Inside Severance Pay

Tito Boeri[a], Pietro Garibaldi[b], Espen R. Moen[c] *

November 4, 2016

Abstract

All OECD countries have either legally mandated severance pay or compensations imposed by industry-level bargaining in case of employer initiated job separations. The paper shows that mandatory severance is optimal in presence of wage deferrals induced by workers' moral hazard. We also establish a link between optimal severance and efficiency of the legal system and characterize the effects of shifting the burden of proof from the employer to the worker. Quantitatively, the welfare effects of suboptimal severance payments vary in general equilibrium between 1 and 3 percent. The model accounts also for two neglected features of the legislation. The first is the discretion of judges in declaring the nature, economic vs. disciplinary, of the layoff. The second feature regards the relationship between severance and tenure. Our theory gives necessary conditions under which optimal severance is increasing with tenure, as generally observed.

Keywords: Severance, Unfair Dismissal, Graded Security, Legal Systems.
JEL codes: J63,J65,J33.

*[a] IGIER-Bocconi University. [b]Collegio Carlo Alberto, University of Torino. [c]Norwegian Business School.

Introduction

Most OECD countries have legally mandated severance pay in case of employer initiated job separations. When rules are not specified by the law, it is collective bargaining at the industry or national level to mandate severance to individual employers. For instance, Kodrzycki [30] reports that 86% of workers in Massachusetts are covered by a severance pay agreement, involving one week's wage per year of service. Such employer-employee transfers are the most important component of individual dismissal costs[1]. The average compensation for unfair dismissals is about two years of pay in case of a worker with at least 20 years of tenure. According to Garibaldi and Violante [24] who estimated the red tape costs of layoffs in Italy, severance pay accounts for about 2/3 of total dismissal costs. Severance pay also accounts for almost 50% per cent of the cross-country variation in the OECD index of the strictness of employment protection legislation (EPL) for regular workers, the reference measure of EPL in the literature. It is mandatory even in countries with negligible firing taxes to be paid to third parties.[2]

A fundamental difference between severance payment and firing taxes is that the latter is paid to a third party, while the former is a pure transfer. Severance pay differs from firing taxes also in that it is generally dependent of tenure. Furthermore, the amount of severance pay depends both on the nature – disciplinary vs. economic – of the dismissal, and on whether it is deemed fair or unfair by a Court ruling.

Reforms of these regulations are high on the policy agenda and have been explicitly requested by IFI (International Financial Institutions) to the so-called program countries during the Eurozone public debt crisis. Distinguished economists active in the US policy debate have been suggesting that severance should be increased at least during downturns, while several labor economists in Europe have been advocating the introduction of a tenure-related security contract involving severance being gradually increasing with tenure.

We still lack a proper framework to evaluate severance pay. The extensive literature on Employment Protection Legislation fails to characterise some of the key features of mandatory severance. In particular, it does not take into account that the size of these transfers typically varies depending on the seniority of the worker, on the nature – economic or disciplinary – of the dismissal and on its legitimacy (fair or unfair), as established in a Court ruling.

The purpose of this paper is twofold. First, we provide a normative framework, aimed at extending and systematizing earlier results as to the efficiency of severance pay. Second, we model the relationship between tenure and severance, and we dig into the features of the judicial systems, notably characterizing how the efficiency of the legal system and the rules concerning the burden of proof affect the optimal design of severance. We show that mandatory severance is optimal even in the absence of risk aversion and when there are no ex-ante rents to be split between the worker and the firm, unlike in the efficiency wage literature. Therefore, the "bonding critique" does not apply. What is sufficient to make severance efficient is wage deferrals motivated by deterrence of opportunistic behavior of workers and constraints to wage renegotiations *ex-post.* We provide a formal argument of why severance should be enforced by a co-ordinating mechanism outside the single firm, which is based on the fact that adverse selection stands on the way of severance schemes introduced by individual employers. This mechanism is coherent with the static model of Levine [34], where efficient hiring levels cannot be achieved when workers are heterogeneous and wages must satisfy a non shirking constraint à la Shapiro Stiglitz. Our model is dynamic and job destruction is properly modeled. In addition, our results hold also in general equilibrium, and we assess quantitatively the impact of removing severance pay. Under reasonable parameter values, the welfare effects of removing severance pay are between 1 and 3 percent, depending on the size of the initial severance pay. Our results suggest also that severance should be increasing in the inefficiency of the legal system. We also provide necessary conditions under which the optimal severance pay is increasing with tenure. Our results are empirically relevant. Legal rules about the severance-tenure profile appear to be positively correlated with the wage-tenure profile that we estimate drawing on longitudinal data. We also find that OECD measures of efficiency of judicial systems are correlated with severance pay for individual economic and unfair dismissals

[1] Group layoffs, that is, collective dismissals involving a discrete number of workers of the same firm, are not considered in this paper

[2] See Postal-Vinay and Turon [48], and Boeri [10] for a theory of severance pay as a device to buy time and avoid paying firing taxes, in presence of on-the-job search.

in a way which is consistent with the implications of the model.

The plan of the paper is as follows. Part one reviews our contribution to the extended literature on EPL. Next, it characterizes two neglected features of EPL, the discretion of judges in setting the level of severance pay depending on whether the individual dismissal is disciplinary, economic, fair or unfair, and the tenure profile of severance pay. Part two presents the model with moral hazard of the employees, and evaluates optimal severance pay under these circumstances. It also provides a formal argument of why severance should be mandated to individual employers. Part three extends the model looking into Court rulings as to the nature of dismissals, endogenizing the probability that not investing workers get severance pay for economic dismissal and that the dismissal is considered unfair. Part four extends the results to the general equilibrium, and provides numerical simulations of the effects of different levels of severance pay on welfare and unemployment. Part five goes back to the data investigating the correlation between severance and efficiency of judicial systems, and the severance tenure profile under different regimes as to wage deferrals and involvement of Courts in layoff procedures. The final section summarizes our key results and concludes.

1 Severance pay: literature and neglected features

1.1 Our contribution to the literature

Employment protection legislation is one of the most widely investigated institutions in the labor market. [3] The theoretical literature, pioneered by Bentolila and Bertola [4], Bertola [6] and Lazear [33], typically treats EPL as a firing tax to be dissipated or paid to a third party by the employer in case of a layoff. Severance pay - a transfer from the employer to the worker contingent on employer initiated separations-[4] is generally not framed in these models, as Lazear [33] neutrality result indicates that, with wage flexibility and risk neutrality, it only affects the tenure profile of wages leaving employment, hiring and separations unaffected. When instead, wages are rigid, severance pay increases unemployment (Garibaldi and Violante [24]).

Why do we need then severance pay then? There are three key rationales for severance pay according to the literature.

The first draws on moral hazard and adopts the standard setup of the efficiency wage models à la Shapiro-Stiglitz [54]. A severance paid to fired workers that did not shirk acts as a commitment device to an employment policy that does not strongly react to negative shocks. By playing this role, severance reduces labor costs. Without the severance, wages would have to be increased to deter shirking. Fella ([18] and [19]) draws on this initial intuition by Saint-Paul ([51]) to show that an optimal severance can be as high as to equalize wages across all possible productivity realizations. These results have been extended by Baumann [3] to the case where even some shirkers can receive severance pay and to double moral hazard (of employers choosing over projects having different levels of risk, in addition to employees deciding as to whether to put effort). This extension builds on the work by Galdon-Sanchez and Guell ([21]) who introduced the possibility that shirkers 'can get away with it' in a standard model of employment protection, but did not evaluate the efficiency properties of severance schemes. Stähler [52] extends the idea of Galdon-Sanchez and Guell to explicitly consider juidical mistakes for disciplinary dismissal, but he also does not consider efficiency issues.

This explanation of severance is subject to the 'bonding critique' which challenges the efficiency wage literature. A cheaper deterrent to opportunistic behavior is for firms to commit to a wage schedule offering initially lower wages (even below the marginal product) and higher wages if confirmed in the firm. In other words, severance pay is not needed when there are wage deferrals allowing for wages increasing with tenure as those documented by the empirical literature estimating Mincer-type wage equations.

The second argument for severance pay rests on risk aversion. Severance pay protects workers against uninsurable labor market risk, just like unemployment benefits do. Under full insurance, Blanchard and Tirole ([9]) show that severance pay has to be preferred to unemployment benefits because it internalizes the costs of layoffs. By the same token, employers could pay themselves the unemployment benefits. There is, in other words, a full substitutability between severance pay and unemployment benefits when the latter can be

[3]See Boeri and vanOurs [11] for a review of this literature.

[4]Our definition of severance clearly does not encompass deferred compensation schemes, such as private pension arrangements, which are paid at retirement or at any separation, including voluntary quits.

experience-rated.[5] When full insurance is not feasible, there is no longer full substitutability between the two institutions. There is always a role for the state in the provision of unemployment insurance and severance is not fully crowded out by unemployment benefits. In this explanation the borders between severance and unemployment benefits are not very well defined. It is debatable whether this is a theory of severance or a theory of unemployment benefits. What is clear is that it involves a substitutability between the two institutions.

A third rationale for severance pay is related to wage rigidities. With constant wages, search frictions and idyosincratic shocks to productivity, severance prevents excess job reallocation, as in the model by Alvarez and Veracierto [1]. This needs once more to posit a counterfactual wage rigidity – the fact that wages are constant over the lifetime – to attribute a welfare enhancing role to severance pay.

In our model, severance pay is efficient even when workers are risk-neutral and entry wages are fully flexible. Still wage deferrals needed to incentivize workers create a rationale for severance pay; once the firm is committed to a wage schedule, it may find it ex-post optimal to layoff a worker even when the job is generating some surplus over the value of unemployment. This wage schedule is not exogenous, but it is optimized at the time of job creation, and cannot be made contingent on a random component of output, as we further discuss in the next section. Furthermore, there is no substitutability between unemployment benefits and severance pay. The two institutions are complementary as unemployment benefits, under a broad set of circumstances, make the wage tenure profile steeper, inducing more inefficient separations when severance pay is absent. In our model, wages are deferred in order to incentivize workers to invest in job-specific productivity. Investment should be interpreted broadly. The key element is that the cost of the activity is realized before the return, so that wages will be deferred. Relevant investments include firm-specific training as well as effort for which the returns materialize at a future point in time. Some evidence coherent with the existence of these types of investment costs is presented in the discussion section.

The underlying assumption is that firms can commit to a future wage schedule (not contingent on individual productivity), but not on the employment relationship. Hence a firm cannot commit not to fire a worker if that is in the firm's interest *ex-post*, or to any payments to workers who are fired. This is a standard assumption in the literature (see Menzio and Moen [39] and the references therein). Under these conditions, severance deals with the moral hazard problem associated with firms firing too frequently senior workers receiving deferred wages. The result is general as wage deferrals are a common feature of labor markets. Moreover, studies measuring both wages and productivity (e.g., Medoff and Abraham [38], Kotlikoff and Gokhale [31] Flabbi and Ichino [27]) suggest that the effects of seniority on wage profiles can be attributed mainly to incentive reasons, and are not necessarily associated with a higher productivity of senior workers. There is also indirect evidence of deferred compensation. For instance, it is consistent with the findings by Lazear and Moore [32], who compared seniority-earning profiles of employees and self-employed (for which no agency problem arises) and by Barth [2], who compared the wage-tenure profile of workers paid piece-rate with that of workers receiving a flat wage.

We also relate to specific features of severance pay. The theoretical literature on EPL typically treats severance as a deterministic transfer from the employer to the employee. In the few cases where stochastic severance is allowed (Garibaldi [22], Malo [37]), it is modeled more as an option to fire (a firing permission) than as a distribution of alternative costs of dismissals. Moreover, no reference is made by this literature to the moral hazard problem related to the distinction between economic and disciplinary dismissals. Two partial exceptions are Galdon-Sanchez [21] and Boeri [12]. However, Galdon-Sanchez [21] operates on a reduced form model and both Boeri [12] and Galdon-Sanchez [21] do not address the efficiency of severance pay, but only consider its effects on unemployment and the layoff behavior of firms of different size.[6]

1.2 Neglected features

Our model rationalizes two neglected features of Severance Pay that affect the cost of individual dismissal. We first highlight these features, and we then provide some cross country evidence.

[5] Pissarides [46] shows in a model with risk aversion that when severance pay is optimally set, exogenous unemployment benefit does not influence equilibrium unemployment.

[6] R*ü*hmann and *Sü*dekum [50] consider the efficiency of severance payments in terms of human capital investment, but do not address the moral hazard problems associated with Court involvement and the severance-tenure profile.

The first relates to the discretion of judges in deciding upon the fairness and the nature (economic vs. disciplinary) of the dismissal.[7] Compensation is generally not offered to workers being fired for disciplinary reasons unless a Court ruling declares that the dismissal is unfair. When the individual layoff is motivated by the economic conditions of the firm, that is, it occurs independently of the behavior of the worker, compensation is typically offered also for fair dismissals, that is, cases where there is no evidence of opportunistic behavior of the employer. In the case of unfair dismissals compensation is higher than the severance for fair economic dismissals. There are countries in which compensation is provided only for unfair dismissals and fair economic dismissals do not involve mandated severance to the workers. In the case of Germany, Goerke and Pannenberg [25] show that actual EPL is extremely difficult to estimate, and the validity of dismissals defined by labor courts significantly affects the incidence and magnitude of severance pay. Due to these wide differences in the levels of compensation related to the nature of dismissals, there are strong incentives for the employee or the employer to bring the case before a Court. Involvement of judges cannot be avoided by state contingent contracts, and the decisions of the judges will be based on limited information. The judicial discretion clearly affects also private settlements out of Court, as such settlements will be based on the expected costs had the case gone to Court. These relevant interactions between EPL and the efficiency of judicial systems have to a large extent been neglected to date by the theoretical literature on EPL although there is evidence (Fraisse, Kramarz and Prost [20]) that the organizational structure of judicial systems does affect significantly labor market outcomes. An exception is Stähler [52], who do consider judicial mistakes, but he does not focus on the burden of proof. In addition, judicial mistakes and uncertainty are empirical relvent. Ichino et al.[28] find that in the case of a large Italian Bank, 22% of all dismissals are taken to Court and 17% of the dismissals are overruled by the latter.

The second neglected characteristic of EPL is the *tenure profile* of severance pay. As documented below and in the Web annex, most countries allow for mandated severance pay to be increasing with tenure. We are not aware of any theory rationalizing these arrangements. Chéron et al. [15] study age dependent employment protection, and argue that firing taxes should be hump shaped with respect to age, but they focus on firing taxes and ignore severance payments. Personnel economics offers explanations for why firms offer *tenured jobs*, that is, positions that cannot be severed under any set of circumstances. Tenure prevents the strategic choice of incumbents of hiring only low quality workers in order to reduce competition with outsiders (Carmichael [14]). These theories explain why employers may decide to commit not to layoff some workers, but do not explain why a *mandated* profile of severance increasing with tenure is chosen for potentially *all* private firms. Moreover, these models do not address problems of commitment: private firms generally cannot credibly commit not to layoff some workers, irrespective of their performance.

Judicial discretion and burden of proof

Statutory severance pay depends on the nature, economic vs. disciplinary, and on the fairness of dismissals. Fairness in the case of *economic dismissals* refers to the behavior of the employer: she should have tried as much as possible to avoid this outcome. Although the definition of fair economic dismissal is country specific, it generally implies that some "genuine and serious" exogenous shocks hit the firm.

In the case of *disciplinary dismissals*, the fairness refers to the behavior of the worker. In fair disciplinary dismissals there is evidence of misconduct on the part of the worker, where "misconduct" is often not defined, and the burden of proof typically falls onto the employer. When the economic or disciplinary dismissal is ruled "unfair", the amount of severance pay is significantly higher. Moreover, the employer, in addition to providing severance pay, typically has to pay the legal costs of the employee and compensate for the foregone months of pay during the legal procedure. In some countries, the employer is forced to reinstate the worker.

The decisions as to the nature of the dismissal and its fairness require some Court ruling. In practice, disputes are mostly settled before the Court decision, taking in consideration the nature of the dismissal, the probability that is considered fair and the severance and additional compensations envisaged under the different circumstances. Thus, in practice the level of severance ultimately depends on decisions made by third parties having limited information on the behavior of workers and employers. For all of these reasons

[7] A notable theoretical exception is the paper by Deffains et al. [16]. They study the effect of the judicial aptitude in labor disputes to balance out the parties' ex-post bargaining power in court.

the actual costs of layoffs are stochastic, and generally depend on the evidence that the employer can provide for a disciplinary or economic dismissal.

Table 1 displays the maximum compensation (severance pay plus notice period) required for fair economic, fair disciplinary, and unfair dismissals in in OECD countries. The table is based on the analysis of the country files used by the OECD in building up the summary measure of strictness of EPL, a report prepared for a European conference of labor lawyers [17], a study by the ILO [7] and a recent survey of Civil Justice also carried out by OECD (Palumbo [45]).

As shown by Table 1, in all countries even fair dismissals command some compensation to the worker, either in terms strictly of severance pay or of a minimum notice period (de facto an extension of pay after the date when the worker is made redundant).[8] The compensation for unfair dismissals (T_U, first column) is, however, always higher than that provided in case of fair dismissals (either economic, T_F^E, or disciplinary, T_F^D, second and third columns). One of the reasons why unfair dismissals cost more than fair dismissals is that in several countries (see Table A2 in the Annex), in addition to a monetary compensation, an unfair dismissal may also be sanctioned with the compulsory reinstatement of the worker in the ranks of the firm.[9] Thus, we estimate the costs of unfair dismissals as given by the statutory notice period (N) and severance (S), which is itself the sum of the severance for fair dismissals S_F and the additional compensation for unfair dismissals S_U, plus, limited to the countries with reinstatement, the average length of the trial period (d) and the compensation for unfair dismissal, which is a proxy for the cost of the reinstatement, the latter two terms multiplied by the likelihood that a reintegration of the worker is actually imposed by the Court. In particular, let (π) be the probability that a reintegration of the worker is imposed by the Court. The costs of unfair dismissals, T_U are given by:

$$T_U = N + S_F + S_U + \pi(d + S_U),$$

where the unit of measurement is monthly wages. As detailed in Table A1 in the web Annex of the paper, we attribute to π the value obtained by standardizing to the unit interval the 0-3 OECD index on the likelihood of the reinstatement, where 0 means never reinstatement and 3 denotes the case where employees can freely decide upon the reinstatement in the case where the dismissal is ruled to be unfair.[10] As shown by the fourth and fifth columns of Table 1, unfair dismissals are generally more than twice as expensive as fair economic dismissals, while the latter are more expensive than fair disciplinary dismissals, which typically involve no compensation, and only a relatively short notice period.

The documented wide differences in the costs of dismissals (and in the compensation for workers) can be summarized in the index of dispersion provided in the web Annex (Table A2). This dispersion creates strong incentives for the involvement of judges in dismissal procedures, hence of uncertainty associated with judicial discretion. Another fact highlighted by the last column on the right-hand-side of Table 1, is the presence of a significant cross-country variation in legal rules concerning the burden of proof. Most countries put the burden on the employer, but there are cases even outside Eastern Europe (e.g., France, Denmark, Switzerland, and Japan) where the burden of proof falls partly on the worker according to the OECD.

The elasticity of severance to tenure

In 25 countries out of 30 there is evidence of severance increasing with tenure (Figure A1 in the web Annex). If we add the notice period (de facto an extension of the contract after the notification of the dismissal giving

[8] Consistently with the literature on employment protection, we consider the notice period as an extension of the pure severance. The notice period is typically used by workers to seek for alternative employment. The empirical literature also finds that severance and notice have the same effects on labor market flows. In any event, tables without the notice period are available, upon request, from the authors.

[9] In these countries the costs of unfair dismissals should include the duration of the trial period, as reinstated workers should be back paid the full wage between the date of the dismissal and that of the Court ruling, and an additional compensation, as the worker and the employer generally agree on a monetary transaction in lieu of an actual reinstatement after the Court ruling. This compensation will be related to the protection provided to job-holders, that is, to the severance in case of unfair dismissals in that specific country.

[10] To give an example, T_U in Sweden is 38 because, as shown in Table A1, maximum notice (N) is 6 months, statutory severance for fair dismissals is 0, and the additional compensation for unfair dismissal is 32 months, while the likelihood of reinstatement is 0. Thus, 32+6=38

Table 1: Judicial discretion over severance pay

Country	T_U	T_E^F	T_D^F	$T_U - T_E^F$	$T_E^F - T_D^F$	Burden of Proof
Australia	13.90	3.80	1.00	10.10	2.80	Employer
Austria	20.29	4.00	4.00	16.29	0.00	Employer
Belgium	31.30	11.15	11.15	20.15	0.00	Employer
Canada(Federal)	-	4.3	2.00	-	2.30	Employer
CzechRepublic	19.99	3.50	2.00	16.49	1.50	Worker
Denmark	19.97	9.00	6.00	10.97	3.00	Worker
Finland	20.00	6.00	6.00	14.00	0.00	Employer
France	27.67	7.40	2.00	20.27	5.40	Worker
Germany	43.58	17.00	7.00	26.58	10.00	Employer
Greece	-	12.00	4.00	-	8.00	Employer
Hungary	27.16	9.00	3.00	18.16	6.00	Worker
Ireland	40.90	6.00	2.00	34.90	4.00	Employer
Italy	40.14	6.00	6.00	34.14	0.00	Employer
Japan	10.16	1.00	1.00	9.16	0.00	Both
Korea	17.81	1.00	1.00	16.81	0.00	Worker
Luxembourg	18.20	12.00	6.00	6.20	6.00	Employer
Mexico	-	-	-	-	-	Employer
Netherlands	16.67	4.00	4.00	12.67	0.00	Employer
NewZealand	12.49	0.50	0.50	11.99	0.00	Employer
Norway	29.61	6.00	6.00	23.61	0.00	Employer
Poland	11.82	6.00	3.00	5.82	3.00	Employer
Portugal	62.85	14.50	2.50	48.35	12.00	Employer
SlovakRepublic	27.79	7.00	3.00	20.79	4.00	Worker
Spain	36.50	12.50	0.50	24.00	12.00	Employer
Sweden	38.00	6.00	6.00	32.00	0.00	Employer
Switzerland	9.00	3.00	3.00	6.00	0.00	Worker
Turkey	32.00	22.00	2.00	10.00	20.00	Employer
UnitedKingdom	17.67	7.60	3.00	10.07	4.60	Employer
United States	-	0.00	0.00	-	0.00	Employer

Notes: All magnitudes are expressed in monthly wages.
Reference is made to a worker with 20 years of tenure.
T_U is compensation for unfair dismissal; T_E^F is compensation for fair economic dismissal, and T_D^F is compensation for fair disciplinary dismissal.
Sources: EPLex; OECD (2013);
See the main text and Table A1 in the Annex for details.

to the worker time to find alternative employment, see Table A3 in the web Annex), only two countries pay the same compensation at all tenure levels, notably Austria and Japan.

Why do regulations in so many countries allow for severance graded with tenure? Is this profile efficient from the standpoint of the individual worker and firm involved? In the model presented in Section 2, we provide necessary conditions under which a privately efficient and positive severance-tenure profile is socially efficient.

2 The model

One worker and one firm have a job opportunity that lasts n periods. The worker and the firm are risk neutral, and both discount the future at rate β, with $\beta < 1$. Workers are infinitely lived, with outside option given by the constant lifetime npv income of U, exogenous in this section. The corresponding per period income is $b = (1-\beta)U$. In period $\{i\}_{i=0}^{n-1}$ the worker faces a specific investment opportunity $s_i = \{0, 1\}$. The investment opportunity costs $\{C_i\}_{i=0}^{n-1}$ to the worker in each period. The investment is private information to the worker in each period.

Conditional on the investment being undertaken in period $i-1$, productivity in the following period will be ϵ_i, where ϵ_i is drawn from a continuous distribution $F(\epsilon_i)$, defined over the support $Z = [\underline{\epsilon}, \overline{\epsilon}]$. The distribution is time invariant. Note that there is no investment in period n. Further, productivity in period 0 is deterministic and fixed at $\tilde{\epsilon}_0$. Productivity is observed only by the firm, hence wages cannot be made contingent upon it. We assume that $\overline{\epsilon_i}$ is sufficiently large for the firm to break even, and for investment to be profitable in all periods (unless the worker is fired).

If the worker shirks in period i, this is observed by the firm in the following period, either directly or indirectly because his productivity is low. In this case, the firm will want to undertake a disciplinary dismissal. In the next section, where we focus on the burden of proof, we will be very specific about the productivity of the shirking worker. At this stage, we just assume that the firm initiates a disciplinary dismissal when it finds out about a non-investing worker. Note that the firm may also want to fire the worker, even if the worker has invested, if the draw of ϵ_i is sufficiently low. We refer to this case as an economic dismissal. The legislation applies different rules to these different types of dismissal.

Definition 1 *Disciplinary Dismissal. In period i, a firm is entitled to freely dismiss a shirking worker who did not invest in the period $i-1$.*

Definition 2 *Economic Dismissal. In period i, when productivity is sufficiently low, a firm is entitled to dismiss a worker by paying a severance T_i.*

Note that disciplinary dismissals, which we refer to as worker's shirking, need a third party intervention to certify the case. In the paper, we refer to the Court of law. From the match standpoint, the Court ruling is stochastic. We assume that there is a probability $1-q_i$ that the Court observes shirking and declares the firing as fair and pure disciplinary. In such a case, the firm is exempted from paying severance payments. Hence, there is a probability q_i that a shirking worker gets away with it and receives the severance payment. Since the realization of q_i is made after the firm has fired the worker, the expected severance payment for the firm when firing a shirking worker is $q_i T_i$. In this section, q_i is exogenous. It will be endogenized in section 3.[11] In what follows we assume that a firm always finds it in its interest to fire shirking workers, and then demonstrate that this is always the case when the severance is optimally set. In the case of an economic dismissal, we assume that the severance payment is always due, hence we abstract from moral hazard on the firm side.

Before we continue and derive the optimal contract, let us comment on our two driving assumptions regarding the contract space. First, a firm cannot commit to a severance payment, and hence that the severance payment T_i is a policy tool. Second, future wages can not be made contingent on future random productivity shocks.

[11] In terms of the definitions used in Section 1, we have that T_F^D=0 and T_F^E=T_i in this setting. T_U is introduced in section 3 below. Here we focus only on moral hazard of the employee.

We rationalize the first assumption, that the firm cannot contract upon T_i, by alluding to an underlying, not-modeled problem of adverse selection that stands on the way of a private contractual arrangement. If a firm unilaterally commits to a severance payment, it would be a victim of negative selection, and would end up hiring less favorable workers, in a mechanism similar to what was pointed out by Levine [34]. Let us be more specific. Suppose that there are two types of workers; ordinary workers as described above and shirkers, with $C = \infty$. Hence the shirkers always shirk. The fraction of the "shirkers" may be small, but strictly positive. Firms cannot distinguish between shirkers and ordinary workers. Consider a two periods situation where $n = 1$, and all firms offer a contract (w_0, w_1, T_1), where w_0 and w_1 are wages in two periods, respectively, and $T > 0$ is a privately imposed severance. We will argue that this cannot be an equilibrium. Consider a firm that deviates and offers a contract $(w_0, w_1', T_1 - \epsilon)$, where $w_1' > w_1$ and ϵ can be arbitrarily small. Since ordinary workers are strictly more willing to trade off severance payment for a higher period 1 wage than are the shirkers, it is possible to chose w_1' so that ordinary workers strictly prefer the new contract and shirkers strictly prefer the old contract. Hence the deviator only attracts the more profitable ordinary workers, and the equilibrium unravels. This argument can be used for any equilibrium candidate in which also ordinary workers receive severance pay. Thus, an arbitrarily small fraction of shirkers drives out severance pay for ordinary workers altogether. A formal treatment of this argument is provided in the Annex (see also [41] and [49]).[12] A mandatory severance solves this co-ordination problem. The realism of this assumption can be assessed considering that severance, wherever it exists, is either legislated or established within collective agreements at the industry, state or national level.

The second restriction, that future wages cannot be made contingent on the random component of output, is rationalized by alluding to asymmetric information. More specifically, we assume that the workers cannot easily observe the random component of productivity, while the firm can. Hence the firm can always pretend that it is low, in which case the worker will not accept a proposal to reduce wages. This is particularly the case if the value of a match depends not only on the workers output measured in volume, but also on a series of variables like the demand structure, input prices, and the efficiency of other parts of the corporation that it is difficult for a worker to obtain hard information about. It is also a matter of fact that wage contracts often are quite simple, and only specify an unconditional wage. Firms also seem reluctant to cut wages in hard times, as pointed out, inter alia, by Bewley [8].

2.1 Optimal contract

A wage contract $\Omega = \{w_i, \epsilon_i^d\}_{i=0}^{i=n}$ specifies sequences of wages and cut-off levels.[13] In each period, the firm will fire the worker if it is in its interest at that stage. We will refer to this as the firm's firing constraint.

The optimal contract $\Omega^\xi = \{w_i^\xi, \epsilon_i^\xi\}_{i=0}^{i=n}$ maximizes the firm's profit given the following constraints

1. worker's incentive compatibility constraint (ICC);

2. worker's participation constraint (PC) ;

3. firm's firing constraint (FC).

Below we first derive the optimal contract for a given seqence $\{T_i\}_{i=1}^{i=n}$ of severance payments, taken as given by the firm. Then we characterize the optimal cut-off levels. Finally we derive the optimal severance sequence $\{T_i^*\}$ as the severance payments that implement the optimal cut-off levels.

The incentive compatibility and participation constraints

At any point in time, the worker is free to end the relationship. Hence, in each period, the future value of the wage contract to the worker must be at least as large as the value of the outside option. This is the worker's participation constraint. The value of a job for an investing worker at period $i - 1$ is

[12] The argument still holds if the worker is risk averse. Risk aversion does imply that a cut in the severance will need a larger wage compensation in order to attract non-shirkers when workers are risk averse. However, this does not alter the argument, as the changes in the wage contract can be made arbitrarily small.

[13] In period 0 the worker is never fired. Hence we fix ϵ_0^d at some number below $\tilde{\epsilon}_0$.

$$W_{i-1} = w_{i-1} - C_{i-1} + \beta\left[(1 - F(\epsilon_i^d))W_i + F(\epsilon_i^d)(U + T_i)\right] \quad (1)$$

where w_{i-1} is the wage and C_{i-1} is the investment cost in period $i-1$ that will affect productivity in period i. W_i is the value of the job in the following period. Note that an investing worker is retained the following period with probability $(1 - F(\epsilon_i^d))$, where ϵ_i^d will be chosen optimally by the firm so as to maximize ex-post profits. With the complementary probability $F(\epsilon_i^d)$ the worker will get the outside option U augmented by the severance payment T_i. Suppose that the incentive constraints hold from period i onward. The value of employment for a shirker in period $i-1$ is $\tilde{W}_{i-1}$ and reads

$$\tilde{W}_{i-1} = w_{i-1} + \beta[q_i(U + T) + (1 - q_i)U] \quad (2)$$

The shirker does not invest C_i, and is fired in period i. Once fired, he will get the outside option U, and with probability q_i he will get also a severance payment T_i.

The incentive compatibility constraint implies that the wage w_{i-1} is determined so that $W_{i-1} \geq \tilde{W}_{i-1}$. In the web Annex we show the following lemma

Lemma 1 *Suppose that $T_i \leq T_i^*$ for all $i \geq 1$. Then the worker's incentive compatibility constraint binds in all periods*

Proof. See the Web Annex

A rough intuition goes as follows: suppose that the contract gives rents to the worker in period i. Suppose that the firm reduces the wage in period i down to the incentive compatibility constraint, and increases the wage in period $i-1$ so that the worker's expected income W_{i-1} (and hence also all earlier periods) stays constant. This lowers the firing threshold of the worker in period i and increases the joint income of the worker and the firm in that period, while the firing thresholds in all other periods are unaltered. Hence it increases the joint income of the match. Since the worker is on his participation constraint, this increase in joint income accrues to the firm. The ICC condition thus writes

$$-C_{i-1} + \beta\left[(1 - F(\epsilon_i^d))W_i + F(\epsilon_i^d)(U + T_i)\right] = \beta(U + q_i T_i)$$

After simple algebra, the incentive compatibility constraint can be written as follows:

$$(1 - F(\epsilon_i^d))(W_i - U) = \frac{C_{i-1} + T_i\beta[q_i - F(\epsilon_i^d)]}{\beta} \quad (3)$$

The rent $R_i = (1 - F(\epsilon_i^d))(W_i - U)$ has to be given to the worker in order to induce her to invest at time $i-1$. Let us give some comments on the optimal rent indicated by equation (3). First, note that if $\epsilon_i^d = \underline{\epsilon}_i$, i.e. if workers who invest are never dismissed, then $R_i = C_{i-1}/\beta + q_i T$. In this case the worker who invests is compensated for her outside option U, her investment cost C_i, and the rents $q_i T_i$ she would get if shirking. Second, the numerator in (3) increases in T_i if $q_i > F(\varepsilon_d)$. This stems from the fact that the worker in this case is more likely to get the severance if shirking than if not shirking. If $q_i < F(\epsilon_i^d)$, the opposite holds.

The wage w_0 is set so as to satisfy the participation constraint U at time zero:

$$W_0 = w_0 + \beta\left[(1 - F(\epsilon_1^d))W_1 + F(\epsilon_1^d)(U + T_1)\right] \geq U \quad (4)$$

For the optimal contract Ω^ξ, the participation constraint (PC) binds and uniquely determines w_0.

The optimal contract in period n

As there is no investment in period n and the employment relationship ends after the period, the worker's value function when employed is $W_n = w_n + \beta U$. Using the the ICC of equation (3) for investments in period $n-1$, and recalling that $b = (1-\beta)U$, yields

$$w_n = b + \frac{C_{n-1} + T_n\beta[q_n - F(\epsilon_n^d)]}{\beta[1 - F(\epsilon_n^d)]} \quad (5)$$

The firm profit is then

$$\Pi_n(\epsilon) = \epsilon - b - \frac{C_{n-1} + T_n\beta[q_n - F(\epsilon_n^d)]}{\beta[1 - F(\epsilon_n^d)]} \tag{6}$$

The optimal reservation rule solves $\Pi_n(\epsilon_n^\xi) = -T_i$. The optimal period-$n$ cut-off level ϵ_n^ξ is thus given by

$$\epsilon_n^\xi = b + \frac{C_{n-1} - (1 - q_n)\beta T_n}{\beta[1 - F(\epsilon_n^\xi)]} \tag{7}$$

Given ϵ_n^ξ, equation (5) determines w_n^ξ. The expected joint surplus of the worker and the firm, equal to the sum of wages and profit less b, is given by $S_n^e = \int_{\epsilon_n^\xi}^{\overline{\epsilon}} (\epsilon_n - b)dF$, where the expectation is conditioned on the relationship being active in period $n-1$.

The optimal contact, period $i < n$

We now proceed by backward induction. Consider period i, and suppose $w_{i+1}^\xi, ..., w_n^\xi$, $\epsilon_{i+1}^\xi, ..., \epsilon_n^\xi$, and $S_{i+1}^e, ..., S_n^e$ are solved for. Profits in period i for a firm that operates at productivity ϵ are given by

$$\Pi_i(\epsilon) = \epsilon - w_i + \beta \left\{ \int_{\underline{\epsilon}}^{\overline{\epsilon}} Max[\Pi_{i+1}(z), -T_i] dF(z) \right\} \tag{8}$$

The expected profit in period i, conditioned on the event that the match continues, is $\Pi_i^e = \int_{\epsilon_i^d}^{\overline{\epsilon}} (\Pi_(\epsilon_i) - w_i)dF$. The firm's continuation policy satisfies the reservation rule, and, at each i, we have that ϵ_i^d solves $\Pi_i(\epsilon_i^d) = -T_i$. Then

$$\Pi_i(\epsilon) = \epsilon - w_i + \beta \left[\Pi_{i+1}^e - F(\epsilon_{i+1}^d) T_{i+1} \right] \tag{9}$$

The surplus at time i is defined as $S_i(\epsilon) = W_i(\epsilon) - U + \Pi_i(\epsilon)$ The expected surplus in period $i+1$, conditioned on the worker being employed in period $S_{i+1}^e = \Pi_{i+1}^e + [W_{i+1} - U](1 - F(\epsilon_{i+1}^d))$.

Subtracting U on both sides of (1) and using that $b = (1-\beta)U$ gives that

$$W_i - U = w_i - b - C_i + \beta[(1 - F(\epsilon_i^d))(W_{i+1} - U) + F(\epsilon_i^d)T_{i+1}] \tag{10}$$

Inserting for (3) gives

$$w_i = b + \frac{C_{i-1} + T_i\beta[q_i - F(\epsilon_i^d)]}{\beta[1 - F(\epsilon_i^d)]} + C_i - \beta \left[(1 - F(\epsilon_{i+1}^d))(W_{i+1} - U) + F(\epsilon_{i+1}^d)T_{i+1} \right] \tag{11}$$

Inserting this into (9), and using the definition of joint surplus, gives that

$$\Pi_i(\epsilon_i) = \epsilon - b - \frac{C_{i-1} + T\beta[q_i - F(\epsilon_i^d)]}{\beta[1 - F(\epsilon_i^d)]} - C_i + \beta S_{i+1}^e \tag{12}$$

The optimal period-i-hiring threshold, ϵ_i^ξ, is given by the equation $\Pi_i(\epsilon_i^\xi) = -T_i$, or

$$\epsilon_i^\xi = b + \frac{C_{i-1} - T_i\beta(1 - q_i)}{\beta[1 - F(\epsilon_i^\xi)]} + C_i - \beta S_{i+1}^e \tag{13}$$

Given ϵ_i^ξ, (11) defines the optimal wage w_i^ξ, and we can use the definition of joint surplus to derive S_i^e. Finally, the entry wage w_0^ξ is determined by the worker's period-zero participation constraint (4). Hence we have derived the optimal contract Ω^ξ. For later reference we denote the period-zero npv profit of the firm by $\Pi_0^\xi(U)$.

Note that both the left-hand side and the right-hand side of (13) are increasing in ϵ_i^d, hence the equation may not have a solution.[14] If investments are sufficiently productive, the equation has a solution, and we

[14] Or it may have multiple solutions, in which case the lowest solution is the relevant one since the firm chooses the lowest possible incentive compatible wage.

say that the investment is *implementable* in that period. To be more precise, suppose that the distribution can be written as $\epsilon = kz$, where z is a stochastic variable on $[a, b]$ with median value of $z^m > 0$ and expected value of $\bar{z} > 0$. The scalar k is a measure of the productivity of the investment.

Lemma 2 *Consider arbitrary sequences of investment cost $\{C_i\}_{i=0}^{n-1}$ and severance $\{T_i\}_{i=1}^{n}$ such that $T_i < T^*$, where T^* is defined below. If the productivity of the investments are sufficiently high (k is sufficiently high), the investments are implementable in all periods, and ϵ_i^ξ is increasing in the investment cost C_i in all periods.*

Proof. See the Annex

2.2 Efficient separation and optimal severance pay

Efficient separation in the last period n is defined simply as a productivity ϵ_n^* that ensures zero surplus, so that $S_n(\epsilon_n^*) = 0$. Given, ϵ_n^*, we can define the expected efficient surplus in the last period, S_n^{*e}. The surplus in the last period is simply $\epsilon_n - b$.[15] We thus have that

$$\begin{aligned} \epsilon_n^* &= b \\ S_n^{*e} &= \int_{\epsilon_n^*}^{\bar{\epsilon}} (z - \epsilon_n^*) dF(z) \end{aligned} \tag{14}$$

For earlier periods, efficient firing and efficient surplus conditional on future efficient firing can be defined recursively. Given S_{i+1}^{*e}, the joint surplus in period i as a function of ϵ_i, reads[16] $S_i^*(\epsilon) = \epsilon_i - b - C_i + \beta S_{i+1}^{e*}$ Optimal firing in period i (defined as $S_i^*(\epsilon_i^*) = 0$) and S_i^{*e} requires then

$$\epsilon_i^* = b + C_i - \beta S_{i+1}^{e*} \tag{15}$$

$$S_i^{*e} = \int_{\epsilon_i^*}^{\bar{\epsilon}} (z - b - C_i + \beta S_{i+1}^{e*}) dF(z) \tag{16}$$

Note that, as we already pointed out, neither wages nor severance payments appear in the joint surplus, as they are transfers between the two parties. Equation (15) suggests that firing is efficient whenever the productivity from the job (ϵ_i^*) falls below the worker's outside option $b = (1 - \beta)U$ augmented by the investment cost C_i, and the continuation surplus S_{i+1}^{*e} .

Definition 3 *The severance payment sequence $\{T_i^*\}_{i=1}^{n}$ is efficient if it is such that $\{\epsilon_i^\xi = \epsilon_i^*\}_{i=1}^{n}$.*

Hence efficient severance payments is obtained when the reservation productivity $\epsilon_i^\xi(T_i^*)$ is identical to its efficient counterpart level. Let us first consider period n. From equation (7) and (14)

$$\epsilon_n^\xi - \epsilon_n^* = \frac{C_{n-1} - (1 - q_n)\beta T_n}{\beta[1 - F(\epsilon_n^\xi)]} \tag{17}$$

The equation immediately tells us that at $T_n = 0$, $\epsilon_n^* < \epsilon_n^\xi$ so that there is too much firing when severance payments are zero. As wages need to pay for the worker's investment effort in earlier periods, the firm has a tendency to over-dismiss a worker who invested in the previous period.

To derive the optimal severance, we again do this recursively. Consider any period i and suppose that $T_j = T_j^*$ for all $j > i$ (if any), so that $S_j^e = S_j^{*e}$ for all $j > i$. From equation (13) and (15), it follows that efficient firing is obtained if and only if

[15]To see this, recall that $S_n = W_n + \Pi_n - U = \epsilon_n + (1 - \beta)U = \epsilon_n - b$

[16]Note that $S_i = W_i + \Pi_i - U = \epsilon_i - b + \beta E(W_{i+1} + \Pi_{i+1} - U) = \epsilon_i - C_i - b + \beta S_{i+1}^e$

$$\frac{C_{i-1} - T_i\beta(1-q_i)}{\beta[1-F(\epsilon_i^{\xi})]} = 0 \tag{18}$$

Or

$$T_i^* = \frac{C_{i-1}}{\beta(1-q_i)} \quad \forall i = 1, ...n \tag{19}$$

To understand this simple formula, first note the following: the firm defers wages in order to incentivize the worker to invest, and hence tends to fire workers too often *ex post*. The optimal severance neutralizes this effect. The higher is the cost of investments in a given period, or the easier it is to get away with shirking, the more wages must be deferred, and the higher must the severance be in order to achieve optimal firing.

To be more detailed, recall that the origin of the problem is that the worker's inside option W_i^{ξ} in any period i is higher than his outside option U. The optimal severance is equal to this difference, $T_i^* = W_i^{\xi} - U$. In other words, with $T_i = T_i^*$, the firm's firing decision has no welfare effects on the worker, and the firing decision is jointly optimal. Furthermore, at T_i^*, the expected income in period i if investing is T_i^* (since the worker is indifferent between being retained and laid off), and if not investing $q_i T_i^*$, hence the difference is $(1-q_i)T_i^*$. This should exactly cover the cost of investing in the previous period, which, when discounting is taken into account, is C_{i-1}/β. Thus, it follows that $(1-q_i)T_i^* = C_{i-1}/\beta$. Solving for T_i^* gives (19).

The expression for the optimal severance turns out to be surprisingly robust. It holds with endogenous Court decisions (endogenous q_i) as shown in the next two sections. In addition, it holds in general equilibrium, as we show in section 4. If $q_i = 1$, the severance does not influence the firm's hiring decision, and is then useless as a policy tool for inducing optimal retention by the firm.

Proposition 1 *For $q_i < 1$, the optimal severance T_i^* is given by* (19)*. It is increasing in the investment cost in the previous period, and in the probability of getting away with it if shirking. It does not depend on investment costs, or the probability of being caught in any other periods.*

It follows that the severance is increasing with tenure if q_i is increasing with tenure or if C_i is increasing with tenure. Both seems reasonable, as discussed in Section 5.

Corollary 2 Necessary conditions for Upward Sloping Severance Payments
The optimal severance payment is strictly increasing with tenure if either q_i or C_i is strictly increasing in tenure and neither is strictly decreasing with tenure.

Proof The proof is a direct consequence of equation (19)

We want to point out the remarkable fact that optimal severance is independent of the distribution of ε_i. Optimal severance pay only depends on q_i, a property of the legal system, and C_{i-1}, the investment costs. It seems natural to assume that q_i is the same for all the firms in a country. The investment cost C_i is probably firm-specific. However, one may think that the average value of C_i may vary from country to country. Hence our theory predicts that countries with a high value of q_i (inefficient judicial system), and where workers tend to have high investment costs, the optimal severance pay is high. We return to this prediction in Section 5.

We also want to study wage profiles. Inserting that $W_i - U = T_i$ in (11) and substituting out T_i and T_{i+1} by the virtue of (19) gives that

$$w_i = b + \frac{C_{i-1}}{\beta}\frac{1}{1-q_i} - C_t\frac{q_{i+1}}{1-q_{i+1}} \tag{20}$$

Let us look at one example. Suppose then that the cost in period C_i is $C_0 + i\Delta$, keeping q_i constant and $\beta = 1$. We start at $i = 0$ for convenience. It follows that the wage is given by

$$w_i = \begin{cases} b - C_0\frac{q}{1-q} & \text{if} \quad i = 0 \\ b + C_0 + (i-1)\Delta - \frac{q}{1-q}\Delta & \text{if} \quad 0 < i < n \\ (C_0 + (n-1)\Delta)\frac{1}{1-q} & \text{if} \quad i = n \end{cases}$$

Hence wages are increasing over tenure by the same amount as the increase in per period investment costs. In the last period, the worker gets a large bonus, and this drags down wages in all earlier periods. This discussion suggests that the wage tenure profile is steeper when the investment cost and the severance pay grows more quickly.

Before we continue, we want to make a point regarding fiscal externalities. Although the optimal severance maximizes the joint income of the worker and the firm, there may still be fiscal externalities associated with firing, as unemployed workers receive unemployment benefits while employed workers pay taxes and firms payroll taxes. This fiscal externality implies that the private value of unemployment exceeds the social value, while the opposite is true for employment. One may think that this is an argument for increasing the severance pay even further. However, this needs not work. When it is privately optimal for the worker and the firm to separate, the optimal contract will prescribe a wage that is higher than w_i^{ξ} derived above, so that the (privately) efficient separation rates will be realized. In this case, the ICC constraint of the worker will not bind in periods $1, ..., n$, and the worker will compensate the firm for the slack through a lower wage in period 0. In this case, a more direct and effective policy tool will be a firing tax, paid by the firm to the Government, reflecting the fiscal externalities associated with firing the worker.

3 Burden of proof and endogenous Court rulings

In this section we dig deeper into the legal system and obtain endogenously the probability that a shirking worker gets away with it. In addition, we define under which circumstances the firm may get away with it. We do this by analyzing separately two different cases, in which the burden of the proof is on the worker and on the firm, respectively. The burden of proof is defined as Court presumption regarding the nature of the dismissal, economic versus disciplinary. We abstract from any preference biases by the Court. Hence, if the burden of proof is one one of the parties, this will always be honored by the Court. The party that can get away with it is the party that has not the burden of proof. To keep things simple, we focus on a two-periods model, so that the relevant workers value functions are W_0 and W_1. Results can be readily generalized to a generic n-periods setting as well as to general equilibrium.

We assume that shirking workers draw a productivity level ϵ in the first period, with ϵ continuously distributed on $[\alpha, \gamma]$ with cumulative distribution function $\tilde{F}(\epsilon)$. An investment in period 0 on the part of the worker shifts the distribution of productivity up by Δ, which is common knowledge. Hence the support of the distribution of a worker who is investing is $[\alpha + \Delta, \gamma + \Delta]$, and the cumulative distribution function can be written as $F(\epsilon) = \tilde{F}(\epsilon - \Delta)$. To make the problem interesting, we assume that the support of the two distribution has an area of overlap (Figure 1), so that $\Delta < \gamma - \alpha$. We assume that the investment cost C_0 is sufficiently small relative to Δ so that the firm always wants to induce the worker to invest.

Whether a firing is economic or disciplinary is settled by a Court ruling. We assume that the Court can observe the realized productivity ϵ, but can not directly observe whether the worker did invest or not in period 0.[17] We return to the issue of Court monitoring in section 3.4.

We assume that if the employer has the burden of proof, severance has to be paid to a worker who is fired as long as the observed productivity is within the support of the productivity of a worker who has invested. It follows that a worker who invests always gets severance if fired, while a shirker gets away with it unless her productivity draw is below $\alpha + \Delta$. The probability that the worker gets away with it is thus $q_1^e = 1 - \tilde{F}(\alpha + \Delta)$. The superscript e indicates that the burden of proof is on the employer. Conversely, if the burden of proof is on the worker, the firm does not have to pay severance as long as the observed productivity is within the support of a shirker. In this case a shirker never gets severance, while the employer in some circumstances gets away with it and hence carries out an *unfair economic dismissal*. The probability that

[17] We retain the assumption that the wage cannot be made contingent on productivity. If it were, the Courts would have to intervene in each period in order to ensure that the wage contract was honored, and this may be prohibitively costly. In principle, when the Court can observe a noisy signal of the worker's productivity, the employer may offer a wage contract that depends on this noisy signal. However, even in this case there is a difference between letting the Court decide on wages and on severance pay in the case of firing, as wages are set every period, while firing takes place less frequently. Going to Court every period to receive the signal may be prohibitively costly. That being said, the threat from the worker of going to Court may discipline the firm, and makes it willing to offer higher wages when output is high. Analyzing this more complicated game is on our agenda for future research.

the firm gets away with it is $q_1^w = F(\gamma) = \tilde{F}(\gamma - \Delta)$, where superscript w indicates that the burden of proof is on the worker. It follows that q_1^e and q_1^w only depend on the productivity shifter Δ and the distribution of ϵ, while other aspects of the legal system do not matter.

Total surplus in period 1 is $S_1(\epsilon) = \epsilon - b$. Efficient separation requires $\epsilon_1^* = b$. To simplify the exposition, we shall operate only around the efficient solution, so that the reservation productivity set by the firm will always be $\epsilon^d(T_1^*) = b$. Hence

$$T_1^* = w_1(T_1^*) - b \tag{21}$$

where w_1 is the wage in period 1 determined by the ICC. Recall that the worker's rent in period 1 is $W_1 - U = w_1 - b$. Equation (21) thus implies that the first-period rent is equal to the severance payment- $W_1 - U = T_1$, a result that we already know from the previous section.

We will consider first the case where the burden of proof is on the employer, and subsequently model a case where the burden of proof is on the employee.[18]

3.1 Burden of proof on the employer

When the burden of proof is on the employer, the Court presumes that the dismissal is economic. An investing worker always gets either wages or severance in period 1. Using the fact that the worker rent in period 1 is $W_1 - U = T_1^*$, W_0 writes

$$W_0 = w_0 - C_0 + \beta[U + T_1^*]$$

A shirking worker will get severance if her productivity is above $\alpha + \Delta$, and may even be hired if $\gamma > b$. A worker is equally well off being retained and being fired with severance, the income in both cases is $U + T$. Hence, a shirking worker gets

$$\tilde{W}_0 = w_0 + \beta\left[U + q_1^e T_1^*\right]$$

The incentive compatibility constraint ($W_0 = \tilde{W}_0$) then implies that

$$T_1^* = \frac{C_0}{\beta(1 - q_1^e)} \tag{22}$$

The expression is identical to the expression for optimal severance derived earlier, given by equation (19), with q_1 replaced by q_1^e.

To complete the model, and for comparison with the specification in the following section, the period zero wage satisfies the participation constraint and solves $W_0 \geq U$ so that

$$w_0 = b + C_0 - \beta T_1^* \tag{23}$$

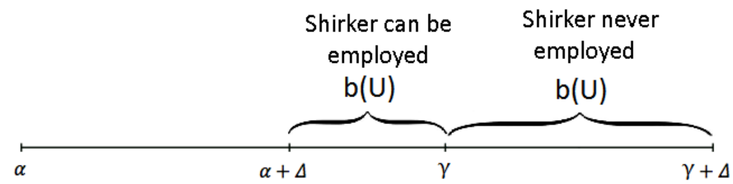
where it is clear that the severance payment is prepaid by the worker in period 0.

3.2 Burden of proof on the worker

When the burden of proof is on the worker, the Court presumes that the dismissal is disciplinary. We distinguish between two cases, based on whether or not it may be efficient to retain a shirker. The same two cases depicted in Figure 1 apply in this context.

[18]The case where $\alpha < b < \alpha + \Delta$ is not interesting. In such an environment, an investing worker is never fired, hence severance is never paid in equilibrium

Figure 1: Outside option of the worker and productivity: 2 cases



Never efficient to retain shirker: $\gamma < b$

An investing worker will be employed if her productivity exceeds the period 1 wage. As explained above, an investing worker is severed if her productivity is below the wage, but above γ, so that she is distinguishable from a shirker. The severance, if paid, is equal to $w_1 - b$. The value functions for a shirker and a non-shirker then read

$$\begin{aligned} \tilde{W}_0 &= w_0 + \beta U \\ W_0 &= w_0 - C_0 + \beta[U + (1 - q_1^w)T_1^*] \end{aligned}$$

Setting $\tilde{W}_0 = W_0$ gives

$$T_1^* = \frac{C_0}{\beta(1 - q_1^w)} \tag{24}$$

Note the similarity with the optimal severance when the burden of proof is on the firm, given by (22) (and also with exogenous q, given by (19)). The only difference is that the q_1^e now is replaced by q_1^w, the probability that the firm (not the worker) gets away with it. To give intuition for the similarities, note that the fact that the investing worker does not get severance when he is indistinguishable from the shirking worker reduces the pay-off of investing with $q_1^w T_1^*$ vis-à-vis the payoff when investing when the burden of proof is on the employer. However, the fact that the shirking worker never gets away with it reduces the value of shirking with $q_1^e T_1^*$ vis-à-vis the payoff from shirking when the burden of proof is on the employer. Note further that if we add the assumption that the distribution of ϵ is symmetric around its expected value (for instance being uniformly distributed), it follows that $q_1^e = q_1^w$. In this case the two effects cancel out, and the optimal severance is the same in the two cases.[19]

The participation constraint reads

$$\begin{aligned} W_0 &= w_0 - C_0 + \beta[U + (1 - q_1^w)T_1^*] \\ &= U \end{aligned}$$

so that the first period wage is

$$w_0 = b + C_0 - \beta T_1^* + q_1^w T_1^* \tag{25}$$

By comparing (23) and (25) it follows that the wage profile is flatter when the burden of proof is on the worker rather than the employer. The intuition is straight-forward: when the burden of proof is on the worker, the worker then obtains a lower expected period-1 profit and is compensated by a higher period-0 wage. Note also that the firing is still efficient. This follows from the assumption that b is outside the support of the productivity of the shirker. Hence, around the efficient firing threshold, the firm has to pay severance, and therefore makes the efficient decision. Investing workers with a lower productivity may not receive severance, but these workers would be fired anyway.

[19]Recall that $q_1^e = 1 - \tilde{F}(\alpha + \Delta)$, while $q_1^w = \tilde{F}(\gamma - \Delta)$. With symmetry, the two are equal. Note that if ϵ is uniformly distributed, it follows that $q_1^e = q_1^w = 1 - \frac{\Delta}{\gamma - \alpha}$.

May be efficient to retain shirker: $b < \gamma$

In this case, an employer can fire a worker with productivity b and get away with it. Since the ICC requires that $w_1^* > b$, it follows that it is impossible to use severance to induce the firm to retain a worker with productivity in the interval $[b, \min[w_1^*, \gamma]]$.

Note that if $\gamma < w_1^*$, shirking workers will never be retained nor severed, while investing workers will be retained or severed and receive a total payoff equal to $U + T_1^*$ if and only if $\epsilon \geq \gamma$. The probability of this event is $1 - F(\gamma) = 1 - q_1^w$. The gain from investing is thus $T_1^*(1 - q_1^w)$, and from the non-shirking condition it follows that

$$T_1^* = \frac{C_0}{\beta(1 - q_1^w)} \tag{26}$$

which is identical to (24).

If, on the contrary, $\gamma > w_1^*$, shirking workers are hired with positive probability, and severance is never paid. The difference in the probability of receiving a wage offer between the shirker and the non-shirker is $1 - F(w_1^*) - (1 - \tilde{F}(w_1^*)) = \tilde{F}(w_1^*) - \tilde{F}(w_1^* - \Delta)$, since $F(w_1^*) = \tilde{F}(w_1^* - \Delta)$. Thus, the gain from investing is $(w_1^* - b)[\tilde{F}(w_1^*) - \tilde{F}(w_1^* - \Delta)]$. Hence, the ICC wage writes

$$w_1^* = b + \frac{C_0}{\beta[\tilde{F}(w_1^*) - \tilde{F}(w_1^* - \Delta)]} \tag{27}$$

Note that if ϵ is uniformly distributed, $\tilde{F}(w_1^*) - \tilde{F}(w_1^* - \Delta) = \Delta/(\gamma - \alpha) = 1 - q_1^w = 1 - q_1^e \equiv 1 - q_1$,[20] in which case

$$w_1^* = b + \frac{C_0}{\beta(1 - q_1)} \tag{28}$$

Hence the period 1 rent to the worker, $\frac{C_0}{\beta(1-q_1)}$, is the same as in the rest of the section.

With $b < \gamma$, the Government may use other instruments to achieve efficient separations. By imposing a firing tax equal to the difference between the period 1 wage and the outside option, efficient firing is restored. For instance, if $\gamma > w_1^*$, a firing tax of $w_1^* - b$ will restore efficiency along the firing margin. However, a firing tax will, in contrast with severance, reduce the profitability of the firm, as the firing tax cannot be offset by a lower wage in the first period. Hence, a firing tax will lead to inefficient entry.

3.3 Double moral hazard and unfair economic dismissals

When the burden of proof is on the worker, the firm has an incentive to declare that a productivity level below $b = (1 - \beta)U$ is always the results of a non-investment on the part of the worker. In this subsection we relax the assumption, and allow the Court to monitor the firm with some probability.

We consider the case in which $b > \gamma$ so that shirking workers are never employed. We are in the efficient solution of equation (24), and the firm wants to fire the worker, so that the productivity is below b. After dismissal, the Court audits the firm with probability λ, and finds out whether the dismissal is economic or disciplinary. If the firm claims it is disciplinary, while the Court finds that it is economic, a severance of T_{U1}^* for *unfair economic dismissal* is due. Hence the expected profit of the firm is $\tilde{\Pi} = -\lambda T_{U1}^*$. Conversely, truth-telling gives the firm a profit of $\Pi = T_1^*$. The incentive compatibility constraint for the firm requires that $\Pi \geq \tilde{\Pi}$. Assuming that the constraint binds, we have that

$$T_{U1}^* = \frac{T_1^*}{\lambda} \tag{29}$$

This means that severance payments for unfair economic dismissals should be higher than severance payments for economic reasons, a property that holds in all countries (see Table 1). Furthermore, we expect T_{U1}^* and $T_1^* = T_E^F$ to be positively correlated, another property which is in the data (the cross-country correlation of

[20] To see this, note that with uniform distribution, $\tilde{F}(w_1^*) - \tilde{F}(w_1^* - \Delta) = \Delta/(\gamma - \alpha)$.

T_U and T_E^F is .6 which is statistically significant at conventional levels). More interestingly, combining the above condition with the no-shirking (and participation) condition for the worker, we have that

$$T_{U1}^{*} = \frac{C_0}{\beta\lambda(1-q_1^e)} \tag{30}$$

which establishes a relationship between optimal severance and efficiency of the judicial system, here defined more precisely in terms of the audit technology. A higher "λ" implies a more efficient legal system and a lower severance payment.

Proposition 3 *In a more efficient legal system, severance payments for unfair dismissals are lower.*

4 Severance pay in general equilibrium

In this section, we close the model by embedding the partial equilibrium contracting problem into a Diamond-Mortensen-Pissarides framework with directed search. Next, we calibrate severance pay and numerically solve the model, looking at the welfare and unemployment effects of suboptimal severance.

We follow the competitive search approach à-la-Moen, and assume that firms post period 0 rents to maximize the value of a vacancy. The matching function is a standard Cobb-Douglas with constant returns to scale. There is a cost K of opening a vacancy. The vacancy operates until it is filled. However, if a worker and an employee separate, the vacancy is lost. We shall indicate with θ the aggregate vacancy unemployment ratio and with $\chi(\theta)$ the probability of finding a worker. Firms advertise wage contracts, and workers calculate the expected discounted income W_0 that the contract yields.[21] The value of a firm with a vacancy can be written as

$$V = -c + \beta\left[\chi(\theta)\Pi + (1-\chi(\theta))V\right] \tag{31}$$

where c is the flow cost of maintaining a vacancy and Π is the NPV profit of hiring a worker. Let z denote the pure flow utility value of unemployment and $\theta\chi(\theta)$ be the worker's probability of finding a new job. The net present income of an unemployed worker reads

$$U = z + \beta\left[\theta\chi(\theta)W_o(\Omega) + (1-\theta\chi(\theta))U)\right] \tag{32}$$

As in Moen [40], the higher is W_0, the quicker is the vacancy filled. Since there is also moral hazard in our model, we follow Moen and Rosen [42], and separate the optimal contracting problem in two parts. First, for a given npv value of the contract W_0, we characterize the optimal contract and the associated value of $\Pi_0(\Omega)$. Second, we chose the level of W_0 that maximizes the value of the vacancy V defined by (31), given the constraint (32).

For any given npv value W_0, the contract Ω prescribes to the worker, lemma 1 still holds, and it follows that for all periods but the first one, the solution to the first step of the maximization problem is the optimal contract derived in partial equilibrium, as prescribed in lemma 2. The period-0 wage does not influence any retention decisions, and this wage is adjusted so as to give the worker a total NPV compensation of W_0.

Lemma 3 *Suppose the firm offers a contract that gives the worker an expected income $W_0 > U$. Then the optimal contract $\Omega^{W_0}(U)$ is identical to the optimal contract $\Omega^{\xi}(U)$ in all periods but period 0, while $w_0^{W_0} = w_0^{\xi} + W_0 - U$. The associated firm profit is given by $\Pi = \Pi_0^{\xi}(U) - (W_0 - U)$, where $\Pi_0^{\xi}(U)$ is the profit associated with $\Omega^{\xi}(U)$.*

The intuition of Lemma 3 is that the firm can always scale up or down W_0 by increasing or decreasing w_0^{ξ}, as this will not influence the incentive compatibility constraint in any period.

[21] Note that as in partial equilibrium, a worker in period 0 only has preferences over the NPV income the contract prescribes, not the time profile.

Definition 4 General Equilibrium
General equilibrium is a wage contract Ω^g, a npv income U^g for unemployed workers, a labor market tightness θ^g, and an npv income W_0^g for a worker becoming employed such that

1. $\Omega^g = \Omega^\xi(U^g)$ *for all periods but period* 0 *and* $w_0^g = w_0^\xi + W_0^g - U^g$.
2. W_0^g *maximizes* V *given by* (31) *subject to* (32) *for* $U = U^g$.
3. *Zero profits,* $V = K$.

In the web annex we show that the equilibrium satisfies

$$K = \frac{-c + \beta\chi(\theta)(1-\eta(\theta))(\Pi_0^\xi(U^g) - K)}{1-\beta} \quad (33)$$

$$U^g = \frac{z + \beta\theta\chi(\theta)\eta(\theta)(\Pi_0^\xi(U^g) - K)}{1-\beta} \quad (34)$$

where $\eta(\theta)$ is the elasticity of the matching function χ with respect to θ. When solving the model, we first solve for the optimal contract and $\Pi^\xi(U)$. Then (33) and (34) together solve for U^g and θ^g. The wage W_0 can then be backed out from (32).

Furthermore, we know from section 2 that the separation decision within each firm is efficient, and that search frictions do not create inefficiencies in competitive search equilibrium.[22] The following proposition thus follows:

Proposition 4 Efficient General Equilibrium
Suppose that the sequence of severance pay is given by a set of T_i *that satisfies equation* (19). *Suppose further that* z *reflects both the private and the social flow value of being unemployed. Then the general equilibrium allocation is efficient.*

Proof: This is just an application of Moen [40] with endogenous destruction.

In order to find the unemployment rate, we have to characterize the separation rate. This is conceptually simple but somewhat tedious. Let N_t be the stock of workers employed in firms with tenure t. Then the separation rate in that period is $F(\epsilon_t^*)$, where ϵ_t^* is given by equation (15). Furthermore, $N_t = (1 - F(\epsilon_{t-1}^*))N_{t-1}$. Hence, the average separation rate $\bar{s}$ in steady state reads

$$\bar{s} = \sum_{t=1}^{n+1} F(\epsilon_t^*)\Pi_{j=1}^t(1 - F(\epsilon_j)) \quad (35)$$

with $F(\epsilon_{n+1}^*) \equiv 1$. The unemployment rate is then $u = \bar{s}/(\bar{s} + p)$.

Numerical simulations

This section performs a set of simulations of the general equilibrium of the model, and provides a quantitative sense to the results of the previous section. We perform three exercises. First, we solve numerically the general equilibrium with optimal severance payments. To get some discipline on our choice of parameters, we assume that the observed severance levels are indeed optimal. We thus calibrate values for the optimal severance payments using the evidence provided in Table 1, and provide values for a "high" and "medium" severance pay economies. The high severance pay economies are coherent with Turkey and Portugal, while the medium economy has a severance pay level coherent with that of Germany, see Table 1. Finally, we ask what are the effects of deviating from the optimal level of severance pay in the two economies. The

[22] We have not given general conditions for existence of equilibrium, as we consider this outside the scope of this paper. One can show that the equilibrium exists and is unique if the expected value of ϵ is sufficiently high, if $\tilde{\epsilon}_0 < b$, and if the sequence C_i is non-increasing or not increasing too rapidly. Note also that existence for specific parameters are implicitly shown through the simulations.

simulations show that with reasonable parameters the welfare loss associated to the elimination of severance varies between 1 percent in the "medium" severance pay economy to a 3 percent loss in the "high" severance pay economy.

The time period is one year and the discount rate is 0.98. The time period of the labor contract is 20 years, so as to be coherent with the severance pay values available in Section 1, and in the web Annex. The rest of the labor market values are reported in Table 2. The most important parameters are the sequence $\{C_i\}$ and $\{q_i\}$. The logic of the calibration of $\{T_i\}$ is the following. In the top part of Table 3 we report values for the "high" severance pay economy in terms of monthly wages, in a way coherent with the evidence provided. With respect to the severance pay reported in Table 1, we aim at capturing the size of T_E^F, i.e. the severance pay for a fair economic dismissal. Given the productivity distribution specified in Table 2, we convert the value of the severance pay in terms of average productivity of our economy.[23] These average productivity levels are reported in the column labeled *Data* in Table 3. Converting severance pay from wages to productivity one-to one is a simplification, since we do not take into account that a share of the output is allocated to the firm, to cover the entry cost.[24] Note that in the column *Data* severance payments increase with tenure.

The logic of the choice of $\{C_i\}$ and $\{q_i\}$ to back out $\{T_i\}$ and matching the data is the following. We work with a time invariant value of C_i, so as to approximately match the average level of the severance payments. We then use the sequence q_i to match their slope. While our calibration is not perfect, Table 3 shows that we obtain reasonable values for both the level and the steepness of severance pay. The overall fitting of the ''high" severance pay economy is better than the corresponding fitting of the "medium" economy. Figure 2 plots the optimal severance payment for the "high" economy as well as the reservation productivity, with respect to tenure. The severance payments has an initial spike in the early years, and thereafter grows smoothly with tenure, driven by the sequence of $\{q_i\}$. From Figure 2, also the reservation productivity increases with tenure. This is the well known labor hoarding effect, and reflects the option value of delaying dismissal typical of dynamic models of labor demand.

The policy experiment is reported in Table 4. Starting from the parameters specified in Table 2, we change only the level of severance payments. Specifically, the severance payments moves from the value reported in Table 3 to a value of zero.[25] This is reported in the row labelled SP at zero in Table 4. In the corresponding columns we report the equilibrium unemployment rate and market tightness. In addition, the table reports the equilibrium value of unemployment U, a natural measure of welfare in our competitive search economy. The exercise shows that when severance payments are taken to zero the equilibrium unemployment rises and market tightness falls. The key question is the effect on welfare of removing severance payments. The table shows that the welfare loss is approximately 1 percent in the case of the "medium" severance economy and close to 3 percent in the "high" severance economy.

Table 2: Baseline Parameters in two Economies

Parameter	Notation	High T_i	Medium T_i
Pure Discount Rate	β	0.980	0.980
Stochastic productivity: mean	$\overline{\epsilon}$	0.400	0.400
Stochastic productivity: variance.	$\sigma\{\epsilon\}$	0.820	0.650
unemployed income	z	0.160	0.100
matching function elasticity	α	0.400	0.400
matching function parameter	A	1.000	1.000
search cost parameter	c	0.100	0.100
entry cost	k	0.900	1.770

Source: Authors' calculation

[23]The average productivity level is $\overline{y} = \frac{\epsilon_0+\sum_{i=1}^{n}\Pi_{j=1}^{i-1}(1-F(\epsilon_j^d))\int_{\epsilon_i^d}^{\overline{\epsilon}} zf(z)}{1+\sum_{i=1}^{n}\Pi_{j=1}^{i}(1-F(\epsilon_j^d))}$

[24]Consider, for example, the 5 monthly wage at five years in the 'high' severance economy reported in Table 3. Since the average yearly effective productivity in the model is 0.6, 5 months of productivity correspond to the value of 0.25 reported in Table 3

[25]Note that we do not consider a policy experiment that increases the level of severance payment beyond the optimal level, since the firm can undo the effect of excess severance by oversatisfying the ICC in the period of the excess severance, simply increase the wage in that period and reducing it in earlier periods.

Table 3: Calibrating Severance Payments

High Severance Pay Economy			
	Data		Model
Tenure/Years	SP in Monthly Wage	SP converted in Monthly Productivity [a]	SP converted in Monthly Productivity [b]
5 years	5	0.251	0.221
10 years	10	0.503	0.296
20 years	20	1.006	0.918
Training Technology	Notation		Model
Training Costs:	C_0		0.180
Prob. of Getting away with it: min	q_0		0.000
Prob. of Getting away with it: max	q_n		0.800

Medium Severance Pay Economy			
	Data		Model
Tenure/Years	SP in Monthly Wage	SP converted in Monthly Productivity [a]	SP converted in Monthly Productivity [b]
5 years	2	0.077	0.134
10 years	5	0.193	0.203
20 years	10	0.386	0.267
Training Technology	Notation		Model
Training Costs:	C_0		0.131
Prob. of Getting away with it: min	q_0		0.200
Prob. of Getting away with it: max	q_n		0.500

[a] Monthly productivity in the model corresponding to $\{T_i\}$ in terms of monthly wage in the data.
[b] Optimal severance pay in the model in terms of monthy productivity

Table 4: Effects of Suboptimal Severance Payments

Sev. Payments	Welfare U	Unemp. u	Tightness θ	Welfare Loss [a]
Economy with High SP:				
SP at optimal level:	10.238	0.066	0.558	
SP at zero:	9.958	0.071	0.488	-0.027
Economy with Medium SP:				
Sp at optimal level:	7.413	0.069	0.524	
SP at zero:	7.338	0.070	0.508	-0.010

[a] Percentage changes in unemployment value U from reducing severance pay to zero.

Figure 2: Severance Payments by tenure in the High Severance Payment Economy

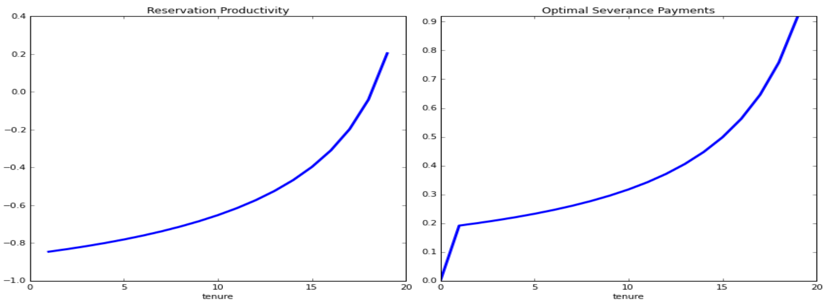
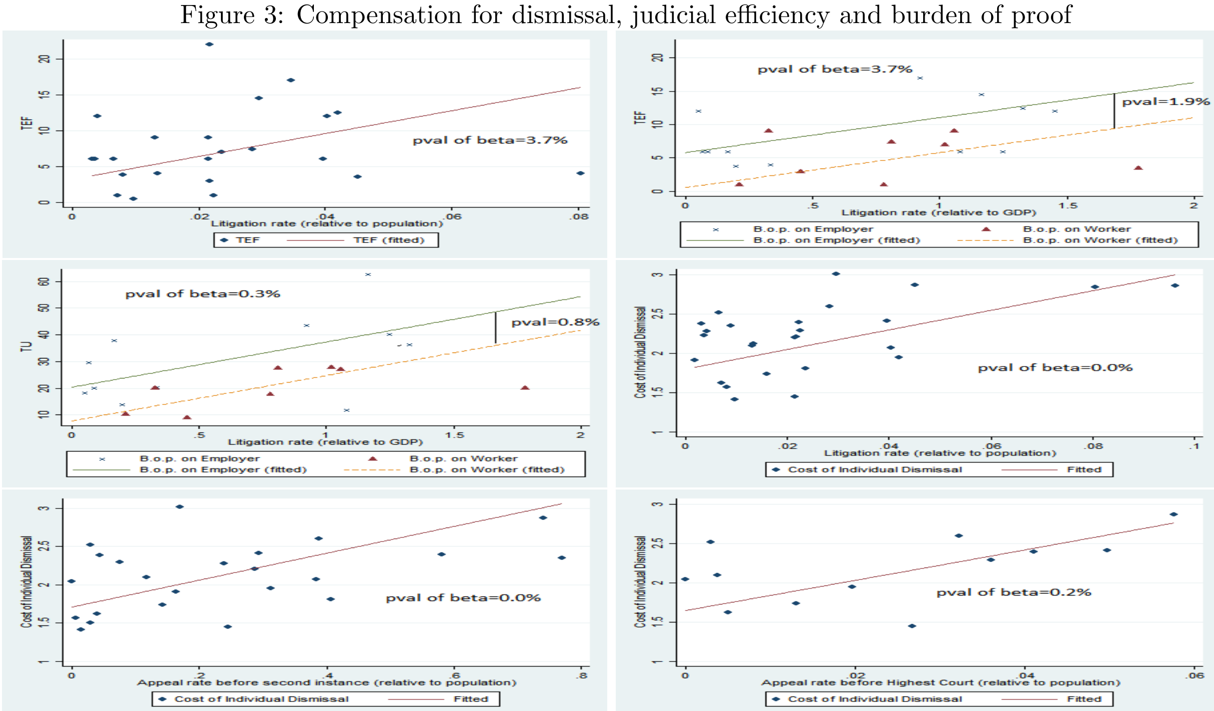


Figure 3: Compensation for dismissal, judicial efficiency and burden of proof



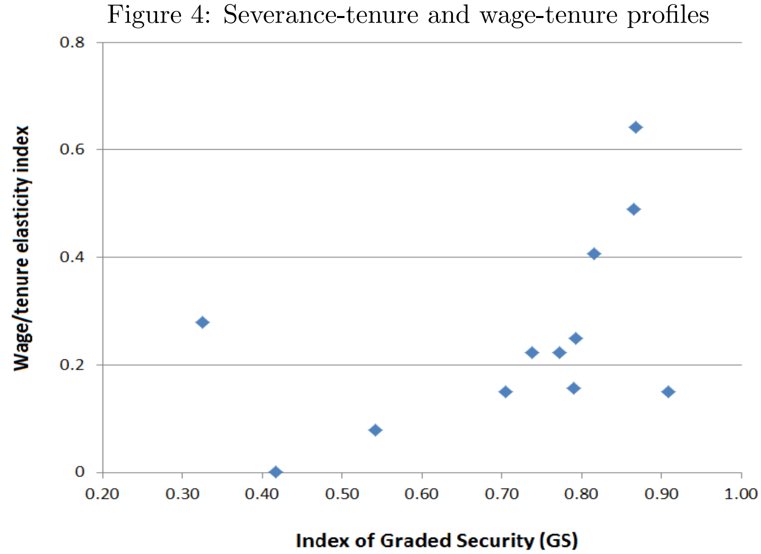
Sources: T_F^E and T_U are as in Table 1.
Cost of Individual Dismissal: Index of Compensation in case of individual dismissal produced by the OECD.
Data on Litigation rate, trial length, appeal rate and burden of proof from Palumbo [45].

5 Discussion

5.1 Severance-tenure and wage-tenure profiles

When wages are deferred and training costs are increasing with tenure, also wages should be increasing with tenure according to our model. Thus, indications as to its empirical relevance may come from the correlation between the severance-tenure and the wage-tenure profiles.[26] Figure 4 displays the apparent elasticities of severance and wages with respect to tenure in all countries for which data are available. In particular, we recover the severance-tenure elasticities from the legal rules as to the mandatory notice period and redundancy payment in the case of fair economic dismissals in the different countries. Cross-country comparable data on EPL specify the level of severance at discrete tenure intervals. Based on this information, we could compute apparent elasticities at different tenure lengths and then aggregate them in the GS index presented in the web Annex.[27] The correlation is positive and statistically significant at conventional (90 per cent) levels. This finding is coherent with our hypothesis that investment costs, or the probability of getting away with it, are increasing with tenure. However, this does not show causality, as other mechanisms may be at play. In any event, the interpretation of the investment as training applies more easily to technical or skilled jobs. Conversely, if the jobs are blue-collar and low-skilled, a more relevant interpretation is that the investment is long-term effort for which the return comes at a lag. In addition, note that with the training interpretation, one may conjecture that the optimal investment level, and hence the cost associated with it, is increasing in the worker's productivity. If productivity increases with tenure in the beginning of the employment relationship, this may rationalize our assumption that the investment cost is increasing with tenure in this period.

Figure 4: Severance-tenure and wage-tenure profiles



Note: The Index of Graded Security is a weighted average of apparent elasticities of severance to tenure at different tenure lengths. The wage/tenure elasticity index is computed in the same way, and at the same durations, based on the empirical estimates of the wage-tenure profile. See the web Annex for details.

[26] No causality is involved here as both severance and wages depend on the costs of training.

[27] For each country of the European Community Household Panel (ECHP), which is particularly suited to this as it is a long panel allowing to identify separately age and tenure, we estimated the following augmented Mincer-type wage equation against micro data on workers' earnings

$$\ln w_i = \alpha + \beta_1 \tau_i + \beta_2 \tau_i^2 + X_i \gamma + \epsilon_i \tag{36}$$

where w denotes hourly wages, τ years of tenure, and the vector X includes educational attainment dummies (tertiary and secondary education), age and gender.

Our model also implies that severance should be increasing with tenure when q is higher. We are not aware of data on Court rulings by tenure of workers. There is some empirical support to the view that judges internalize the re-employment probabilities of workers being laid-off: both the percentage of cases being brought to Courts and the fraction of labor disputes ending with a Court ruling favourable to the worker appear to be higher during cyclical downturns and in relatively depressed labor markets [7], an indication that judges are more protective of workers under these circumstances. Insofar as senior workers face lower re-employment probabilities than junior workers, the legal system may turn out to allow for a greater probability of getting away with it as tenure progresses. Our model also provides an explanation as to why in most countries severance pay is lower for small firms than for large firms: the presence of threshold scales below which the strictest employment protection provisions do not apply can be rationalized in terms of the better monitoring of workers in small organizations.[28]

5.2 Severance and the legal system

Our model has predictions about the relationship between employment protection and the efficiency of the legal system. In particular, it suggests that we should expect to observe higher levels of severance in the countries where the judicial system is less efficient and where the burden of proof is entirely on the employer. Previous work had found that the organization of legal systems, notably the legal origin of countries played an important role in labor market outcomes Botero et al. [13]. Our model can provide an indirect explanation for this, which is based on the effects of the legal system on employment protection regulation. Moreover, we can directly evaluate the empirical relevance of the link between severance and the judicial system implied by the model. Based on recent work done by the OECD in creating cross-country comparable data on legal systems, we can indeed analyse the cross-country correlation between, on the one hand, compensation for fair and unfair dismissals, and, on the other hand, characteristics of judicial systems.

The first two panels at the top of the Figure 3 display the correlation between, on the one hand, T_F^E, and, on the other hand, the litigation rate, that is, the number of the new civil cases commenced in any given year normalized by the population or GDP. This indicator captures congestion, and, per given supply of services, a longer duration and lower quality of judicial services. We allow for separate intercepts for countries where the burden of proof is entirely on the employer, and for countries where it is also partly on the worker. The litigation rate, normalized either by population or GDP, is positively correlated with the mandated months of severance in case of fair economic dismissals. Furthermore, as suggested by the top diagram on the right-hand-side of Figure 3, the level of compensation for fair economic dismissals is higher in countries where the burden of proof is entirely on the employer than in countries where it also falls on the worker (all the countries of this second group are below the regression line for the other group and the intercept is statistically significant at conventional levels).

The other four panels of Figure 3 look at the compensation in case of unfair dismissals (T_U) as well as to a broader measure of the compensation to employees in the case of fair and unfair dismissals. They show that T_U is positively correlated with the litigation rate. Alternative measures of the efficiency of legal systems are trial lenght, and appeal rates before the second instance or higher Courts. These indicators of the efficiency of the judiciary are positively correlated with subjective evaluations of public opinion as to the quality of the legal system collected within the World Value Survey (Palumbo [45]). We find that severance in case of unfair dismissals, T_U, is strongly and positively correlated with trial length, as well as appeal rates before the second instance or higher Courts, but it can be a spurious correlation as appeal rates and trial length appear in our measure of the costs of unfair dismissals. Thus, in the remaining three diagrams we consider a global measure of the compensation for fair and unfair dismissals from regular contracts produced by the OECD, which does not draw on information on trial length and appeal rates. The correlation is once more positive and statistically significant.

[28] While there exists an empirical literature of the effects of EPL on firms of different size, the existence of such treshold is always taken as given in such work. See Garibaldi et al. [23], Boeri and Jimeno [12] and Schivardi e Torrini [53]

6 Final remarks

Research on employment protection fails to account for the relevance of mandatory severance pay in OECD countries. It also neglects two critical features of EPL: the tenure profile of severance pay and the fact that dismissal costs are not only stochastic, but also vary depending on whether they are motivated by economic or disciplinary reasons. In this paper we provide a normative theory of tenure-related severance pay which draws on the involvement of third parties in the decision about the nature, fair or unfair as well as disciplinary or economic, of dismissals.

We show that under a rather broad set of circumstances, and without having to impose any restriction on the distribution of productivity shocks, a severance scheme which is increasing in firm-specific investment costs and in the inefficiency of the legal system is privately efficient in that it avoids separations of jobs that are still originating a positive surplus. This result, which is new for the literature on employment protection, is in line with the reported correlation between, on the one hand, mandatory severance pay, and, on the other hand, OECD indicators of the inefficiency of the legal systems. It implies that reforms of the judiciary can be more effective than labor market reforms in reducing the level of employment protection. For instance, a wider use of workplace arbitration may reduce the informational asymmetries motivating strict employment protection. This result is also important from the standpoint of the sequencing of reforms. If a Government, perhaps under the pressure of IFI, is pressed to reduce EPL, it may find it convenient to at first enhance the scope of workplace arbitration and then intervene on statutory severance.

We also find empirical support for the key rationalization provided by the paper for a positive tenure profile of severance pay, that is, for the fact that investment costs or the probability of getting away with it are increasing with tenure.

The results of this paper are important in evaluating proposals to introduce mandatory compensation increasing steadily with tenure in countries characterized by "contractual dualism", that is, the coexistence of a highly protected segment of the workforce and one segregated into temporary jobs providing low, if any, employment protection. It is also informative as to the optimal slope of the severance tenure profile, depending on the way in which Courts typically protect senior workers and on the costs of training for older workers. Our theory is therefore particularly useful in assessing the scope for "insertion contracts", involving mandatory compensation increasing steadily with tenure. Such "unifying" contracts have been advocated in a number of countries as a measure to reduce "contractual dualism". They have also been recently implemented in the Italian so-called Jobs Act. The theory presented can be used to assess and evaluate some of the reforms currently undertaken or under discussion in various Southern European parliaments. Moreover, our theory suggests that tenure-related severance is efficient even under the typical conditions faced by "temporary workers", that is, under flexible wages, provided that agreed compensation is deferred and that the employer cannot commit not to layoff the worker who has invested in training.

References

[1] Alvarez, F. and Veracierto, M. 2000. Labor Market Policies in an Equilibrium Search Model, *NBER Macroeconomics Annual*, 14, 265-304.

[2] Barth, E. 1997, Firm-Specific Seniority and Wages, *Journal of Labor Economics*, 15 495-506.

[3] Baumann, F. 2010. Severance Payments as a Commitment Device, *Journal of Institutional and Theoretical Economics*, 166, 715-734.

[4] Bentolila, S., and Bertola, G. 1990. Firing Costs and Labor Demand: How Bad is Eurosclerosis?, Review of Economics and Statistics,57, 381-402.

[5] Bentolila, S. 1997. Las decisiones judiciales sobre despidos en Espaa: una perspectiva econmica, Cuadernos de Informacin Econmica, 128-129.

[6] Bertola, G. 1990. Flexibility, Investment and Growth. CEPR Discussion Papers 422.

[7] Bertola, G., Boeri, T., and S. Cazes. 2000. Employment protection in industrialized countries: The case for new indicators. *International Labour Review*, 139(1):57-72.

[8] Bewley, T., 1999 *Why Wages Don't Fall during a Recession* Harvard University Press. Cambridge: Massachusetts

[9] Blanchard O.J. and Tirole J. 2008. The Joint Design of Unemployment Insurance and Employment Protection: A First Pass. *Journal of the European Economic Association*, MIT Press, vol. 6(1), pages 45-77, 03.

[10] Boeri, T. 1999. Enforcement of employment security regulations, on-the-job search and unemployment duration, European Economic Review, 43, 65-89.

[11] Boeri, T. and vanOurs, J. 2013. The Economics of Imperfect Labor Markets, Princeton University Press.

[12] Boeri, T. and Jimeno, J. 2005. The effects of employment protection: Learning from variable enforcement, *European Economic Review*, 49 (8), 2057-77.

[13] Botero, S, Djankov, S., La Porta, R., Lopez-de-Silanes, F. and Shleifer, A. 2004. The Regulation of Labor, *Quarterly Journal of Economics 119*, 1339-1382.

[14] Carmichael, H.L. 1988. Incentives in Academics: Why Is There Tenure?, *Journal of Political Economy*, 96(3): 453-72, June

[15] Chéron,A., J.-O. Hairault, F. Langot, 2011, *Age-dependent employment protection*, *Economic Journal*, 212, 1477-1505

[16] Deffains, Bruno, Gabuthy, Y., Lambert, E., 2010 Labour Disputes, investment decisions and the judiciary, *Labour Economics* Volume 17, 2, 424-433

[17] European Labour Law Network. 2011. Dismissal - particularly for business reasons - and Employment Protection. Thematic Report, Contract No. VC/2010/1636.

[18] Fella, G. 2000. Efficiency wage and efficient redundancy pay, *European Economic Review*, 44, 1473-1490.

[19] Fella, G. 2012. Matching, wage rigidities and efficient severance pay, *The B.E. Journal of Economic Analysis and Policy* 12.

[20] Fraisse, H, Kramarz, F. and Prost, C. 2009. Labor Court Inputs, Judicial Cases Outcomes and Labor Flows: Identifying Real EPL. Working Paper Banque de France, n.404.

[21] Galdon-Sanchez, E. J. and M. Guell. 2003. Dismissal conflicts and unemployment. *European Economic Review*, Elsevier, 47(2):323–335.

[22] Garibaldi, P. 1998. Job flow dynamics and firing restrictions. *European Economic Review* 42, No. 2: 245-75.

[23] Garibaldi, P., L., Pacelli (2004) Employment Protection Legislation and the Size of Firms, *Giornale degli Economisti e Annuali di Economia*. Num. 4, 33-68

[24] Garibaldi, P. and Violante G.L. 2005, The Employment Effects of Severance Payments with Wage Rigidities. *The Economic Journal*, 115: 799832

[25] Goerke L, and Pannenberg, M., 2010, An Economic Analysis of Dismissal Legislation: Determinants of Severance Pay in West Germany,*International Review of Law and Economics*, Vol. 30(1), 71-85.

[26] Guerrieri, V., Shimer, R., and Wright, R. (2010). Adverse selection in competitive search equilibrium. Econometrica, 78(6), 1823-1862.

[27] Flabbi,L. and A.Ichino. 2001. Productivity, seniority and wages: new evidence from personnel data, *Labour*, Elsevier, vol 8(3), pp 359-387.

[28] Ichino, A., Polo, M., and Rettore, E. (2003). Are judges biased by labor market conditions?, *European Economic Review* 47, 913-944

[29] Ilmakunnas, P., vanOurs, J. and V., Skirbekk (2010), Age and Productivity, in Garibaldi, P., Oliveira-Martins, J. and vanOurs, J. (eds.) *Ageing, Health and Productivity*, Oxford University Press.

[30] Kodrzycki, Y. K. (1998), Effects of employer-provided severance benefits on reemployment outcomes,*New England Economic Review*, pp.41-68.

[31] Kotlikoff, L. and Jagadeesh G. 1992. Estimating a Firm's Age-Productivity Profile Using the Present Value of Workers' Earnings. *Quarterly Journal of Economics*, 107, 1215-1242.

[32] Lazear, E.P. and Moore R.L. 1994. Incentives, Productivity and Labor Contracts. *Quarterly Journal of Economics*, 99, 275-296.

[33] Lazear, E.P. 1990. Job security provisions and employment. *Quarterly Journal of Economics*, 105:699–726.

[34] Levine, D.I. Just-cause employment policies in the presence of worker adverse selection *Journal of Labor Economics* 9, 249-305

[35] Malo, M.A. 1998. Las indemnizaciones por despido, un problema de negociacin. Editorial ACARL, Madrid.

[36] Malo, M.A., and Toharia, L. 1999. Costes de despido y creacin de empleo. Ministerio de Trabajo y Asuntos Sociales, Madrid

[37] Malo, M.A. 2000. A simple model of severance pay determination: The case of individual dismissals in Spain. *Labour*, 14(2):269–290.

[38] Medoff, J.L. and Katharine G. A. 1980, Experience, Performance and Earnings, *Quarterly Journal of Economics*, 95, 186-216.

[39] Menzio, G., and Moen, E.R. 2010. "Worker Replacement". *Journal of Monetary Economics*, 57, 623-636

[40] Moen, E.R. (1997). "Competitive Search Equilibrium". *Journal of Political Economy* 1997, vol 105, No.2, 385-411.

[41] Moen, E. R. and Rosen, Å. 2005. Incentive Contracts and Adverse Selection, *Scandinavian Journal of Economics*, 107(2), 279 - 298.

[42] Moen, E.R. and Rosen, Å(2011). Incentives in Competititve Search Equilibrium (with sa Rosen). *Review of Economic Studies* 2011 vol 78 (2) 733-761.

[43] Mortensen, D., and C. Pissarides, 1994, Job Creation and Job Destruction in the Theory of Unemployment. *Review of Economic Studies*

[44] Mortensen, D., and C. Pissarides, 1999, New development in models of search in the labor market, Elsevier.

[45] Palumbo, G. et al. 2013. The Economics of Civil Justice: New Cross-country Data and Empirics. OECD *Economics Department Working Papers* 1060, OECD Publishing.

[46] Pissarides, C.A. (2001) Employment Protection, *Labour Economics* Vol. 8. 131-159

[47] Pfann, G. and Frenk, M. (2014) Firing Costs and Labor Demand in Recessions, mimeo.

[48] Postel-Vinay, F. and Turon, H. 2013. The Impact of Firing Restrictions on Labor Market Equilibrium in the Presence of On-the-job Search, Economic Journal, 124, 31-61.

[49] Rothschild, M. and Stiglitz, J. 1976. Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information. *Quarterly Journal of Economics 17*, 557-66.

[50] Rühmann, P. and Südekum, J. 2001. Severance Payments and Firm-Specific Human Capital, Discussion Paper n. 111 of the University of Gòttingen.

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

[52] Stähler, N. 2008 Firing Costs, Severance Payments, judicial mistakes and unemployment, Labour Economics, Volume 15, 6, 1162-1178

[53] Schivardi, F., and Torrini, R., 2008, Identifying the Effects of Firing Restrictions through Size-Contingent Differences in Regulation, with R. Tor Labour Economics, Vol. 15, pp. 482511.

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

[55] Venn, D. 2009. Legislation, Collective Bargaining and Enforcement: Updating the OECD Employment Protection Indicators. OECD *Social, Employment and Migration Working Papers* 89, OECD Publishing.