Available online at www.sciencedirect.com
SCIENCE DIRECT
Labour Economics 11 (2004) 669–686
LABOUR ECONOMICS
www.elsevier.com/locate/econbase

Hierarchies in organisations and labour market competition

Gianni De Fraja*

University of York, Università di Roma "Tor Vergata", and C.E.P.R., UK

Received 1 March 2003; received in revised form 1 July 2003; accepted 1 November 2003
Available online 14 January 2004

Abstract

This paper studies the endogenous determination of hierarchies in firms. Firms can design a hierarchy with a continuum of ranks, one for each ability level. Nevertheless, in the market equilibrium, they choose to have only a finite number of ranks, so that in each rank there are workers of different abilities, who produce different output, and receive the same wage. It is also shown that an increase in the extent of labour market competition reduces the number of ranks in the hierarchy.

JEL classification: L23; J30
Keywords: Labour market competition; Hierarchies; Delayering; Organisation theory

1. Introduction

This paper takes the view that a firm's internal hierarchy is endogenously determined, and studies the effects of the strength of labour market competition on the number of ranks in a firm's hierarchy. Existing theories of the internal organisation of firms have paid surprisingly little attention to the influence of labour market competition: they either assume an exogenously given job ladder and address the problem of optimally allocating workers to tasks (Calvo and Wellisz, 1979; Rosen, 1982; Waldman, 1984a; Ricart i Costa, 1988; Bernhardt, 1995; Gibbons, 1998 for a survey), or see the emergence of hierarchies as a consequence of exogenous factors such as bounded control span and economies of scale in production (Williamson, 1967; Mirrlees, 1976; Calvo and Wellisz, 1978). These

* Department of Economics, University of York, Heslington, York YO10 5DD, UK. Fax: +44-1904-433759.
E-mail address: gd4@york.ac.uk (G. De Fraja).

0927-5371/$ - see front matter
doi:10.1016/j.labeco.2003.11.001

"technological" explanations of hierarchies in organisations may not however account for changes in the hierarchical structure the extent of which cannot be attributed to changes in the underlying exogenous factors alone, such as the recent trends towards delayering,[1] nor for radical organisational differences in similar firms in different countries, as documented for example in Dore's celebrated comparison between an English and a Japanese factory (Dore, 1973).

Our analysis is built around an idea originally proposed by Waldman (1984b): workers differ in ability, and firms have an informational advantage: a firm knows the ability of its workers, but the workers do not know their own ability. Outside employers cannot observe output, but can observe the position in the hierarchy of a given worker. While in Waldman's model there are only two given ranks, we let firms choose the number of ranks to maximise their profit, allowing potentially a continuum of ranks. Our main result is that if there is sufficient substitutability between outside workers and a firm's own employees, then the equilibrium hierarchy displays a finite set of ranks. We also show that an increase in this substitutability can only reduce the number of ranks. In our model, the degree of substitutability between workers in different firms is a measure of the firm's specific human capital, and therefore of the strength of labour market competition, in that it is inversely related to a firm's monopsonistic power. In this sense, a possible interpretation of our analysis is the provision of a theoretical link between the trend towards delayering mentioned above and the increase in the extent of labour market competition.[2]

The intuition underlying our results is that the shape of a firm's hierarchy is determined by a trade-off between assigning workers to their most suitable positions, where they are most productive, but thus signalling their value to the market, and assigning them to positions where they are less productive, or need to be paid more, but where they are less attractive to outside employers. In other words: on the one hand, the more ranks a firm has, the more it reveals the ability of its workers to the market: outside employers will be willing to increase their wage offer to a given worker, and his wage must be raised to prevent him from leaving. On the other hand, bunching workers is also costly, because workers are required to exert an inefficient level of effort. The wider a rank, the lower the wage that is needed to stop workers being poached, but the bigger the distortion in effort. When labour market competition increases, workers become more valuable to other employers, and must therefore be paid more to stop them leaving: this reduces the gain from the efficient choice of effort, and therefore ranks become wider and thus fewer.

The model is presented in Section 2 and Section 3 contains the results of the paper: Proposition 2 shows that the hierarchy has a finite number of ranks, Proposition 3 that this number can only be reduced as a consequence of a decrease of the current employer's monopsonistic power.

---

[1] "Many large companies are learning that they can do better with four layers of management than with 12" Tom Peters, (in The World in 1994, The Economist, 1993, p. 120). Note that this is by no means new: in 1914, Ford "reduced the number of basic wage rates at Ford from sixty-nine to eight" (Jacoby, 1984, p. 37).

[2] "A rise in inter-firm mobility" is one of the effects that has been associated with the trend towards delayering (Littler et al., 2003, p 227; this paper contains an exhaustive survey of the recent literature on delayering), though the causal direction is the opposite of that hypothesised here.

2. The model

There is a continuum of workers. They are all identical, except for their exogenously determined skill level: a worker is characterised by a parameter $\mu \in [\underline{\mu}, \bar{\mu}]$. Let $\int_{\underline{\mu}}^{\mu} g(\tilde{\mu})\mathrm{d}\tilde{\mu}$ be the number of workers with skill parameter no greater than μ. In other words, $\frac{g(\mu)}{\int_{\underline{\mu}}^{\bar{\mu}} g(\tilde{\mu})\mathrm{d}\tilde{\mu}} \in [0,1]$ is the density of the workers' ability. Workers live two periods: we refer to a worker in his first period as "young" and to a worker in his second period as "old". In each period, workers have the option not to work and enjoy leisure only. This gives utility normalised to zero. If they do work, their utility is given by $u(w,e) = w - v(e)$, where w is their monetary income and $e \geq 0$ is the effort exerted on the job; v satisfies $v', v'' > 0$: workers are risk neutral and have increasing marginal disutility of effort. A worker's lifetime utility is given by the undiscounted sum of utilities.[3]

The economy comprises a large number of infinitely lived, profit maximising firms, which use labour to produce output, sold at a given price in the market. For simplicity, we assume that the distribution of the workers hired by each firm corresponds to a scaled down version of the economy-wide distribution: the density of workers hired by a representative firm, $f(\mu)$, is given by $f(\mu) = \lambda g(\mu)$, for some $\lambda > 0$, for every $\mu \in [\underline{\mu}, \bar{\mu}]$. Without further loss of generality, we normalise the firm's size in such a way that $\int_{\underline{\mu}}^{\bar{\mu}} f(\mu)\mathrm{d}\mu = 1$.

Assumption 1 below describes the relationship between the skill and effort of a worker and his output. It captures the creation of firm specific human capital: a worker of a given skill exerting a given effort produces more in the firm where he has already worked than he would produce in a different firm. We rule out externalities among workers, so that the output of a worker does not depend on the skill of his fellow workers. The firms' technology displays therefore constant returns to scale in the specific sense that the output of a firm is given by the sum of the output of all its workers. Formally:

Assumption 1: The value of the output of a worker of skill $\mu \in [\underline{\mu}, \bar{\mu}]$ who exerts effort $e \geq 0$ is given by $y(\mu,e)$ if he has not worked in the firm before and by $(1+s)y(\mu,e)$, with $s > 0$, if he has. $y(\mu,e)$ is twice differentiable and satisfies: $y_{\mu} > 0$, $y_e > 0$, $y_{\mu\mu} \leq 0$, $y_{ee} \leq 0$, $y_{\mu e} > 0$.

Subscripts denote derivatives. The next assumption specifies who observes what, and therefore it determines the constraints on the labour contracts that can be signed.

Assumption 2: The effort exerted by a worker is observed both by himself and by his employer. The worker's output is observed only by his employer, at the end of each period, and his skill is not observed by anyone. An employment contract specifies the exertion of a given level of effort by the worker and the payment of a given wage by the employer. Employment contracts are observed by the market.

As in ,Waldman (1984b) the current employer has an informational advantage over both all the other employers in the economy *and* the worker himself. We have in mind

[3] This avoids the need to consider the additional parameter given by the discount rate, which would also account for the possibility of exogenous turnover. Since all young workers are treated in the same way, this possibility would not affect the nature of our results, even if the exogenous turnover were not independent of skill. All is necessary for our analysis is that firms know the distribution of "old" workers.

situations where output and effort are difficult variables to measure: the worker knows the effort he has put on the job, but only the employer can gauge exactly[4] the effect of his effort on the firm's performance. Only the employer, for example, can ascertain the market value of a worker's personal finishing touches to the final product, of his advertising slogans, sales techniques, popularity with customers and so on. In situations where production is by teams, μ may measure the worker's ability to motivate effectively his subordinates and get on well with his peers and superiors[5]: our assumption that a worker's output depends only on his skill and his effort implies that a worker's marginal contribution is independent of the rest of the team's output.

Assumption 2 implies that a worker's employer can infer his skill from her observation of his effort and his output. With regard to the assumption that effort is contractible, a reputation argument can be invoked to justify it. If the firm is (implicitly or explicitly) committed to reward effort according to some standard, and if, plausibly, other workers can also observe the effort level exerted by a given worker, an employer who is seen not to reward a worker's effort adequately would make the rest of her workers less inclined to remain with her, or do so only for a higher wage than suggested by competitive considerations alone (see Carmichael, 1984, for a model built on a reputational mechanism).

Assumption 3 describes formally the timing of the game.

Assumption 3

3.1 At the beginning of each period, each employer offers a contract (i.e. a wage-effort pair) to each of her newly hired young workers.
3.2 At the end of the period, having learned the ability of each of her workers, each employer offers them a contract (a wage-effort pair) for the second period of their working age.
3.3 Subsequently, all employers in the market observe the contract offered to all the (old) workers. Every employer can offer a contract to any of the (old) workers employed by any other employer.
3.4 Each old worker who has received more than one offer selects his preferred offer (ties are split in favour of the current employer[6]).

[4] The crucial assumption for our results to obtain is simply that the employer has better information than the worker. Therefore, we may confidently conjecture that conclusions similar to our own would be obtained in a stochastic environment along the lines of Ricart i Costa (1988). For example, suppose that a worker's effort translates into output with a random noise. The employer observes output better than the worker, and thus has a better signal of a worker's skill than the market and the worker. Our deterministic set-up is the extreme case of this scenario, where the employer has a perfect signal and the worker a totally uninformative signal.

[5] In sports, the selection of a national team requires more skill in the cases where output is difficult to measure. Compare soccer with track and field events: in the latter, selection is often automatically based on recent past performance, while in the former is entrusted to experienced (and highly paid) managers. An important component of their skill is the ability to judge each individual player's contribution to the team's chance of success.

[6] We do not allow the current employer to make counteroffers. De Fraja (1995) obtains very similar results as the present paper in a model where employers can make counteroffers. The main difference is that, in equilibrium, wages and effort levels are lower; this is due to the extreme form of winner's curse (by which an outside employer whose offer is not matched infers that she has contacted a relatively low ability worker) identified in the present situation by Milgrom and Oster (1987) (see footnote 8 below).

We do not allow workers and firms to write lifetime contracts, as, in practice, contractual terms are a small fraction of a worker's lifetime. Assumption 3.3, the full publicness of wage offers, may appear quite strong, and, constituting as it does one of the central assumptions of our model, requires some discussion. If it is true that remuneration packages are not normally public, it is also the case that grade assignments are: it is usually known whether somebody is a director, an executive, a controller, a manager, a supervisor, a shopfloor worker and so on. Moreover, there often exist reliable guides of grade and salary levels: for example, the model of the company car assigned to an employee is often correlated with his overall remuneration.[7] In addition, the role of head-hunters is to spread information about people's position in their current organisation. More subtly, as long as, as is the case both in our model and in practice, wage is correlated with the value of a worker to his employer, and consequently to his skill, it is in the interest of a worker to reveal the value of his compensation package to any other potential employer, in order to receive offers worth more than this. There is no common interest of employer and employee to hide any payment the former makes the latter, as there would be, for example to the taxman. Assumption 3.3 also implies that outside employers also observe the hierarchy in each firm, that is the number of ranks and the number of workers in each rank.

3. Hierarchies in a competitive environment

Given a set $M\subseteq[\underline{\mu},\bar{\mu}]$, let $L(M)$ be the proportion of workers whose ability is in M: $L(M)=\frac{\int_{\mu\in M}f(\mu)\mathrm{d}\mu}{\int_{\mu\in[\underline{\mu},\bar{\mu}]}f(\mu)\mathrm{d}\mu}$. For a set $M\subseteq[\underline{\mu},\bar{\mu}]$ with $L(M)>0$, let e_M^s be the solution in e of: $\frac{1}{L(M)}\int_{\mu\in M}(1+s)y_e(\mu,e)f(\mu)\mathrm{d}\mu=v'(e)$; for a subset $\{\mu\}\subseteq[\underline{\mu},\bar{\mu}]$ let $e_{\{\mu\}}^s$ be the solution in e of $(1+s)y_e(\mu,e)=v'(e)(y_{ee}(.)<0$ and $v''(.)>0$ imply that e_M^s is uniquely defined. e_M^s is the optimal level of effort for workers in the set M given that all workers in the set M exert the same level of effort. It is straightforward to establish that all young workers are required to exert effort $e_{[\underline{\mu},\bar{\mu}]}^0$ and they are paid wage $(e_{[\underline{\mu},\bar{\mu}]}^0)-E(u_2)$, where *where* where $E(u_2)$ is the expected utility of an old worker.

A useful benchmark case is the absence of labour market competition. This obtains when s is so high that an employer need not worry about the presence of potential outside employers.

Proposition 1: *Let $y(\bar{\mu},e_{\{\bar{\mu}\}}^0)\leq v(e_{\{\bar{\mu}\}}^0)$. Moreover, let s^0 satisfy: $(1+s^0)y(\underline{\mu},e_{\{\underline{\mu}\}}^{s^0})-v(e_{\{\underline{\mu}\}}^{s^0})=0$. For every $s\geq s^0$, the optimal hierarchy is such that effort is $e_{\{\mu\}}^s$ and wage is $v(e_{\{\mu\}}^s)$ for every $\mu\in[\underline{\mu},\bar{\mu}]$.*

In the absence of labour market competition, an employer who knows the ability of a worker will simply ask him to exert the effort level which maximises the difference

[7] Note that we are not imposing the requirement that workers are paid a fixed wage: their remuneration may be in the form, for example of shares options. What is necessary for our results is that all workers in the same rank receive the same compensation package.

between output and the wage, which makes a worker indifferent between accepting to work for his current employer and remaining unemployed. Effort is $e^s_{\{\mu\}}$ and wage $v(e^s_{\{\mu\}})$. As shown in the proof, the condition in the statement of the proposition implies that workers are at most indifferent between working for the current employer and for a new employer (the indifference only occurring when $\mu=\bar{\mu}$). An immediate consequence of the proposition follows from the observation that $\mu_1>\mu_2$ implies $e^s_{\{\mu_1\}}>e^s_{\{\mu_2\}}$. Therefore, in the absence of labour market competition, more able workers are required to exert more effort, and paid accordingly more. We will refer to $e^s_{\{\mu\}}$ as the *efficient level of effort* for workers of skill μ.

The equilibrium concept appropriate for the case of labour market competition is sequential equilibrium; loosely speaking, upon observation of a firm's offers, outside employers form beliefs about each of its workers. In equilibrium these beliefs must be consistent with that firm's strategy. Specifically, a firm offers contracts to its workers which are described by a set of triples $\{(w_i,e_i,M_i)\}_{i\in I}$, where I is an index set, not necessarily finite, and w_i and e_i are the wage and the effort offered to workers in the set $M_i\subseteq[\underline{\mu},\bar{\mu}]$, with $M_{i_1}\cap M_{i_2}=\emptyset$ for every $i_1,i_2\in I$, and $\bigcup_{i\in I}M_i\subseteq[\underline{\mu},\bar{\mu}]$, that is, the sets $\{M_i\}_{i\in I}$ form a partition of the set of workers who are retained. Because the market only observes the *proportion* of workers who are offered a given contract, the market observes $\{(w_i,e_i,L(M))\}_{i\in I}$, and forms beliefs $r_k(\{(w_i,e_i,L(M_i))\}_{i\in I})$, $k\in I$, where $r_k\subseteq[\underline{\mu},\bar{\mu}]$, with $r_{k_1}\cap r_{k_2}=\emptyset$ for every $k_1,k_2\in I$. In equilibrium $r_i=M_i$ for every $i\in I$. In general, out-of-equilibrium beliefs can affect the set of equilibria of a given game, so that it might be necessary to impose restriction to ensure some required property of the equilibrium set. In this model, the following is sufficient to ensure our results.

Assumption 4: Let $\{(w_i,e_i,M_i)\}_{i\in I}$ and $\{(w_j,e_j,M_j)\}_{j\in J}$ be two contracts with $I=J$. Let there exist a one-to-one correspondence between I and J, $g:I\rightarrow J$, $j=g(i)$, with the following properties: (i) $L(M_i)=L(M_{g(i)})$ for every $i\in I$, and (ii) $w_{i_1}\gtreqless w_{i_2}$ implies $w_{g(i_1)}\gtreqless w_{g(i_2)}$ and $e_{i_1}\gtreqless e_{i_2}$ implies $e_{g(i_1)}\gtreqless e_{g(i_2)}$ for every $i_1,i_2\in I$. Then $r_k(\{(w_i,e_i,L(M_i))\}_{i\in I})=r_k(\{(w_{g(i)},e_{g(i)},L(M_{g(i)}))\}_{g(i)\in J})$ for every $k\in I$.

In words, Assumption 4 implies that, if a firm changes marginally the wage or effort in one of its ranks, but changes neither (i) the number of workers in that rank, nor (ii) the order of this rank in the overall wage and effort hierarchy, then the market does not change its beliefs with regard to the way in which the firm assigns workers to rank. This seems natural, and it entails that the beliefs of the market are affected only by the hierarchy.

We can now present the main result of this paper, Proposition 2, which characterises the hierarchical structure of the firm.

Proposition 2: *Let $s>0$. Let $y(\underline{\mu},e^0_{\{\underline{\mu}\}})\geq v(e^0_{\{\underline{\mu}\}})$. Let $\{(w_i,e_i,M_i)\}_{i\in I}$ be the profit maximising strategy of the representative firm. Then I is a finite set, $I=\{1,2,\ldots n\}$, there exists a finite set of points $\mu_0,\mu_1,\ldots,\mu_{n-1},\mu_n\in[\underline{\mu},\bar{\mu}]$, with $\underline{\mu}=\mu_0<\mu_1<\ldots<\mu_{n-1}<\mu_n=\bar{\mu}$, with $M_i=[\mu_{i-1},\mu_i)$, $e_i=e^s_{[\mu_{i-1},\mu_i]}$, and $w_i=\frac{1}{L(M_i)}\int_{\mu\in M_i}y(\mu,e^0_{M_i})f(\mu)d\mu-v(e^0_{M_i})+v(e^s_{M_i})$, for almost all $\mu\in[\mu_{i-1},\mu_i)$, for $i=1,\ldots,n$.*

The proofs of the results are in Appendix A. In equilibrium, all workers are retained and they are offered a wage schedule and an effort schedule, which are non-decreasing step functions, each step corresponding to a rank, with all the workers in the same rank receiving the same wage and exerting the same effort. The effort is the optimal effort for workers in that rank, and the wage is what is necessary to make the worker indifferent between this effort-wage pair and accepting an offer which gives zero expected profit to an outside employer.[8]

To understand what determines this result, suppose that the current employer separated fully her workers by creating a hierarchy of the type described in Proposition 1. In this case, effort would be at the efficient level, and the wage would be set to match the output with the efficient effort in a new firm, and to compensate the worker for his effort. When instead workers of different skill are bunched at the same effort level, their output, net of the reward for effort, is lower than it would be if each exerted the efficient effort: $\int_{\mu\in M}[(1+s)y(\mu,e^s_M)-v(e^s_M)]f(\mu)\mathrm{d}\mu<\int_{\mu\in M}[(1+s)y(\mu,e^s_{\{\mu\}})-v(e^s_{\{\mu\}})f(\mu)\mathrm{d}\mu$, but wage is reduced even more, because effort would be reduced further in an outside employer, whose productivity is lower since she does not benefit from the firm's specific human capital accrued during the worker's youth. Put it differently, widening the ranks from zero to a small positive width determines a loss in output, which is a second order effect, and a reduced wage, which is a first order effect. Therefore it pays an employer to create ranks and bunch workers.[9] As the proof shows, the crucial assumption is that effort and ability are complements ($y_{e\mu}(.)>0$). This implies that, in any interval, average effort is lower when workers are bunched than when they are separated. The process of bunching, however, need not go to the extreme, and it is possible to have more than one rank: this is so because the output loss due to the effort distortion might overcome the wage saving if ranks are wide enough.

Fig. 1 illustrates the situation. It compares output and wage as determined in the optimal hierarchy in Proposition 2 (solid lines) with their level when effort is efficiently set (dashed lines). Note that some workers are required to exert *more* effort than it is efficient (this does not happen in Waldman (1984b). Workers earn a strictly positive rent (this is shown as the grey area in Fig. 1): they receive a wage and are required a level of effort

[8] Note that the firm might well knowingly be paying a worker (one whose ability is low within the rank) above his marginal product: if a worker who receives more than his marginal product were demoted to a lower rank, then the remaining workers would have to paid more, and this might reduce the firm's profit. This is a consequence of an employer's inability to match an outside offer: she must make sure that the average worker in a rank is retained. One might therefore think that, if the current employer were allowed to make counteroffers, she could select the workers to whom a counteroffer is made and retain only the higher ability workers within the rank. This is not so, however: the outside employers would take into account this rational response of the current employer when evaluating the profitability of making offers to another employer's worker. This is an instance of the winner's curse: the outside employer makes an offer such that she makes zero profit when the least able worker in a given rank accepts her offer. (Milgrom and Oster, 1987; De Fraja, 1995).

[9] The existence of a finite number of ranks corresponds to Laffont and Tirole's results that separating equilibria are non-feasible (respectively non-optimal) in a dynamic principal agent setting with no commitment (respectively, with commitment and renegotiation) (Laffont and Tirole, 1993, part IV). Given the different information structure of the two models, this suggests that bunching contracts are pervasive in dynamic environments where one party obtains a rent for her informational advantage, which is destroyed by separation.

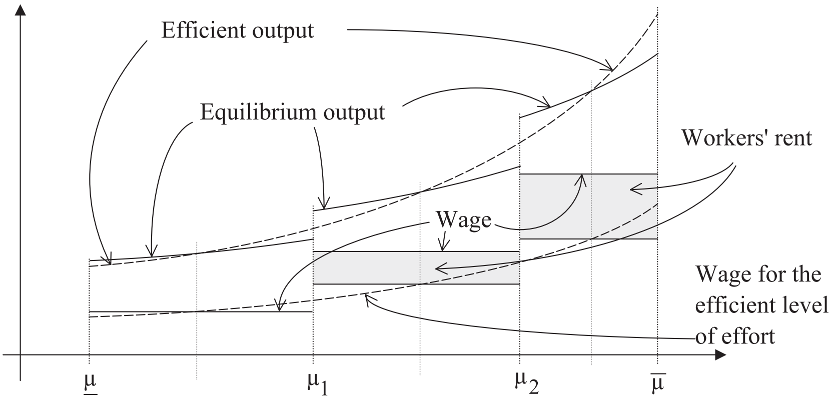

Fig. 1. Efficient output and equilibrium output, equilibrium wage, wage for the efficient level of effort and workers' rent.

such that the resulting utility is above what they would achieve were they unemployed. This rent is *not* due to asymmetric information between workers and employers, but is exclusively the effect of labour market competition. Moreover, as Fig. 1 suggests, and as we establish formally in Lemma 1, this rent is increasing with skill and therefore with wage.

Lemma 1: $w_i - v(e^s_{[\mu_{i-1},\mu_i]}) < w_{i+1} - v(e^s_{[\mu_i,\mu_{i+1}]}),\ i = 1,\ldots,n-1.$

This tallies with casual observation, for example, with the stylised facts that it is more costly to give up a high level job, even if it requires a high level of effort, and that better paid workers are better off. The intuition for this conclusion is that better workers are more valuable to outside employers, and the latter will therefore be willing to give these workers more rent, (in fact all their potential rent): therefore workers must receive more rent from their current employer to remain with her. MacLeod and Malcomson (1988) also study the endogenous determination of hierarchies, and find that as a worker progresses in his career, his rent *decreases*. This marked contrast with Lemma 1 is due to the fact that, in their paper, workers have an information advantage, and firms use promotion as a device to motivate able workers. But as a worker approaches the rank appropriate to his ability, his information advantage is reduced, and consequently so is his rent. Note also that, by Lemma 1, all workers receive a utility greater than their reservation utility for the second period. This implies that, in their first period, they will be willing to work for a utility level lower than the reservation utility for that period. This of course corresponds to the observation, as well as the theoretical analysis (Milgrom and Roberts, 1992), of people working harder in the their youth.

The inefficient choice of effort implies that firms and workers would be better off if the firms could commit to a given hierarchical structure for their old workers. This ability to commit could be given by a reputation argument, similar to that invoked to justify the firms' ability to commit to a given effort level. The first best, implying effort $e^s_{\{\mu\}}$ for a worker of ability μ, could then be achieved. The difficulty with this argument is that labour contracts are between each individual worker and his employer, and they do not typically specify the assignment of all the other workers. Moreover, this would not be robust to technological changes or the relaxation of the hypothesis that the firm knows the distribution of its young workers: in both these cases, the firm does not know when it hires young workers what the first best hierarchy will be when they become old. A further difficulty is that, for this mechanism to operate, workers would need to make an up-front payment (the firm "sells" the future rents to its young workers). Liquidity constraints and limited liability may preclude these up-front payments.

Note also that the presence of labour market competition creates an externality between workers of different ability. Because a firm optimises over the entire distribution of workers' ability, changes in the distribution of workers' abilities affect the hierarchical structure of the firm and a worker's compensation (this is not the case in MacLeod and Malcomson, 1988).

Note that it is possible that both conditions $y(\bar{\mu},e^0_{\{\bar{\mu}\}}) \leq v(e^0_{\{\bar{\mu}\}})$ and $y(\underline{\mu},e^0_{\{\underline{\mu}\}}) \geq v(e^0_{\{\underline{\mu}\}})$ are violated, so that neither Proposition 1 nor Proposition 2 apply. In this case, the hierarchy in the firm is "split". There is a value $\mu^* \in [\underline{\mu},\bar{\mu}]$ such that the hierarchy described in Proposition 2 applies to workers with high skill, $\mu \in (\mu^*,\bar{\mu}]$, and the hierarchy described in Proposition 1 to workers with low skill, $\mu \in [\underline{\mu},\mu^*]$.[10]

We end the paper with the question of the effects of labour market competition on the number of ranks in the hierarchy. We capture the effects of competition by a reduction in s, the degree of firm specific human capital: the workers of an outside employer become better substitutes for any employer's experienced workers. That a reduction of s also implies a reduction in the second period output does not matter in the partial equilibrium set-up considered here (the model can be embedded in a general equilibrium framework: see De Fraja, 1995).

Proposition 3: *Let s_L, s_H be given, with $s_L < s_H$. Let n_i be the number of ranks determined in the stationary equilibrium when $s = s_i$, $i = H,L$. Then $n_L \leq n_H$.*

That is, an increase in labour market competition can only reduce the number of ranks: if workers acquire human capital which is less firm specific, then they will tend to be placed in a flatter hierarchy, one with fewer steps. This also contrasts with MacLeod and Malcomson (1988), where a reduction in labour market friction—measured by the length of the minimum period of employment—increases the number of ranks and reduces the salary gap between them. The theme of the complexity of the hierarchy recurs in the literature on internal labour markets: in this perspective, Proposition 3 is consistent with the observation that "internal labour market are likely to be more elaborate where (...) tasks require a high degree of firm specific knowledge" (Kanter, 1984, p. 127). The results

[10] A rigorous proof of this result is in De Fraja (1995). We note that the internal organisation of the firm implied in this case is suggestive of a dual labour market: the lower ability workers are paid according to their output (piece rate reward) and do not earn any rent; the higher ability workers are put in a hierarchical structure, are given a wage independent of output and earn a strictly positive rent. This reflects the organisation of the internal labour market in many firms, where the lower ranks are often appointed on a casual basis, or hourly only when needed, or contracted out to outside firms, with weaker unions and employment protection (Doeringer and Piore, 1971). Higher rank workers, by contrast, tend to have more formal contracts.

of this section also suggest a direction for empirical research aimed at determining the cause of the recent trend towards delayering: to the extent that labour market competition (especially for middle management) is keener than product market competition, then one would expect differences in the degree to which direct monitoring of effort is possible in different industries, and in the substitutability of workers, s, to affect the extent of delayering in these different industries.

The intuition for the result in Proposition 3 is straightforward. The reason why an employer bunches workers together is to hide them from the outside market. The benefit is that they can be paid less, the cost is that their effort is distorted with respect to the efficient level: if ranks are wider, then the wage is lowered (because the information conveyed to the outside employers is reduced), but the effort distortion is increased. When s is high, distorting effort is relatively more costly, and the firm will be more willing to incur the higher wage cost necessary to fine tune effort to skill, and reduce the effort distortion.

Proposition 2 holds for the case $s>0$. It is easy to verify that if we had $s=0$, then *any* hierarchical structure would an equilibrium, and that at any equilibrium the current employers are completely unable to extract a profit from her second period workers. If an employer made a profit from her old workers, then any outside employer could exactly match the hierarchy offered by the current employer, increasing the wage by an arbitrarily small ε, attract all the workers and still make a positive profit. Given the outcome of no period 2 profits for the current employer, then she will be indifferent with regard to the hierarchical structure in her firm. This implies that there is a discontinuity in the equilibrium behaviour of the model as $s \to 0$: the hierarchy is uniquely determined for any strictly positive s, but it "explodes" into a very big set when s reaches 0.[11] Discontinuities of this type are fairly common in game theoretic interactions (see for example the bargaining problem as the players' impatience goes to 0, Rubinstein, 1982), and here we follow the standard practice of analysing the limiting behaviour of the model rather than the special case where frictions are exactly 0.

4. Conclusion

We study in this paper the endogenous determination of hierarchies in organisation. The model is sufficiently general to capture several important features of employment relationships. Firm-specific human capital is accumulated with experience. The information structure we assume reflects the plausible hypotheses that employers know more than workers about the latter's contribution to the firms' output, that outside employers know less than the current employer about a worker's ability, and that outside employers can observe a worker's position in the hierarchy.

Our main result is that firms choose an endogenous hierarchical structure with a finite number of ranks. We also show that delayering occurs as a result of labour market

[11] In technical terms, defining the equilibrium correspondence as a correspondence from the set of possible values of s, viz. the non-negative real line, into the set of the subsets of $[\underline{\mu},\bar{\mu}]$, the argument in the text implies that this correspondence is not lower semicontinuous at $s=0$.

competition: when own and outside workers become better substitutes, the hierarchy has fewer ranks. Although the model is quite rich, we do leave out many other important aspects of employment relationships. In particular our model gives workers an exceedingly passive role: they simply do as they are told. This passive role is clearly unrealistic, and further analysis of the interaction between labour market competition and hierarchies should try to incorporate strategic behaviour on the workers' part. However, the basic mechanism which underlies our findings, that firms incur the efficiency cost of creating hierarchies in order to hide their employees' ability, seems likely to continue to operate in a more general set-up where workers have some private information and use it strategically.

Acknowledgements

I would like to thank John Moore, Peter Simmons, Pierre de Trenqualye and two referees of this journal for helpful comments on earlier drafts. Part of this research was carried out while the author was visiting Chuo University, Tokyo and Sonderforschungs-bereich 303 at the University of Bonn. Preliminary versions of the paper were presented at seminars in York, Manchester, Milan (Catholic University), Edinburgh, Nottingham and Warwick.

Appendix A. Proofs of the results

Proof of Proposition 1: The employer's problem, for a worker of known skill μ, is the choice of a wage-effort pair which solves her profit maximisation problem, subject to the worker not choosing to become unemployed, and no outside employer being able to obtain a positive offer from a worker whose skill is known to be μ:

$$\max_{t,e}(1+s)y(\mu,e)-t$$

s.t.:

$$t-v(e)\geq 0$$

$$y(\mu,e^{0}_{\{\mu\}})-v\leq 0$$

The first constraint ensures that the worker prefers to be employed by the current employer, and the second that the value of the worker to an outside employer is non-positive. Ignore the second constraint to obtain $e=e^{s}_{\{\mu\}}$ and wage $t=v(e^{s}_{\{\mu\}})$. To end the proof simply note that the condition $y(\bar{\mu},e^{0}_{\{\bar{\mu}\}})-v(e^{0}_{\{\bar{\mu}\}})\leq 0$ simply says that when $\mu=\bar{\mu}$, the second constraint is satisfied. Since $y(\mu,e^{0}_{\{\mu\}})-v(e^{0}_{\{\mu\}})$ is increasing in μ, the second constraint is satisfied for every $\mu\in[\underline{\mu},\bar{\mu}]$. Finally, it must be the case that the current employer does prefer to employ all her workers. This is ensured by the condition $(1+s^{0})y(\underline{\mu},e^{s^{0}}_{\{\underline{\mu}\}})-v(e^{s^{0}}_{\{\underline{\mu}\}})=0$, which states that it is worth the employer's while to employ the least able workers, and again by the fact that $(1+s^0)y(\mu,e^{s^0}_{\{\mu\}})-v(e^{s^0}_{\{\mu\}})$ is increasing in μ. This completes the proof. □

Proof of Proposition 2: We begin by establishing the following lemma.

Lemma 0: *For any beliefs held by the market satisfying Assumption 4, the equilibrium wage schedule $\omega(\mu)$ and the equilibrium effort schedule $\eta(\mu)$ are non-decreasing functions, except at most for a subset of measure zero.*

Proof: Suppose, by contradiction, that the salary schedule is decreasing: there are positive measure sets $M_{i_1},M_{i_2}\subseteq[\underline{\mu},\bar{\mu}]$, $\mu_1\in\mathrm{int}\,M_{i_1}$, and $\mu_2\in\mathrm{int}\,M_{i_2}$ with $\mu_1<\mu_2$, satisfying the contradiction hypothesis, $w_1>w_2$. Then there exists $\varepsilon>0$ such that it is possible to take open neighbourhoods $M_j\subseteq M_{i_j}$, with $\mu_j\in M_j$, and $L(M_j)=\varepsilon$, $j=1,2$. That is, there are ε workers both in M_1 and M_2. Now, let the firm "swap" the contracts offered to workers in these two sets. Instead of the original contract $\{(w_i,e_i,M_i)\}_{i\in I}$, let the firm offer the contract $\{(w_{i_1},e_{i_1},M_{i_1}\setminus M_1\cup M_2),(w_{i_2},e_{i_2},M_{i_2}\setminus M_2\cup M_1),\{(w_i,e_i,M_i)\}_{i\in I\setminus\{i_1,i_2\}}\}$. This is observationally equivalent to $\{(w_i,e_i,M_i)\}_{i\in I}$, and therefore the market's beliefs cannot be changed by the swap (irrespective of Assumption 4). Because $\{(w_i,e_i,M_i)\}_{i\in I}$ is optimal, this swap cannot improve profit:

$$\begin{aligned}&\int_{\mu\in M_1}(1+s)y(\mu,e_1)f(\mu)\mathrm{d}\mu-\varepsilon w_1+\int_{\mu\in M_2}(1+s)y(\mu,e_2)f(\mu)\mathrm{d}\mu-\varepsilon w_2\\&\quad\geq\int_{\mu\in M_1}(1+s)y(\mu,e_2)f(\mu)\mathrm{d}\mu-\varepsilon w_2+\int_{\mu\in M_2}(1+s)y(\mu,e_1)f(\mu)\mathrm{d}\mu-\varepsilon w_1\end{aligned}\tag{1}$$

or

$$\int_{\mu\in M_1}[y(\mu,e_1)-y(\mu,e_2)]f(\mu)\mathrm{d}\mu-\int_{\mu\in M_2}[y(\mu,e_1)-y(\mu,e_2)]f(\mu)\mathrm{d}\mu\geq 0$$

$$\begin{aligned}&\int_{\mu\in M_1}\int_{e_2}^{e_1}y_e(\mu,e)\mathrm{d}ef(\mu)\mathrm{d}\mu-\int_{\mu\in M_2}\int_{e_2}^{e_1}y_e(\mu,e)\mathrm{d}ef(\mu)\mathrm{d}\mu\\&\quad=\int_{e_2}^{e_1}\left[\int_{\mu\in M_1}y_e(\mu,e)f(\mu)\mathrm{d}\mu-\int_{\mu\in M_2}y_e(\mu,e)f(\mu)\mathrm{d}\mu\right]\mathrm{d}e\geq 0\end{aligned}$$

by the mean value theorem, there exists $\tilde{\mu}_i\in M_i$, $i=1,2$, such that the above can be written:

$$\int_{e_2}^{e_1}[y_e(\tilde{\mu}_1,e)-y_e(\tilde{\mu}_2,e)]de=\int_{e_2}^{e_1}\int_{\tilde{\mu}_2}^{\tilde{\mu}_1}y_{e\mu}(\mu,e)\mathrm{d}\mu de\geq 0\tag{2}$$

since $y_{e\mu}(.)>0$ and $\tilde{\mu}_2>\tilde{\mu}_1$, Eq. (2) implies $e_1\leq e_2$. Therefore, we have $w_1>w_2$ and $e_1\leq e_2$: workers in M_{i_1} are paid more and exert less effort than workers in M_{i_2} and therefore they are strictly better-off. Let the firm shade the wage offer to workers in M_{i_1} to $w_1-\delta>w_2$. Moreover, let δ be small enough so that the order of w_1, relative to all other wage levels, is not changed. Then, by Assumption 4, the market beliefs are not changed by this deviation, and workers in M_{i_1} continue to be better-off than the unemployment utility and the firm's profit is higher. Therefore, the proposed strategy cannot be an equilibrium. The contradiction shows that the wage schedule must be non-decreasing.The argument used in the previous paragraph also shows that effort must also be non-decreasing and establishes Lemma 0. □

Lemma 0 implies that a firm does not manipulate the markets beliefs about the ranking of the workers : the market knows that if a worker has a higher wage than another, then he must also be abler. This implies $r_i = M_i$ for every $i \in I$. Thus, for every $M_i \subseteq [\underline{\mu}, \bar{\mu}]$ with $L(M_i) > 0$, the current employer will

$$\max_{w_i, e_i} \left\{ \frac{1}{L(M_i)} \int_{\mu \in M_i} (1+s) y(\mu, e_i) f(\mu) \mathrm{d}\mu - w_i \right\} \tag{3}$$

s.t.:

$$w_i - v(e_i) \geq \frac{1}{L(M_i)} \int_{\mu \in M_i} y(\mu, e^0_{M_i}) f(\mu) \mathrm{d}\mu - v(e^0_{M_i}) \tag{4}$$

and

$$w_i - v(e_i) \geq 0 \tag{5}$$

The RHS in Eq. (4) is the expected utility that a worker receives from the outside market, given by his wage, reduced by the cost of effort which the worker will be required to exert. Constraint (5) ensures that the workers do not wish to retire and enjoy reservation utility. Ignore Eq. (5) and substitute w_i from Eq. (4) to get:

$$\max_{e_i} \left\{ \frac{1}{L(M_i)} \int_{\mu \in M_i} \{(1+s) y(\mu, e_i) - v(e_i) - y(\mu, e^0_{M_i}) + v(e^0_{M_i})\} f(\mu) \mathrm{d}\mu \right\} \tag{6}$$

This gives $e_i = e^s_{M_i}$, and w_i is obtained immediately, establishing the last parts of Proposition 2.

We now continue with the proof of Proposition 2, by establishing that the wage and effort schedules are step functions. Define $p(\mu,e,s) = [(1+s)y(\mu,e) - v(e)] f(\mu)$ and suppose the current employer had no ranks but a strictly increasing function, fully separating in $[\underline{\mu}, \bar{\mu}]$. Her profit in this case would be $\pi^s = \int_{\mu \in [\underline{\mu}, \bar{\mu}]} p(\mu, e^s_{\{\mu\}}, s) \mathrm{d}\mu - \int_{\mu \in [\underline{\mu}, \bar{\mu}]} p(\mu, e^0_{\{\mu\}}, 0) \mathrm{d}\mu$. Let $\left\{ \{\mu^t_i\}^{n_t}_{i=0} \right\}^{\infty}_{t=1}$ be a sequence of divisions of $[\underline{\mu}, \bar{\mu}]$, with $\{\mu^t_i\}^{n_t}_{i=0} = \{\mu^t_0, \mu^t_1, \ldots, \mu^t_{n_t}\}$, $\underline{\mu} = \mu^t_0 < \mu^t_1 < \ldots < \mu^t_{n_t} = \bar{\mu}$ and $\lim_{t \to \infty} \left(\max_{i=1,\cdots,n_t} \{L([\mu_{i-1}, \mu_i])\} \right) = 0$. We have:

$$\pi^s = \lim_{t \to \infty} \sum_{i=1}^{n_t} \left\{ \int_{\mu \in [\mu_{i-1}, \mu_i]} p(\mu, e^s_{[\mu_{i-1}, \mu_i]}, s) \mathrm{d}\mu - \int_{\mu \in [\mu_{i-1}, \mu_i]} p(\mu, e^0_{[\mu_{i-1}, \mu_i]}, 0) \mathrm{d}\mu \right\} \tag{7}$$

That is, as the ranks shrink, they tend to separating schedule. Let ε be the minimum interval length at the profit maximising hierarchical structure. We now show[12] that $\varepsilon>0$. This shows that the optimal hierarchy has a finite number of ranks, and, together with Eq. (7), establishes that a hierarchy is better than a fully separating equilibrium. To show that $\varepsilon>0$, we show that the average profit increases with ε for ε sufficiently small. The average profit at a generic interval of length ε is:

$$\pi([\mu_i,\mu_i+\varepsilon]) = \int_{\mu_i}^{\mu_i+\varepsilon} \frac{1}{\varepsilon}[p(\mu, e^{s}_{[\mu_i,\mu_i+\varepsilon]}, s) - p(\mu, e^{0}_{[\mu_i,\mu_i+\varepsilon]}, 0)]\mathrm{d}\mu$$

by the mean value theorem for integrals, there exists s_i such that:

$$\pi([\mu_i,\mu_i+\varepsilon]) = \int_{\mu_i}^{\mu_i+\varepsilon} \frac{1}{\varepsilon} s_i \frac{\mathrm{d}p(\mu, e^{s_i}_{[\mu_i,\mu_i+\varepsilon]}, s_i)}{\mathrm{d}s}\mathrm{d}\mu$$

where $\mathrm{d}p/\mathrm{d}s$ is given by $y(\mu,e)f(\mu)+[(1+s)y_e(\mu,e)f(\mu)-v'(e)]\mathrm{d}e/\mathrm{d}s$, and hence:

$$\pi([\mu_i,\mu_i+\varepsilon]) = \int_{\mu_i}^{\mu_i+\varepsilon} \frac{s_i}{\varepsilon} y(\mu,e)f(\mu)\mathrm{d}\mu + s_i\left[\int_{\mu_i}^{\mu_i+\varepsilon}\frac{1}{\varepsilon}(1+s)y_e(\mu,e)f(\mu)\mathrm{d}\mu - v'(e)\right]\frac{\mathrm{d}e}{\mathrm{d}s}$$

the term in the square bracket is zero because of the optimality of e in $[\mu_i,\mu_i+\varepsilon]$. Hence: $\pi([\mu_i,\mu_i+\varepsilon]) = \int_{\mu_i}^{\mu_i+\varepsilon}\frac{s_i}{\varepsilon}y(\mu, e^{s_i}_{[\mu_i,\mu_i+\varepsilon]})f(\mu)\mathrm{d}\mu$. Finally:

$$\begin{aligned}\frac{\mathrm{d}\pi([\mu_i,\mu_i+\varepsilon])}{\mathrm{d}\varepsilon} &= \frac{s_i}{\varepsilon^2}\left\{\left[y(\mu_i+\varepsilon, e^{s_i}_{[\mu_i,\mu_i+\varepsilon]})f(\mu_i+\varepsilon) + \int_{\mu_i}^{\mu_i+\varepsilon} y_e(.)\frac{\mathrm{d}e}{\mathrm{d}\varepsilon}f(\mu)\mathrm{d}\mu\right]\varepsilon - \int_{\mu_i}^{\mu_i+\varepsilon} y(.)f(\mu)\mathrm{d}(\mu)\right\} \\ &= \frac{s_i}{\varepsilon}\left\{\int_{\mu_i}^{\mu_i+\varepsilon} y_e(.)\frac{\mathrm{d}e}{\mathrm{d}\varepsilon}f(\mu)\mathrm{d}\mu + \left[y(\mu_i+\varepsilon, e^{s_i}_{[\mu_i,\mu_i+\varepsilon]})f(\mu_i+\varepsilon) - \frac{1}{\varepsilon}\int_{\mu_i}^{\mu_i+\varepsilon} y(.)f(\mu)\mathrm{d}\mu\right]\right\} \end{aligned} \tag{8}$$

[12] From a technical point of view, note that the problem is the choice of a discrete variable, the number of ranks. This discrete variable cannot be treated as a continuous one and, in order to obtain the result that the number of ranks is finite, we show that the shortest rank has a strictly positive length. This technique can potentially be applied to other analogous problems.

The term in the square bracket of Eq. (8) can be written as:

$$\left\{ y(\mu_i+\varepsilon, e^{s_i}_{[\mu_i,\mu_i+\varepsilon]})f(\mu_i+\varepsilon) - y(\mu_i, e^{s_i}_{[\mu_i,\mu_i+\varepsilon]})f(\mu_i) - \frac{1}{\varepsilon}\int_{\mu_i}^{\mu_i+\varepsilon} y(.)f(\mu)\mathrm{d}(\mu) \right\}$$
$$+\, y(\mu_i, e^{s_i}_{[\mu_i,\mu_i+\varepsilon]})f(\mu_i)$$

For ε close to zero, the term in the curly brackets is approximated by zero, and hence the term in the square bracket in Eq. (8) is approximated by $y(\mu_i,e^{s_i}_{[\mu_i,\mu_i+\varepsilon]})f(\mu_i)>0$. The first term in Eq. (8) is positive since $y_e(.)>0$ and, from total differentiation of the first order condition for problem (6):

$$\frac{\mathrm{d}e}{\mathrm{d}\varepsilon} = (1+s)\frac{y_e(\mu_i+\varepsilon,e)f(\mu_i+\varepsilon) - \frac{1}{\varepsilon}\int_{\mu_i}^{\mu_i+\varepsilon} y_e(.)f(\mu)\mathrm{d}\mu}{\varepsilon v''(e) - \int_{\mu_i}^{\mu_i+\varepsilon}(1+s)y_{ee}(.)f(\mu)\mathrm{d}\mu}$$

The denominator is positive as $v''(e)>0$, and $y_{ee}(.)<0$, and the numerator is positive as $y_{e\mu}(.)>0$. Therefore, Eq. (8) is positive, that is $\frac{\mathrm{d}\pi([\mu_i,\mu_i+\varepsilon])}{\mathrm{d}\varepsilon}>0$. This establishes that the shortest interval has a strictly positive length, and hence that the number of ranks is finite. To end the proof of the proposition, we only need to establish that it was a legitimate procedure to ignore the second constraint in problem (3). To this aim, we establish Lemma 1 in the text.

Proof of Lemma 1: The utility of a worker in rank M_i is given by:

$$\frac{1}{L(M_i)}\int_{\mu\in M_i} y(\mu, e^0_{M_i})f(\mu)\mathrm{d}\mu - v(e^0_{M_i})$$

$e^0_{M_i}$ can be written as: $e^0_{\{\tilde{\mu}_i\}}$ for some $\tilde{\mu}_i \in M_i$. Differentiation with respect to $\tilde{\mu}_i$ yields:

$$\left[\frac{1}{L(M_i)}\int_{\mu\in M_i} y_e(\mu, e^0_{M_i})f(\mu)\mathrm{d}\mu - v'\,(e^0_{M_i})\right]\frac{\mathrm{d}e^0_{\{\tilde{\mu}_i\}}}{\mathrm{d}\tilde{\mu}_i} + \frac{1}{L(M_i)}\int_{\mu\in M_i} y_\mu(\mu, e^0_{M_i})f(\mu)\mathrm{d}\mu$$

the term in the square bracket is zero because of the optimality of $e^0_{M_i}$, and the second is positive. This shows that higher $\tilde{\mu}_i$ implies higher utility. This establishes the lemma. □

Therefore, the lowest utility is obtained by the workers in the first rank, and their utility is bounded below by $y(\bar{\mu},e^0_{\{\bar{\mu}\}}) - v(e^0_{\{\bar{\mu}\}})$, which is the limit value as the first interval tends to $\{\bar{\mu}\}$. This completes the proof of Proposition 2. □

Proof of Proposition 3: The proof proceed by induction on n_H.

(i) in the first step, we assume that $n_H=1$ and show that $n_L \leq n_H$ (from which, clearly, $n_L=1$).

(ii) in the second step, we assume to have established that if $n_H=j$, $j=1,\ldots,n-1$, then $n_L \leq n_H$ and, taking $n_H=n$, we show that $n_L \leq n$.

Consider the first step. Given s, let π^1 be the employer's profit level when she has one rank and $\pi^2(\hat{\mu})$ the profit when she has two ranks, $[\underline{\mu},\hat{\mu}]$ and $[\hat{\mu},\underline{\mu}]$.

$$\pi^1 = \int_{\mu\in[\underline{\mu},\bar{\mu}]} [p(\mu, e^s_{[\underline{\mu},\bar{\mu}]}, s) - p(\mu, e^0_{[\underline{\mu},\bar{\mu}]}, 0)]\mathrm{d}\mu$$

$$\pi^2(\hat{\mu}) = \int_{\mu\in[\underline{\mu},\hat{\mu}]} [p(\mu, e^s_{[\underline{\mu},\hat{\mu}]}, s) - p(\mu, e^0_{[\underline{\mu},\hat{\mu}]}, 0)]d\mu + \int_{\mu\in[\hat{\mu},\bar{\mu}]} [p(\mu, e^s_{[\hat{\mu},\bar{\mu}]}, s) - p(\mu, e^0_{[\hat{\mu},\bar{\mu}]}, 0)]\mathrm{d}\mu$$

Note that $\pi^1 = \pi^2(\underline{\mu}) = \pi^2(\bar{\mu})$ and therefore that π^2 has an interior stationary point. Because, by assumption, when $s=s_H$, the optimal number of ranks is $n_H=1$, the highest value of $\pi^2(\hat{\mu})$ at a stationary point is less than the value at the endpoints $\underline{\mu}$ and $\bar{\mu}$, (otherwise the optimal hierarchy for $s=s_H$ would have at least two ranks. To establish (i), all we need to show is that the difference $\pi^2(\hat{\mu}) - \pi^1$ increases with s, at the stationary point for $\pi^2(\hat{\mu})$. This shows that, as s decreases from s_H to s_L, the profit with two ranks, $\pi^2(\hat{\mu})$, cannot become greater than the profit with one rank, π^1, and establishes the first step of the induction process. Differentiate $\pi^2(\hat{\mu}) - \pi^1$ with respect to s to get:

$$\frac{\mathrm{d}(\pi^2(\hat{\mu}) - \pi^1)}{\mathrm{d}s} = \int_{\underline{\mu}}^{\hat{\mu}(s)} y(\mu, e^s_{[\underline{\mu},\hat{\mu}]})f(\mu)\mathrm{d}\mu + \int_{\hat{\mu}(s)}^{\bar{\mu}} y(\mu, e^s_{[\hat{\mu},\bar{\mu}]})f(\mu)\mathrm{d}\mu - \int_{\underline{\mu}}^{\bar{\mu}} y(\mu, e^s_{[\underline{\mu},\bar{\mu}]})f(\mu)\mathrm{d}\mu \tag{9}$$

Where $\hat{\mu}(s)$ is the stationary point of $\pi^2(\hat{\mu})$ with the highest value in the set of the stationary points (non necessarily a global maximum, since this could be reached at the extremes of the interval). The above is obtained as the terms in de/ds and those in d$\hat{\mu}$/ds all cancel out by the envelope theorem. Now notice that we have:

$$\int_{\underline{\mu}}^{\hat{\mu}(s)} y(\mu, e^s_{[\underline{\mu},\hat{\mu}]})f(\mu)\mathrm{d}\mu - L([\underline{\mu},\hat{\mu}])v(e^s_{[\underline{\mu},\hat{\mu}]}) + \int_{\hat{\mu}(s)}^{\bar{\mu}} y(\mu, e^s_{[\hat{\mu},\bar{\mu}]})f(\mu)\mathrm{d}\mu - L([\hat{\mu},\bar{\mu}])v(e^s_{[\hat{\mu},\bar{\mu}]}) \geq \int_{\underline{\mu}}^{\bar{\mu}} y(\mu, e^s_{[\underline{\mu},\bar{\mu}]})f(\mu)\mathrm{d}\mu - L([\underline{\mu},\bar{\mu}])v(e^s_{[\underline{\mu},\bar{\mu}]})$$

Since the RHS is the value of the maximand in e_1 and e_2 on the LHS with the additional constraint $e_1=e_2$. Substitute into Eq. (9):

$$\frac{\mathrm{d}(\pi^2(\hat{\mu}(s)) - \pi^1)}{\mathrm{d}s} \geq L([\underline{\mu},\hat{\mu}])v(e^s_{[\underline{\mu},\hat{\mu}]}) + L([\hat{\mu},\bar{\mu}])v(e^s_{[\hat{\mu},\bar{\mu}]}) - L([\underline{\mu},\bar{\mu}])v(e^s_{[\underline{\mu},\bar{\mu}]}) > 0$$

The last equality follows from $e^{s}_{[\underline{\mu},\hat{\mu}]}<e^{s}_{[\underline{\mu},\bar{\mu}]}<e^{s}_{[\hat{\mu},\bar{\mu}]}$ and $v_{ee}(.)>0$. This establishes the first step of the induction process.

Next, consider the second step. With no loss in generality, assume by contradiction that $n_L=n+1$. Begin by noticing that if the optimal hierarchy has a cut-off at $\hat{\mu}$, then it is identical to the outcome obtained by separately optimising in $[\underline{\mu},\hat{\mu}]$ and in $[\hat{\mu},\mu]$. If we had $\mu_L^i=\mu_H^j$ for some i and j, then a contradiction would be obtained since either in $[\underline{\mu},\mu_H^j]$ or in $[\mu_H^j,\bar{\mu}]$ the subdivision with $s=s_H$ has less than n ranks and fewer than the subdivision with $s=s_L$. Therefore, $\mu_L^i\neq\mu_H^j$ for every i and j. Assume $\mu_L^1<\mu_H^1<\mu_L^2$. Let $\underline{\mu}<\mu_H^1<\mu_H^2<\ldots<\mu_H^n=\bar{\mu}$ be the optimal division for $s=s_H$; and let $\underline{\mu}<\mu_L^1<\mu_L^2<\ldots<\mu_L^n<\mu_L^{n+1}=\bar{\mu}$ be the optimal division for $s=s_L$, with e_L^2 and w_L^2 the optimal effort and wage for workers in $[\underline{\mu},\mu_L^1]$. Notice that, when $s=s_L$, the profit obtained in the optimal n-rank hierarchy, call it $\pi^*(n,[\underline{\mu},\bar{\mu}])$, is at least as large as the profit obtained as follows

$$\pi^*(1,[\bar{\mu},\mu_H^1])+\pi^*(n-1,[\mu_H^1,\bar{\mu}]) \tag{10}$$

that is the profit with the (trivial) optimal one-rank hierarchy in $[\underline{\mu},\mu_H^1]$ plus the profit with the optimal $(n-1)$-rank hierarchy in $[\mu_H^1,\bar{\mu}]$. Now consider the first of the two terms in Eq. (10):

$$\pi^*(1,[\underline{\mu},\mu_H^1]) \geq \begin{array}{l}\text{profit obtained}\\ \text{from interval}\\ [\bar{\mu},\mu_H^1]\text{ with the}\\ \text{optimal two-rank}\\ \text{division}\end{array} \geq \begin{array}{l}\text{profit obtained from}\\ [\bar{\mu},\mu_H^1]\text{ with the}\\ \text{sub-optimal}\\ \text{two-rank division}\\ \{\bar{\mu},\mu_L^1,\mu_H^1\}\text{ and wage}\\ \text{and effort optimally}\\ \text{adjusted in each rank}\end{array} \geq \begin{array}{l}\text{profit obtained}\\ \text{from }[\bar{\mu},\mu_H^1]\text{ with}\\ \text{the division}\\ \{\underline{\mu},\mu_L^1,\mu_H^1\}\text{ and}\\ \text{wage and effort}\\ \text{exogenously set}\\ \text{at }w_{\mathrm{L}}^2\text{ and }e_{\mathrm{L}}^2\end{array}$$

The first inequality follows from the first step of the induction process, the second from the fact that the optimal division in any given interval must give more profit than any other, and the third from the fact that the optimal wage effort schedule must give more profit than any other exogenous schedule. Analogously for the second term in Eq. (10):

$$\pi^*(n-1,[\mu_H^1,\bar{\mu}]) \geq \begin{array}{l}\text{profit obtained}\\ \text{from interval}\\ [\mu_H^1,\bar{\mu}]\text{ with the}\\ \text{optimal } n-\text{rank}\\ \text{division}\end{array} \geq \begin{array}{l}\text{profit obtained}\\ \text{from }[\mu_H^1,\bar{\mu}]\text{ with}\\ \text{the sub-optimal}\\ \text{n-rank division}\\ \{\mu_H^1,\mu_L^2,\mu_L^3,\cdots,\\ \mu_L^n,\bar{\mu}\}\text{and wage}\\ \text{and effort optimally}\\ \text{adjusted in each}\\ \text{rank}\end{array} \geq \begin{array}{l}\text{profit obtained}\\ \text{from }[\mu_H^1,\bar{\mu}]\\ \text{with the division}\\ \{\mu_H^1,\mu_L^2,\mu_L^3,\cdots,\\ \mu_L^n,\bar{\mu}\}\text{and wage}\\ \text{and effort}\\ \text{exogenously}\\ \text{set at }w_L^2\text{ and }e_L^2\\ \text{in }[\mu_H^1,\mu_L^2]\end{array}$$

The first inequality follows from the induction hypothesis, the other analogously to the first chain of inequalities. Finally note that the profit obtained with the $(n+1)$-rank division in $[\underline{\mu},\bar{\mu}]$ with $s=s_L$ is given by the sum of the last two items in each of the above chains of inequalities, and therefore it cannot exceed $\pi^*(n,[\underline{\mu},\bar{\mu}])$. This contradiction establishes the proposition for $\mu_L^1 < \mu_H^1 < \mu_L^2$. Other cases are treated analogously. This completes the proof. □

References

Bernhardt, D., 1995. Strategic promotion and compensation. Review of Economic Studies 62, 315–339.

Calvo, G.A., Wellisz, S., 1978. Supervision, loss of control and the optimal size of the firm. Journal of Political Economy 86, 943–952.

Calvo, G.A., Wellisz, S., 1979. Hierarchy, ability and income distribution. Journal of Political Economy 87, 991–1010.

Carmichael, H.L., 1984. Reputations in the labor market. American Economic Review 74, 713–725.

De Fraja, G., 1995. Hierarchies in Organisations and Labour Market Competition. University of York, York, UK. Discussion Paper 95/9.

Doeringer, P., Piore, M., 1971. Internal Labor Markets and Manpower Analysis. D.C. Heath, Lexington, MA.

Dore, R., 1973. British Factory, Japanese Factory: The Origins of National Diversity in Industrial Relations. Allen and Unwin, London.

Gibbons, R., 1998. Incentives in organizations. Journal of Economic Perspectives 12/4, 115–132.

Jacoby, S.M., 1984. The development of internal labor markets in American manufacturing firms. In: Osterman, P. (Ed.), Internal Labour Markets. MIT Press, Cambridge, MA, pp. 23–69.

Kanter, R.M., 1984. Variations in managerial career structures in high-technology firms: the impact of organizational characteristics on internal labour market patterns. In: Osterman, P. (Ed.), Internal Labour Markets. MIT Press, Cambridge, MA, pp. 109–131.

Laffont, J.J., Tirole, J., 1993. A Theory of Incentives in Procurement and Regulation. MIT Press, Cambridge, MA.

Littler, C.R., Wiesner, R., Dunford, R., 2003. The dynamics of delayering: changing management structures in three countries. Journal of Management Studies 40, 225–256.

MacLeod, W.B., Malcomson, J.M., 1988. Reputation and hierarchies in dynamic models of employment. Journal of Political Economy 96, 832–854.

Milgrom, P., Oster, S., 1987. Job discrimination, market forces and the invisibility hypothesis. Quarterly Journal of Economics 102, 453–476.

Milgrom, P., Roberts, J., 1992. Information, Economics and Management. Prentice-Hall, Englewood Cliffs, NJ.

Mirrlees, J., 1976. The optimal structure of incentives and authority within an organization. Bell Journal of Economics 7, 105–131.

Ricart i Costa, J., 1988. Managerial task assignments and promotions. Econometrica 56, 449–466.

Rosen, S., 1982. Authority, control, and the distribution of earnings. Bell Journal of Economics 13, 311–323.

Rubinstein, A., 1982. Perfect equilibrium in a bargaining model. Econometrica 50, 863–894.

Waldman, M., 1984a. Worker allocation, hierarchies and the wage distribution. Review of Economic Studies 51, 95–109.

Waldman, M., 1984b. Job assignments, signaling, and efficiency. Rand Journal of Economics 15, 255–267.

Williamson, O.E., 1967. Hierarchical control and optimum firm size. Journal of Political Economy 75, 123–138.