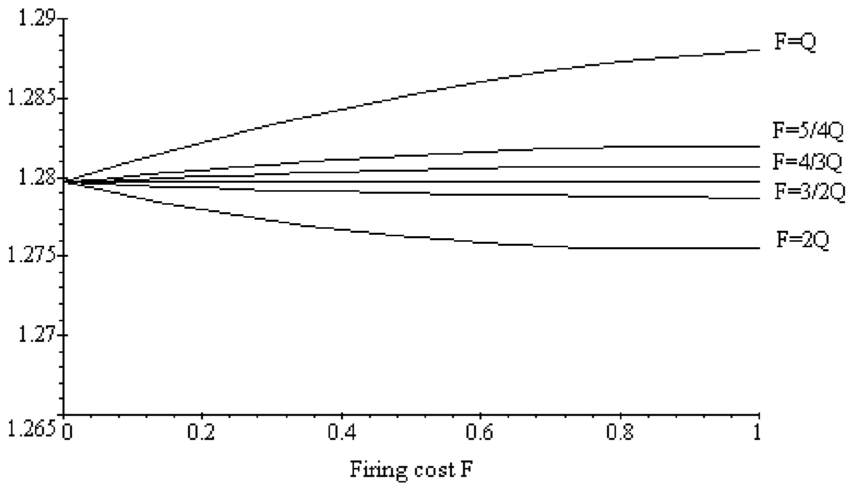

Figure 1: Effect of firing costs on welfare for different deadweight component

to a negative shock: $l_g = 1.2$ and $l_b = 0.6$. The qualitative results are unchanged for any parameterization.

The results are reported in figure 1. On the horizontal axis is the size of the payment F born by the firm. Its maximum value of one corresponds to e/q, the value which ensures efficient separation.

The highest curve in figure 1 corresponds to the case of a pure severance payment ($F = Q$). The others correspond, in decreasing order, to the cases in which the unit deadweight loss equals respectively one fourth, one third, one half and the whole severance payment Q which accrues to the worker. Apart from the highest curve, all curves have a discontinuity at $F = e/q$. When the cost born by the firm equals workers' mobility cost, the parties can economize on third-party payments by labelling the separation a quit. If this is anticipated by the agents involved, aggregate welfare is the same as under a pure severance payment equal to e/q.

It is apparent that even a small deadweight component significantly reduces the efficiency enhancing effect of severance payments. Yet, even for the very high employment variability implied by our parameterization, firing costs reduce welfare below its *laissez faire* level only if they entail a rather sizeable unit deadweight loss in the range of half the severance payment Q or higher.

From a policy perspective, the relevant question is then how large the deadweight loss is in practice. The answer is not easy as the loss encompasses all wastes of resources associated with dismissal regulations: higher monitoring and litigation costs, paperwork, etc.[15] Among European countries it seems safe to say that the loss is unlikely to be high in countries like the UK and Ireland where job security takes mainly the form of a pure (and not very generous) severance payment and litigation is infrequent. This is somehow confirmed, and the mechanism outlined in the paper supported, by the fact that in a sizeable number of UK firms negotiated severance payments significantly exceed statutory minima, as documented in Booth (1987). On the other hand, dismissal regulations involve significant deadweight losses in some Mediterranean countries. Greece and Portugal are characterized by quantitative constraints on the number of workers that can be fired. In Italy, the interaction of an ambiguous discipline and courts' bias against workforce reductions result in wasteful bargaining and delays. Fella (1997) contains a detailed discussion of job security provisions in OECD countries.

5 Conclusion

The simple model in this paper highlights important dynamic implications of efficiency wage theories. If firms cannot commit on future firing, job destruction is inefficiently high in the

[15] These costs need not be additive. For example, compulsory negotiation with workforce representatives is likely to reduce the probability of litigation in court. Furthermore, the parties can always agree on transfers that induce voluntary separation and save on third-party payments.

decentralized equilibrium, as firms do not take into account the negative externality, in the form of foregone rents, that they impose on dismissed workers. The price mechanism cannot internalize the externality, due to both spot and future markets incompleteness, as moral hazard prevents workers to credibly offer to take a wage cut despite a positive rent from continuation and firms' *ex ante* pledges to restrain firing in downturns are not time-consistent.

Firing costs not only unambiguously increase employment, but, acting as a Pigovian tax on firing, they make firms internalize the externality. Thus, they can be Pareto improving, if they at least partly accrue to workers.

Clearly, redundancy pay can improve welfare for other reasons. In Bertola (1998) it reduces income uncertainty, thus resulting in higher welfare for risk-averse, borrowing constrained individuals. Also, severance payments could improve firm/worker matches. Being sunk, they would not affect the search behaviour of borrowing unconstrained individuals, but would allow wealth constrained workers to be more selective. The present paper, though, highlights an additional channel which is relevant for any model of the labour market featuring involuntary separation.

Acknowledgements

This paper is an abridged version of a chapter from my PhD thesis at LSE. I am indebted to Charlie Bean for his constant advice and encouragement. My thinking on this paper has greatly benefited from Giuseppe Bertola's insightful remarks and suggestions. Andrea Ichino, Claudio Michelacci, seminar participants at Research Strategy Seminars at LSE and two anonymous referees provided useful comments. Needless to say, I hold full responsibility for all remaining shortcomings.

Appendix: Proof of proposition 1

Proposition 1. *If $Q = F$, the change in the frictionless allocation induced by severance payments is Pareto improving for a size of the severance payment lower or equal to the privately optimal size e/q.*

Proof. Given that the three constraints are satisfied as equalities, the Pareto optimality condition can be rewritten as

$$\frac{R'(l_b;\alpha_b)-e}{R'(l_g;\alpha_g)-e}=\frac{A-\left(R'(l_b;\alpha_b)-w_b+p\frac{e}{q}\right)}{B-\left(R'(l_g;\alpha_g)-w_b-p\frac{e}{q}\right)}, \tag{24}$$

where

$$A=\frac{2ep\,(l_b+l_g)\left(l_g-1\right)}{q\left(2-l_g-l_b\right)^2} \tag{25}$$

and

$$B=\frac{2ep\,(l_b+l_g)\,(1-l_b)}{q\left(2-l_g-l_b\right)^2}. \tag{26}$$

Equation (24) states that at the social optimum a social indifference curve must be tangent to an isoprofit line in the (l_b, l_g) space. The left-hand-side of (24) is the social marginal rate of substitution and the right-hand-side the slope of an isoprofit line. If the right-hand-side of (24) were greater than the left-hand-side, it would be possible to increase welfare by exchanging l_b for l_g at a rate that left profits unchanged. By evaluating the two sides of (24) at the privately chosen allocation in the presence of firing cost we can prove the result. We can replace for $R'(l_b;\alpha_b)$ and $R'(l_g;\alpha_g)$ using (20) and (21) and remembering that $Q = F$ and $r = 0$. We guess that the left-hand-side of (24) is not smaller than its right-hand-side; i.e.

$$\frac{w_b-e-pF}{w_b-e+p\frac{e}{q}}\geq\frac{A-p\left(\frac{e}{q}-F\right)}{B}. \tag{27}$$

Replacing for w_b and after a little of algebraic manipulation (27) can be rewritten as

$$p\left(\frac{e}{q}-F\right) \geq -p\left(\frac{e}{q}-F\right)\left(\frac{2-l_g-l_b}{l_g+l_b}\right) \tag{28}$$

which proves that our guess is right for $F \leq e/q$. So the exchange of l_b for l_g induced by firing costs is welfare improving for any value of F up to e/q. QED ■

References

Bentolila, S. and G. Bertola, 1990, Firing Costs and Labour Demand: How Bad is Eurosclerosis, Review of Economic Studies 57, 381-402.

Bertola, G., 1990, Job Security, Employment and Wages, European Economic Review 34, 851-886.

Bertola, G., 1992, Labour Turnover Costs and Average Labour Demand, Journal of Labor Economics 10, 389-411.

Bertola, G., 1998, Uninsurable Risk, Efficiency and Labor Market Rigidities, Mimeo, forthcoming on Economic Policy.

Booth, A. L.,1987, Extra-statutory Redundancy Payments in Britain, British Journal of Industrial Relations 25, 401-419.

Cabrales, A. and H. A. Hopenhayn, 1995, Job Dynamics, Correlated Shocks and Wage Profiles, Mimeo.

Fella, G., 1997, Firing Restrictions in OECD Countries, Mimeo.

Greenwald, B. C. and J. E. Stiglitz, 1986, Externalities in Economies with Imperfect Information and Incomplete Markets, Quarterly Journal of Economics 101, 229-264.

Hopenhayn, H. A. and R. Rogerson, 1993, Job Turnover and Policy Evaluation: a General Equilibrium Analysis, Journal of Political Economy 101, 915-938.

Lazear, E. P., 1990, Job Security Provisions and Employment, Quarterly Journal of Economics 105, 699-726.

MacLeod, W. B. and J. M. Malcomson, 1993, Wage Premiums and Profit Maximization in Efficiency Wage Models, European Economic Review 37, 1223-1249.

Millard, S. P. and D. T. Mortensen, 1997, The Unemployment and Welfare Effects of Labour Market Policies: a Comparison of the US and the UK, in: Dennis J. Snower and Guillermo de la Dehesa, eds., Unemployment Policy: Government Options for the Labour Market (Cambridge University Press, Cambridge: UK), 545-571.

Osterman, P. and T. A. Kochan, 1990, Employment Security and Employment Policy: an Assessment of the Issues, in: Katherine J. Abrahams and Robert B. McKersie, eds., New Developments in the Labor Market (The MIT Press, Cambridge: Massachusetts), 155-184.

Piore, M. J., 1986, Perspectives on Labour Market Flexibility, Industrial Relations 85, 146-166.

Saint-Paul, G., 1995, Efficiency Wage, Commitment and Hysteresis, Annales d'Economie et de Statistique 37-38, 39-53.

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