NBER WORKING PAPER SERIES

SENIORITY

DongIk Kang
Miles S. Kimball

Working Paper 31563
http://www.nber.org/papers/w31563

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
August 2023

Seniority
DongIk Kang and Miles S. Kimball
NBER Working Paper No. 31563
August 2023
JEL No. J0,J31,J32

ABSTRACT

This paper studies the optimal wage structure of a firm with imperfect monitoring of worker effort. We find that when firms can commit to (implicit) long-term contracts, imperfect monitoring leads to optimal wage profiles that reflect worker seniority. We provide a precise measure of seniority as a ratio of co-state variables and illustrate how this measure of seniority evolves over the worker's tenure with the firm and how it affects wages, effort, monitoring intensity and separation rates. We also show how earnings loss from unemployment reflects seniority and how optimal monitoring intensity, amenities and on-the-job training evolve with seniority.

DongIk Kang
336 Sicheong-daero
Sejong-si
South Korea
dikang@kipf.re.kr

Miles S. Kimball
Economics Building
University of Colorado Boulder
261 UCB
#212
Boulder, CO 80302
and NBER
miles.kimball@colorado.edu

1 Introduction

In many workplaces, a worker's wage increases as they gain tenure with the firm. A growing list of studies suggest that this reflects not only general experience but also tenure with a specific firm. In addition, these studies show that the effect of worker seniority on wages may be quite large.[1] Recent research has focused mostly on firm-specific human capital to explain tenure effects on the wage profile of workers. However, evidence from studies that relate firm characteristics to wage profiles suggest that deferred compensation schemes, designed to motivate worker effort, also play an important role in the determination of seniority wage structures.[2]

Building on the pioneering work of Lazear(1979, 1981) and others,[3] we model firms as deferring compensation in order to motivate worker effort; effort is imperfectly observable[4] and firms hire workers by offering them implicit long-term contracts. Unlike Lazear(1979, 1981), in which the exact path of wage profiles is indeterminate as long as it provides enough motivation for workers,[5] the model presented here provides a determinate path of wages at

[1]While results from earlier research on returns from tenure are mixed, (for a detailed discussion, see Altonji and Shakotko (1987), Abraham and Farber (1987), Hutchens(1989), Topel (1991), Neal (1995), and Altonji and Williams (2005) among others), recent research shows large tenure effects on wages. See Stevens (2003), Dustmann and Meghir (2005), Buchinsky et al. (2010), Pavan (2011), Zwick (2011), Kwon and Milgrom (2014), and Buhai et al. (2014) among others.

[2]For example, Zwick (2011) finds that firms with steeper seniority wage profiles have lower initial wages and that the correlation between the initial wage level and the gradient is weakly negative or insignificant. Daniel and Heywood (2007) find that firms with pensions and steep wage profiles hire fewer older workers. Their findings support deferred compensation as an important determinant of seniority wage structures. Other papers that find evidence of deferred compensation include Barth (1997) and Heywood et al. (2010).

[3]See, for example, Shapiro and Stiglitz (1984), Bulow and Summers (1986), and Akerlof and Katz (1989) for notable efficiency wage models. Studies such as Krueger and Summers (1988), Cappelli and Chauvin (1991), Blackburn and Neumark (1992) and somewhat more recently, Kugler (2003) and Huck et al. (2011) provide empirical evidence for the presence of efficiency wages.

[4]"Imperfectly observable" could be weakened to "imperfectly verifiable with constraints on what can be done without verification."

[5]Lazear(1979, 1981) allows for explicit bonding: workers posting upfront bonds to the firm that may later be confiscated by the firm. This assumption has been criticized on the grounds that it is seldom observed in the real economy. Akerlof and Katz (1989) argue that when one does not allow explicit bonding, firms should pay higher-than-market wages. They claim that this dominates deferred payment. However, we find that when the effort choice of workers is continuous and not binary, upward sloping wage profiles will still be used by firms to motivate workers.

every point in time. This is because we allow workers to make a continuous effort choice at every instant in time instead of the binary (work or shirk) effort decision typical of the literature.[6] In our model, the wage goes up gradually with seniority. However, seniority in our framework reflects more than just the amount of time employed at the firm: it reflects the value of a worker's future utility to the firm. A worker's future utility is valuable to the firm because it motivates effort, which raises the worker's productivity.

We extend this setup to include the interaction of seniority with other aspects of the firm-worker relationship. First, the seniority wage structure has interesting implications for earnings losses from job loss. Previous research has generally found large and persistent losses in wages that are difficult to reconcile with existing theories. Here, we show that seniority wage structures provide a channel through which a job loss results in large earnings losses for a worker, as well as higher separation rates from subsequent jobs. Second, we consider the choices of firms with variable monitoring intensity and give conditions for optimal monitoring intensity to fall with seniority. Third, we consider the case of firms that provide various amenities to workers in addition to wages. Additively separable amenities behave much like wages, but amenities that make effort easier behave in a more complex way. Finally, we consider on-the-job training; we show that despite firm-specific human capital being subject to depreciation, optimal training expenditures will make firm-specific human capital continually rise with seniority.

We aim in this paper to fill an important void in the debate over seniority wages. By providing a well-defined measure of the importance of seniority and predictions about the

[6] Dealing with a continuous effort choice is made possible in part by the solution technique we apply to solve the agents' problems. The agents in our model solve an optimization problem where the expectations of other agents are a forward-looking state variable. The structure of this problem makes it difficult to utilize the usual techniques of dynamic optimization. The "retrograde" solution technique for control theory problems, which is underappreciated in economics, allows us to reformulate the problem so that standard optimal control methods can be used. The retrograde approach is also valuable for other situations that have a forward-looking variable as a state variable, such as in studying monetary policy and studying capital taxation. We consider the problem we solve a good example for learning the retrograde approach.

interaction of seniority with other aspects of the firm-worker relationship, our framework provides important insight into seniority wages and helps show a path toward testing competing theories about firm-worker relationships.[7] This paper is purely theoretical; a single paper that treated *both* theory and empirics in this area as they should be treated would be much too long. But we view a theory that makes the evolution of the wage with the accumulation of seniority determinate as a big step toward empirical implementation of the theory of deferred compensation.

Understanding the determinants of the empirically increasing tenure-wage relationship is related to a long-standing question in economics: whether spot wages always track productivity. In our model, wages do not directly reflect the firm-specific current productivity of the worker; they reflect the productivity of a worker over the worker's entire expected tenure with the firm: past, present and future. Failure of equality between current marginal revenue product and the wage has important implications for labor economics but also macroeconomics, since it implies that the observed wage may not be the allocative wage.[8]

In addition to providing insight into seniority wages, this paper can be seen as contributing to a revival and deepening of the study of efficiency wage models. Although efficiency wage models once provided much insight into involuntary unemployment (Shapiro and Stiglitz, 1984) and industrial policies (Bulow and Summers, 1986), notable recent research on the topic is much scarcer. We view the decline in prominence of efficiency wage research as unfortunate. Efficiency wage models have the potential to provide important new insights

[7] Different theories of seniority wages—as well as empirical studies designed around these theories—reach different conclusions about the relationship between productivity and wages. The deferred compensation view of seniority wages suggests that the wage does not necessarily coincide with productivity at a given moment and that increases in wages are larger than increases in productivity. Theories of firm-specific human capital accumulation argue that seniority wages reflect increased productivity of workers. (Recent efforts in theoretical research that focus on firm-specific human capital accumulation leading to seniority wage structures include Stevens (2003), Lazear (2009), and Burdett and Coles (2010).)

[8] Oversimplifying, an "allocative wage" is a wage that determines the quantity of labor demanded and the quantity of labor supplied. In an efficiency wage model, the observed wage does neither.

into the sources and consequences of unemployment. The current discourse on unemployment and its fluctuations is dominated by the search and matching framework; a qualitative and quantitative exploration into motivational unemployment may enhance our understanding of unemployment and its fluctuations.

This paper also belongs to a broader literature that studies the implications of variable worker effort. Research has shown that accounting for variable effort is critical to understanding various economic phenomena. Basu and Kimball (1997) and Basu, Fernald and Kimball (2006) demonstrate that accounting for variable effort is critical in the measurement of technical change. Strobl and Walsh (2007) argue that accounting for the role of variable effort is central to understanding the relationship between monitoring and wages. Epstein and Kimball (2014) puts forth a new theory examining worker effort and disutility of work to study why working hours remains constant in the long-run. Various papers have also attempted to tease out the factors that determine effort. For example, Zivin and Neidell (2012) study the interaction between worker effort and pollution for agricultural workers with piece rate contracts. In a controlled setting, Abeler et al. (2011) find that a person's expectation about their affects the amount of effort they provide. Dellavigna and Pope (2018) find that monetary incentives are most effective, behavior and psychological incentives are somewhat but comparatively less effective in eliciting effort. The focus on asymmetric information distinguishes the efficiency wage literature from the rest of the literature on variable effort. Nevertheless, the theme of variable effort creates a strong bond between the efficiency wage literature and these other papers.

The paper is organized as follows. In Section 2, we present and solve the worker's and the firm's problems. Section 3 analyzes the optimal contract graphically. Section 4 considers extensions and additional applications of the framework. Section 5 concludes. (Appendix A discusses the steady state equilibrium.)

2 Model

We present a dynamic model in which firms hire workers in a competitive labor market by offering workers (implicit) long-term contracts. The model is presented in continuous time and is set in a partial equilibrium framework where interest rates are exogenous, capital is fixed and output is otherwise a function of effective labor input alone. All workers and firms are ex-ante identical.

A firm's implicit contract with the worker specifies a wage path and dismissal policy, both of which are allowed to depend on macroeconomic conditions. A worker supplies a fixed number of hours per unit time. However, workers are able to choose how much effort they exert at every instant in time. This effort level is only partially observable by the firm. Firms can only observe serious blunders by the worker, the frequency of which depends on the worker's effort level.[9]

Because firms cannot observe effort directly, they motivate effort by punishing blunders through dismissal. The worker's effort level z and the firm's monitoring intensity m jointly determine the instantaneous probability of dismissal $m(1-z)$. Together with the exogenous quit rate q this makes up the attrition rate $a(m,z) = q + m(1-z)$. The expression for the probability of dismissal $m(1-z)$ implicitly assumes that the probability of making an observable blunder is linear in effort.

In theory, there may be alternative methods of punishing workers who make serious mistakes. One way to punish workers that has been explored in the literature is to force them to post a performance bond when they are hired and to confiscate this bond when they make blunders. However explicit bond posting is seldom observed in actual labor markets. In this paper, the assumptions we make regarding capital market imperfections prevent explicit

[9] Alternatively, only serious blunders are verifiable by a third party, and the reputational consequences of firing a worker for an unverifiable reason are very costly.

bonding. We assume that workers cannot borrow or lend in financial markets. The inability to borrow at reasonable rates is plausible in a world in which workers have no collateral, and the danger of losing their job because of low effort creates moral hazard problems for lenders worried about bankruptcy. Being unable to save is less realistic, but workers who are highly impatient will not want to do much saving. As we discuss in Section 2.4, allowing workers to save would complicate the model, but would not alter the qualitative implications of this paper, unless it leads to a system with performance bonds.

2.1 The Worker's Problem

A worker chooses a path of planned effort $\{z_t\}$ to maximize life-time expected discounted utility. Workers receive utility $u(w)$ from consumption (equal at all times to the wage), disutility $v(z)$ from effort, and an exogenously determined reservation utility $\bar{u}$ when unemployed. Thus, the worker's flow utility is $u(w) - v(z)$ when employed and $\bar{u}$ when unemployed.

For the worker's utility from wage $u(w)$, we assume $u'(w) > 0, u''(w) < 0$, and the Inada conditions $\lim_{w \to 0} u'(w) = +\infty$ and $\lim_{w \to \infty} u'(w) = 0$ on the domain of non-negative reals. The disutility of effort function $v(z)$ combines elements of worker preferences and the production function. Effort z is measured as the effectiveness of effort in a particular job relative to the effectiveness of maximum effort in that job. Therefore if one job requires more care and attention than another in order to avoid mistakes, this would be expressed as a high level of disutility necessary to achieve a reasonable level of effectiveness z. In a model in which worker preferences and the characteristics of jobs vary, the two elements of $v(z)$ should be analytically separated. However, for the purposes of this paper the distinction is without consequence. We assume that $v'(z) > 0$, $v''(z) > 0$, $v'''(z) > 0$, and also that $\lim_{z \to 1} v(z) = +\infty$, so that the worker always chooses $z < 1$.

Denoting the expected discounted utility of an unemployed worker at time t as $U_u(t)$, the worker chooses a path of effort $\{z_t\}$ to maximize their expected discounted utility at time t_0. The worker solves

$$\max_{\{z_t\}} \quad U_E(t_0) = \int_{t_0}^{\infty} e^{-\int_{t_0}^{t}[\rho + a(m,z_{\tilde{t}})]d\tilde{t}}[u(w_t) - v(z_t) + a(m,z_t)U_u(t)]dt \tag{1}$$

where ρ is the time discount factor of the worker. The path of the wage $\{w_t\}$ and the expected discounted utility of unemployment $\{U_u(t)\}$ are taken as given. At time t, the worker receives flow utility $u(w_t) - v(z_t)$ if the worker remains employed. If the worker is fired they receive expected discounted utility $U_u(t)$ as the total expected discounted utility from everything after time t for an unemployed worker. The instantaneous probability of separation—including being fired—at time t is $a(m, z_t)$. (The mnemonic is a for "attrition.") Finally, the effective discount factor for flow utility while employed reflects not only the worker's utility discount rate ρ but also the probability that the worker remains employed until time t—which is determined by the integral of the separation rate $a(m, z)$.

Solving the maximization problem in (1) is not trivial, most notably due to the fact that the control variable z is present in the discount factor $e^{-\int_{t_0}^{t}(\rho+a(m,z_{\tilde{t}}))d\tilde{t}}$. The problem cannot be solved using the usual optimal control techniques because the control variable affects the present value of all future flow utilities. However, we show that the problem can be solved using standard Hamiltonian methods by reformulating the problem—by viewing the problem as if time were reversed, treating life-time expected discounted utility $U_E(t)$ as a state variable that accumulates backwards in time, and the objective as maximizing this "retrograde" state variable, life-time expected discounted utility, at time t_0. We use the same technique to solve the firm's problem as well. We describe this "retrograde approach" in more detail below.

2.2 The Retrograde Approach

The definition of the word "retrograde" that we have in mind, as defined by the Merriam-Webster dictionary, is "moving, occurring, or performed in a backward direction." The method we propose treats the endogenous variables as if they were in retrograde motion, traveling backwards in time. This perspective allows us to reformulate the problem so it is possible to use standard Hamiltonian methods. Let us demonstrate this method first by using the worker's problem, (1).

The first key insight is that U_E itself can be thought of as a state variable by reformulating the problem as

$$\max_{\{z_t\}} \quad U_E(t_0) \tag{2}$$

$$s.t. \quad U_E(t) = \int_t^{\infty} e^{-\int_t^{t'}[\rho + a(m, z(\tilde{t}))]d\tilde{t}}[u(w_{t'}) - v(z_{t'}) + a(m, z_{t'})U_u(t')]dt' \qquad \forall t \geq t_0,$$

so that the endogenous discounting is no longer a problem. Here, U_E as a *forward-looking* state variable in the maximization problem with its law of motion described by an integral equation. The objective is to maximize the value of U_E at the initial point in time t_0.

The problem still has a different structure than the usual optimization problems in economics. Economists are accustomed to the case of a backward-looking state variable accumulating over time. (In the paradigmatic case, consumption is the control variable and capital is the state variable.) To convert the problem in (2) to this familiar form, Leibniz's rule can be used to convert a forward-looking integral into a differential equation plus an end-point constraint (here at $+\infty$, or more precisely at T, as $T \to +\infty$ in the limit). By reversing time in the differential equation, the forward-looking state variable in (2) can be thought of as accumulating backwards in time. Acting as if time flowed backwards, the retrograde state variable U_E "accumulates" according to the law of motion:

$$\frac{dU_E(t)}{-dt} = (u(w_t) - v(z_t) + a(m, z_t)U_u(t)) - (\rho + a(m, z_t))U_E(t).$$

Thus, the worker's problem can be thought of as a standard optimization problem where the direction of time is reversed, $U_E(t_0)$ is the objective function, and U_E is the retrograde state variable.

From the retrograde perspective, the worker is trying to maximize expected discounted utility by choosing effort z, conditional on the retrograde state variable U_E. The worker accumulates this utility stock U_E by collecting the instantaneous utility benefit $u(w_t) - v(z_t) + a(m, z_t)U_u(t)$ over time with U_E "depreciating" at rate $\rho + a(m, z_t)$.

The complication that makes the retrograde approach helpful is that the worker's choice of effort in the present and near future affects the present value of utilities gained over all future periods. Thus the worker must always account for how effort choice today will affect the present value of future expected utility, making effort choice dependent on this value.

While the retrograde perspective offers intuition for the workers problem, the direction of time flow itself is irrelevant to the solution of an optimal control problem. We can simply write

$$\max_{\{z_t\}} \quad U_E(t_0)$$

$$s.t. \quad \dot{U}_E(t) = -\Big((u(w_t) - v(z_t) + a(m, z_t)U_u(t)) - (\rho + a(m, z_t))U_E(t)\Big)$$

and solve the problem using the Hamiltonian equations. The Hamiltonian has the form,

$$\mathcal{H} = -\xi\Big([u(w) - v(z) + a(m, z)U_u] - [\rho + a(m, z)]U_E\Big) \tag{3}$$

where z is the control variable, U_E is the retrograde state variable and ξ is the retrograde co-state variable. The negative sign gives ξ an interpretation appropriate to the time-reversed retrograde perspective.[10] The objective function does not show up in the Hamiltonian because the objective function is expressed as a retrograde stock of utility with no flow component.

The Hamiltonian equations are textbook (Kamien and Schwartz, 1991). The derivative of the Hamiltonian with respect to the control variable is set to equal zero ($\frac{\partial \mathcal{H}}{\partial z} = 0$), the derivative with respect to the state variable is set to equal $-\dot{\xi}$ ($\frac{\partial \mathcal{H}}{\partial U_E} = -\dot{\xi}$), the derivative with respect to the co-state variable is set to equal $\dot{U}_E$ ($\frac{\partial \mathcal{H}}{\partial \xi} = \dot{U}_E$), and the transversality condition, which becomes an initial condition from the retrograde perspective, is $\lim_{t \to \infty} \xi_t U_E(t) = 0$. The only potentially unfamiliar condition from the retrograde perspective is a free "end point" condition with "salvage value," $\xi(t_0) = 1$.

Using these equations, the first-order condition for effort z from the worker's problem is:

$$-a_z(U_E - U_u) = v'(z). \tag{4}$$

In words, a worker exerts effort up to the point where the marginal disutility of effort is equal to the marginal reduction in the probability of job loss ($-a_z$) multiplied by the cost of job loss, the utility differential $U_E - U_u$.

Denoting the utility differential by $\Delta \equiv U_E - U_u$, equation (4) can be written as

$$m\Delta = v'(z), \tag{5}$$

since $a_z = -m$. Then, by assuming $v'(z)$ is strictly increasing in z ($v'' > 0$ is sufficient),

[10] Nothing substantive would change if the co-state variable were defined with the opposite sign. But it aids intuition to have most co-state variables be positive.

equation (5) can be inverted to yield

$$z = v'^{-1}(m\Delta). \tag{6}$$

Thus, effort depends only on the product of monitoring intensity and the cost of job loss. If m is treated as exogenous, z can be written as a function of the expected discounted utility differential $z = z(\Delta)$.

The life-time expected discounted utility of unemployed workers is determined by two components, the flow utility one receives when unemployed $\bar{u}$ and the probability of becoming employed again with a consequent jump up in expected discounted utility. The expected discounted utility of an unemployed worker is therefore:

$$U_u(t) = \int_t^{\infty} e^{-\rho(t'-t)}[\bar{u} + \frac{\theta_{t'}}{N_{t'} - L_{t'}}\Delta_{0,t'}]dt', \tag{7}$$

where $\frac{\theta}{N-L}$ is the instantaneous probability of being reemployed, equal to the aggregate rate of hiring, θ, divided by those in the labor force who are not among the employed, $N - L$. The increase in utility from getting a job and becoming employed is the utility differential Δ for a newly hired worker, denoted Δ_0. This can be simplified by defining

$$\psi \equiv \bar{u} + \frac{\theta}{N - L}\Delta_0 \tag{8}$$

as the pseudo-utility of being unemployed. This pseudo-utility accounts for the potential benefit an unemployed worker may gain from being rehired as well as the flow utility when unemployed, $\bar{u}$. In the model, ψ functions as the reservation utility which a firm must on average exceed to elicit effort from workers. Note that ψ is exogenous to the firm. A small firm cannot affect the value of an unemployed worker's opportunities. The Δ_0 that appears in the pseudo-utility of being unemployed is also outside of a firm's control because it is about other firms hiring workers that have left one's own firm.

By Leibniz's rule, the differential equations for expected lifetime utility when currently employed and unemployed respectively are:

$$\dot{U}_E = (\rho + a(m,z))U_E - a(m,z)U_u - [u(w) - v(z)] \tag{9}$$

$$\dot{U}_u = \rho U_u - \psi. \tag{10}$$

By subtraction,

$$\dot{\Delta} = \dot{U}_E - \dot{U}_u = (\rho + a(m,z))\Delta - [u(w) - v(z) - \psi]. \tag{11}$$

Together with the transversality condition, this implies:

$$\Delta_t = \int_t^{\infty} e^{-\int_t^{t'} [\rho + a(m, z_{\tilde{t}})] d\tilde{t}} [u(w_{t'}) - v(z(\Delta_{t'})) - \psi_{t'}] dt'. \tag{12}$$

From this expression, one can see that $u(w) - v(z) - \psi$ is the worker's instantaneous surplus from employment. This surplus utility must be discounted by the attrition rate $a(m,z)$ as well as by the impatience parameter ρ since with instantaneous probability $a(m,z)$ the worker becomes unemployed again and worker surplus drops to zero.

2.3 The Firm's Problem

A firm hires many workers and produces according to the production function

$$p_t f(Z_t),$$

where p_t is the technological component of productivity times a demand shifter, Z_t is the *total* effective labor for the firm by all of its workers, and f is an increasing and concave function of Z_t ($f' > 0, f'' < 0$). This can be thought of either as the firm having a fixed amount of capital, or as facing a downward-sloping demand curve due to monopolistic competition. A firm chooses the wage path for a worker, taking as given the path of the marginal revenue product of effort ($\mu_t \equiv p_t f'(Z_t)$) and the fact that the firm will only fire for cause. For an individual worker μ_t is exogenous, as μ_t is only infinitesimally affected by what any one worker does.

Firms hire workers by offering each worker an implicit long-term contract. By assumption, firms can commit to a wage path (that is conditional on the macroeconomic environment). This assumption is important because, as will be shown, firms will have the incentive to renege on their promises as the expected discounted value of the surplus the worker generates for the firm falls over time. Because it is beneficial for firms to defer payments to incentivize effort, workers' wages will exceed their productivity in the later segments of their careers at a given firm.

Thus, it is fair to question whether firms will be willing to honor their contracts. One possible interpretation of the model is that these long-term contracts, although implicit, are legally enforceable and firms are bound by law to keep their promises. However, even without such legal protection firms might be willing to honor their implicit contracts for reputational reasons. A firm's ability to hire workers and entice them to exert effort depends on a worker's belief that the firm will honor its promises. If a worker believes that a firm is untrustworthy after it observes a firm renege on a contract, the firm's ability to extract effort will be compromised. Formally, this incentive could be modeled as a game in which the firm's optimal strategy is to honor all contracts when the workers' punishment strategy for broken promises is stringent enough. Although we do not explicitly model the punishment for promise-breaking firms in this paper, it seems reasonable to believe that some firms are sensitive enough to their reputation to be willing to honor their implicit contracts.

The optimal contract for the firm maximizes the marginal profit gained from a worker over the worker's career with the firm. Denoting the expected value of discounted profit gained from a worker from time t_0 on as π_{t_0}, the optimal contract is the solution to:

$$\max_{\{w_t\}} \pi_{t_0} = \int_{t_0}^{\infty} e^{-\int_{t_0}^{t}[r_{\tilde{t}}+a(m,z_{\tilde{t}})]d\tilde{t}}[\mu_t z(\Delta_t) - w_t]dt \tag{13}$$

such that Δ_t satisfies (12) for all $t > t_0$. The firm wishes to maximize the difference between μz, the product gained from a worker, and wages w, discounted by the interest rate r and the separation rate $a(m, z)$.

The retrograde approach can be used to reformulate the problem as

$$\max_{\{w_t\}} \pi_{t_0} \tag{14}$$

$$\begin{aligned} s.t. \qquad \dot{\pi}_t &= -\Big([\mu_t z(\Delta_t) - w_t] - [r_t + a(m, z(\Delta_t))]\pi_t\Big) \\ \dot{\Delta}_t &= -\Big([u(w_t) - v(z(\Delta_t)) - \psi_t] - [\rho + a(m, z(\Delta_t))]\Delta_t\Big) \end{aligned}$$

and proceed to solve the problem using the Hamiltonian equations. The Retrograde Hamiltonian for the firm's problem is then:

$$\mathcal{H} = -\lambda\Big([\mu z(\Delta) - w] - [r + a(m, z(\Delta))]\pi\Big) - \zeta\Big([u(w) - v(z(\Delta)) - \psi] - [\rho + a(m, z(\Delta))]\Delta\Big), \tag{15}$$

where λ is the retrograde co-state variable for π, and ζ is the retrograde co-state variable for Δ. Expressed as (14) and (15), it becomes clear that the firm's problem is equivalent to choosing a path of wages to maximize the expected profit from worker at time t_0 (π_{t_0}), with an eye toward the effect of wages on the expected future path of the worker's surplus—and hence of firm profits, which depend on the worker motivation provided by that surplus.

The first-order condition for the worker's wage path is:

$$\frac{1}{u'(w_t)} = \frac{\zeta_t}{\lambda_t}. \tag{16}$$

The ratio $s_t \equiv \frac{\zeta_t}{\lambda_t}$ is a measure of "seniority rights" accumulated by the worker during their tenure with the firm. The co-state variable ζ_t is the marginal value of the utility differential of the worker Δ_t to the firm and λ_t is the marginal value of the future expected profit. Thus s_t represents the value of the worker's utility to the firm in units of future expected profit. Equation (16) shows that the higher a worker's seniority s_t, the more a firm is willing to pay for an extra unit of that worker's utility. As a result, the wage is a monotonic function of seniority. For example, $u(w) = \frac{w^{1-(1/\sigma)}}{1-(1/\sigma)}$ implies $w_t = s_t^{\sigma}$ where σ is the elasticity of intertemporal substitution.

The Hamiltonian equations for the state variables in (15) imply that seniority s_t obeys:

$$s_t = \int_{t_0}^{t} e^{-\int_{t'}^{t}[\rho - r_{\tilde{t}}]d\tilde{t}}[\mu_{t'} + m\pi_{t'}]z_{\Delta,t'}dt', \tag{17}$$

where $\pi_{t'}$ is the expected value of profits obtained from the worker after time t' and $z_{\Delta,t'}$ is the partial derivative of effort z with respect to the utility differential Δ at time t'. Equation (17) indicates that s_t is an accumulation of the value of worker utility to the firm. The value of worker utility to the firm reflects the marginal effect in the past of the utility differential Δ on effort ($z_{\Delta,t'}$), multiplied by the value of marginal effort in terms of firm profit ($\mu_{t'} + m\pi_{t'}$). The latter combines the direct effect of increased output per extra unit of effort $\mu_{t'}$ with the indirect effect $m\pi$ stemming from the reduction in the probability of worker dismissal, and therefore on the probability the firm is obliged to continue with the implicit contract.

The inclusion of $z_{\Delta,t'}$ in the integrand makes it clear why it is optimal for the firm to make wages depend on seniority. Recall from equation (12) that Δ_t is the discounted sum of the flow utility of the worker $\big(u(w) + v(z) - \psi\big)$ from time t onward. Thus, if the firm gives a worker a unit of utility (via the worker's wage) at time t, this will be reflected in the value of Δ from the time the worker is hired at t_0 to the moment the utility is realized at time t. This means that a unit of utility promised to the worker at future date t acts as a carrot until that time; increased worker effort is derived not from utility in the present, but from expected future utility. By integrating the marginal benefit of effort ($z_{\Delta,t'}$) over the interval $[t_0, t]$, equation (17) captures the marginal benefit to the firm from promising the wage w_t at time t as the accumulation of benefits the firm receives from the moment the worker is hired to the moment the wage is paid out at time t.

The discounting in (17), viewing time t utils from time t', is given by $e^{-\int_{t'}^{t}(\rho-r)d\tilde{t}}$ because utility at time t has to be discounted at the rate $\rho + a(m,z)$ back to t' to reflect the effect on the cost of job loss Δ at t', while future dollars written into the contract at time t become cheaper at the rate $r + a(m,z)$. The effects of the attrition rate $a(m,z)$ cancel out, because this affects the discounting of utils and dollars equally. Beyond that, utils at t are worth fewer utils at time t' if the worker's discount rate ρ is high, and are worth more in dollars at time t if those dollars are easy to come by because r is high.

The optimal long-term contract that results from the firm's problem is time-inconsistent in the sense that the optimal contract that maximizes the expected profit of the firm from a worker at hiring π_0, does not maximize expected profit to the firm at future dates. In other words, if the firm were able to renegotiate the contract at future dates without any consequence, it would be eager to do so. However, our seniority measure s_t is able to encode the value of having promised higher wages in the past.

The next section analyzes the optimal implicit contract in depth. Let us preview the

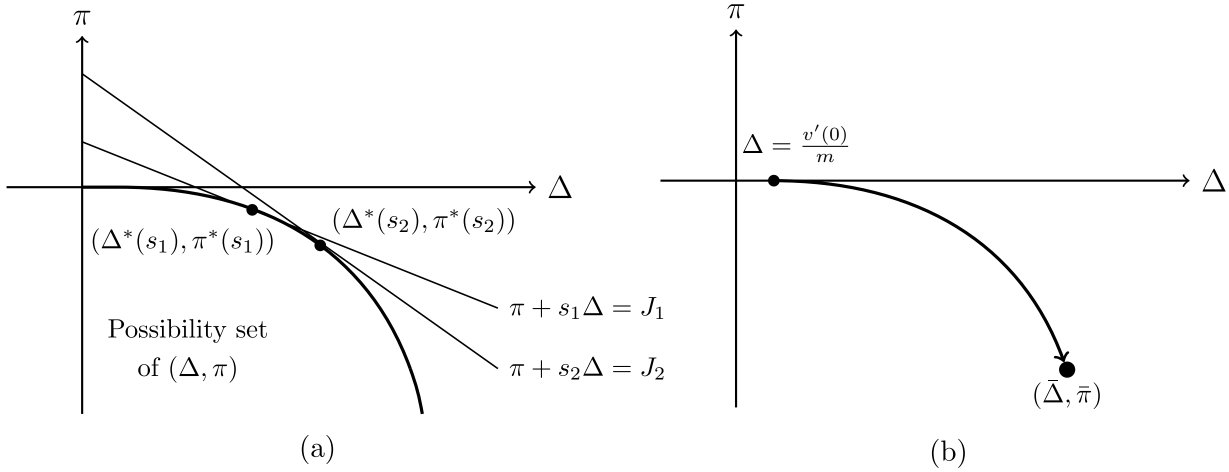


Figure 1: Optimal contract in (Δ, π)

central intuition here. Seniority s_t captures the value of worker utility to the firm. In other words, Δ_t is valued by the firm at s_t dollars per unit of utility on the margin. The continuation of the optimal contract at any time t after a worker is hired can be interpreted as the contract maximizing at time t the "time-consistent objective," $J_t = \pi_t + s_t\Delta_t$ at each point in time, which combines the value of expected future profits from worker π_t with the value captured by the term $s_t\Delta_t$ of keeping the utility differential high for the sake of motivating the worker at previous moments of time. That is, the continuation of the contract from time t on solves

$$\max_{w_t} \pi_t + s_t\Delta_t, \tag{18}$$

where s_t is taken as given, Δ_t satisfies (12), and π_t satisfies (13).

In panel (a) of Figure 1, we illustrate how the optimal contract can be determined given (18). First, there will be some bounded-above set of attainable combinations of Δ and π at any time t, given the expected path of the three key variables exogenous to the contract: μ, ψ and r.[11] The optimal contract will always lie at a point on the efficient frontier of this (Δ, π)

[11]Note that the set of possible (Δ, π) pairs will always be in the southeast quadrant. If π were greater than zero the firm would immediately hire more workers until the resulting decline in the marginal product μ brought π back down to zero. If Δ were less than zero, then working would be less attractive to workers

possibility set. In Appendix B.1, we prove that the relevant part of the efficient frontier will be strictly downward-sloping and strictly concave in steady-state. In panel (a) of Figure 1, the efficient frontier is represented by the thick curved line.[12] Now, note that for a given level of seniority s, the time-consistent objective $\pi + s\Delta$ can be shown on the $\Delta - \pi$ plane as a straight line with the slope $-s$.

In steady state, firms maximize the "time-consistent objective" $\pi + s\Delta$ in steady state because strict concavity of the efficient frontier guarantees that that there is only one point with the slope $-s$ needed to satisfy the local necessary condition for trade-offs between π and Δ. Thus, we can think graphically for a tangency between the line $\pi + s\Delta = J$ where J is some constant, and the efficient frontier. For example, the points $(\Delta^*(s_1), \pi^*(s_1))$ and $(\Delta^*(s_2), \pi^*(s_2))$ on the efficient frontier in Figure 1 represent the tangents between the efficient frontier and the lines $\pi+s\Delta = J$ for the highest possible $J's$ at two different seniority levels s_1 and s_2.

In steady state, the expected present value of future profits from senior workers is negative. In equilibrium, a firm will hire workers up to the point where π_{t_0}, the expected discounted profit from the worker at the time of hiring, is zero. The firm backloads wages, pushing wages further into the future than productivity. Therefore, after some time the firm will face relatively more future wages than future productivity from the worker and the worker ends up being a contractual burden on the firm. As a result, the fact that a worker is more likely to stay with the firm with greater effort diminishes the firm's incentive to defer wages indefinitely.

---

than unemployment.

[12]The efficient frontier is a subset of the upper right-hand boundary of the convex hull of attainable (Δ, π) pairs. The efficient frontier can be disconnected if the region of possible (Δ, π) pairs is non-convex. Non-convexity may arise if firms are able to place workers on temporary leave. If Δ falls such that the implied level of effort z is less than zero, it is beneficial for the firms to temporarily lay off the workers (with pay) and get zero effort instead. This arrangement would not be a breach of the long-term contract as the firms would still honor the terms of their agreement. In any case, in steady state, layoffs are unnecessary and the (Δ, π) possibility set will be convex.

Panel (a) of Figure 1 shows that the optimized value of the time-consistent objective $\pi^*(s) + s\Delta^*(s)$ is non-decreasing in seniority.[13] Because the efficient frontier is downward sloping, the tangency point representing the optimal (Δ, π) for a given level of seniority s moves southeast along the efficient frontier as s increases. The value of $\pi + s\Delta$ is equal to the π-intercept of the tangency line, which increases with seniority s. However, π itself is decreasing in s: the firm is willing to trade off profit for a higher expected present value of worker utility as seniority s rises. At a firm that has workers at various levels of seniority, senior workers will be more of a burden on the firm (π is more negative), will value their jobs more (Δ is higher), will have higher productivity (z is higher), and are less likely to be fired ($a(m, z)$ is lower) than junior workers.

2.4 Discussion

Let us pause to take stock of our contributions so far, and how it lays the foundation for further investigation. First is deriving a precise definition of seniority s_t and showing that a worker's wage profile reflects this measure. Unlike previous studies on deferred compensation, our framework provides not only the general intuition that compensation is deferred to motivate workers, but also the logical structure of the compensation scheme and its determinants.

Second, our results allow us to interpret the gradient of the tenure-wage profile. We show that the increase in wages with seniority depends on the effect of the future wage on the path of past effort multiplied by the marginal product of effort along that path. This allows for a much richer interpretation of the estimates governing tenure-wage relationships, opening up the possibility for empirical decompositions of the components of seniority wages. A key reason we are able to provide determinate expressions for the path of seniority wages (unlike past studies) is that we allow effort to be a continuous choice variable, while past

[13] See Proposition 2 for a proof.

work on effort choice restricts the worker to a binary work/shirk choice (e.g. Lazear 1979, 1981, Akerlof and Katz, 1989). We are able to solve and interpret this more complex model thanks to the Retrograde approach.

Admittedly, assuming that the worker's utility is monotonically increasing and concave in the wage is also critical to the determinacy of the wage profile. As noted previously, we rationalize this assumption by regarding the worker as being liquidity constrained and impatient, making the worker unable to borrow and unwilling to save. But note that even with more complex assumptions about borrowing and saving, the worker's first-order condition for effort (4) would remain unchanged and the firm's first-order condition (13) would remain unchanged. What would change would be the expression for Δ in equation (12) and the expression for seniority in (17). Most of the key insights from the analysis here would remain intact.

Third, our results point to many other things besides wages that are affected by seniority. As we will show in Section 4, our framework allows for easy extensions: we can easily incorporate additional features into the firm-worker relationship and analyze the evolution of the additional variables.

3 Digging Further into the Optimal Contract

In this section, we analyze more deeply the evolution of the optimal contract over the worker's tenure with the firm. First, we show that the contract that maximizes the "time-consistent objective" is equal to the optimal contract.

Proposition 1. *The contract that maximizes the time-consistent objective* (18) *every period is identical to the optimal contract from the firm's problem in* (14).

To see this, notice that the solution to the problem in (18) gives the first-order condition for the path of wages from (14) as given by (16). The first-order condition is

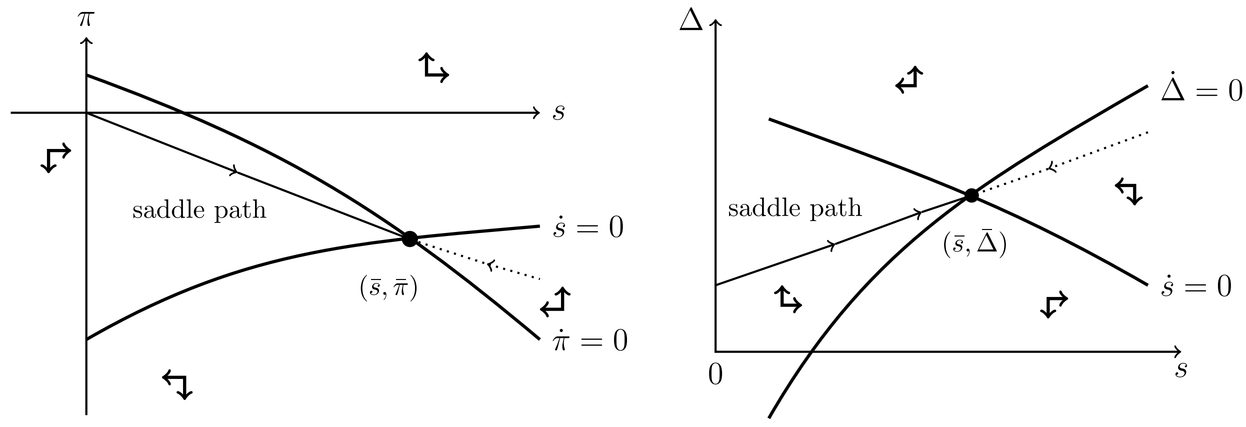
$$0 = \frac{\partial \pi_t}{\partial w_t} + s_t \frac{\partial \Delta_t}{\partial w_t},$$

where $\frac{\partial \pi_t}{\partial w_t} = 1dt$ and $\frac{\partial \Delta_t}{\partial w_t} = u'(w_t)dt$, resulting in an equation identical to (16) at time t. The strict concavity in steady state of the (Δ, π) Efficient Frontier guarantees there is only one point on the efficient frontier with this slope. Thus, solving for the optimal wage w_t from (18) at each point in time will result in the optimal contract from (14).

Proposition 2. *In steady state, Δ is (weakly) increasing in worker seniority s, π is (weakly) decreasing in s and the time-consistent objective $\pi + s\Delta$ is (weakly) increasing in s.*

To see that $\pi + s\Delta$ is increasing in worker seniority consider the following. Recall from Figure 1, that in steady state the possibility frontier of (Δ, π) is constant over time. Then, given two seniority levels $s_2 \geq s_1$, let (Δ_1, π_1) maximize $\pi + s_1\Delta$ and (Δ_2, π_2) maximize $\pi + s_2\Delta$ in the feasible set. First, note that $\pi_1 + s_1\Delta_1 \geq \pi_2 + s_1\Delta_2$ and $\pi_2 + s_2\Delta_2 \geq \pi_1 + s_2\Delta_1$ by definition. Subtracting the two conditions, we get $(s_2 - s_1)\Delta_2 \geq (s_2 - s_1)\Delta_1$, which implies that $\Delta_2 \geq \Delta_1$. Therefore, Δ is non-decreasing in seniority. Next, from $\pi_1 + s_1\Delta_1 \geq \pi_2 + s_1\Delta_2$ we get $\pi_1 \geq \pi_2 + s_1(\Delta_2 - \Delta_1) \geq \pi_2$. Thus, π is non-increasing in seniority. Lastly, $\pi_2 + s_2\Delta_2 \geq \pi_1 + s_2\Delta_1 \geq \pi_1 + s_2\Delta_1 - (s_2 - s_1)\Delta_1 = \pi_1 + s_1\Delta_1$ which shows $\pi + s\Delta$ is non-decreasing in seniority.

Using Figure 1b we can examine the evolution of the optimal contract more closely over the worker's tenure with the firm. As will be shown in Proposition 3, any worker who neither quits nor is fired will gain seniority s and move down along the (Δ, π) Efficient Frontier. The value of the utility differential Δ when the worker is hired is $\Delta_0 = \frac{v'(0)}{m}$. This follows from equation (5) and the fact that in equilibrium, the worker's effort must be zero when hired because if effort were strictly positive when π was zero, then it would be possible for the firm to gain strictly positive profit by adding a period of zero wages to the front end of



(a) Optimal contract in (s, π)

(b) Optimal contract in (s, Δ)

Figure 2: Evolution of the optimal contract.

the implicit contract and get a positive expected present value of profits overall. But in equilibrium, the marginal profit from a worker must be zero when a worker is hired. Panel (b) also shows that the optimal contract in steady state will eventually reach a stationary point depicted in the figure as $(\bar{\Delta}, \bar{\pi})$, which we explain in more detail below.

Up to this point, we have implicitly assumed that workers gain seniority with tenure. In the following propositions, we describe the evolution of worker seniority and effort with worker tenure at the firm.

Proposition 3. *In steady state, seniority "s" starts at zero when the worker is initially hired, and increases with tenure, until it reaches a stationary level $\bar{s}$ that is the asymptotic limit for seniority if a worker has a very long tenure with a firm.*

Proposition 4. *In steady state, the utility differential from hiring Δ increases with worker tenure until it reaches a stationary point $\bar{\Delta}$.*

Corollary 1. *In steady state, the worker's effort z increases with worker tenure until it reaches a stationary point.*

To show Propositions 3 and 4, we describe the dynamics of the optimal contract using phase diagrams given by the law of motion of the three variables π, Δ, and s. The optimal contract lies on the three dimensional saddle path of the dynamics described by the steady state differential equations,

$$\begin{aligned}
\dot{\pi} &= \Pi(\pi, \Delta, s) = -\Big((\mu z(\Delta) - w(s)) - (r + a(m, z(\Delta)))\pi\Big) \\
\dot{\Delta} &= \nabla(\pi, \Delta, s) = -\Big((u(w(s)) - v(z(\Delta)) - \psi) - (\rho + a(m, z(\Delta)))\Delta\Big) \\
\dot{s} &= S(\pi, \Delta, s) = (\mu + m\pi)z_\Delta(\Delta) - (\rho - r)s.
\end{aligned} \tag{19}$$

The partial derivatives of the accumulation functions Π, ∇, S, denoted by subscripts, are:

$$\begin{aligned}
\Pi_\pi &= r + a(m, z(\Delta)) > 0 \\
\Pi_\Delta &= -(\mu + m\pi)z_\Delta(\Delta) < 0 \\
\Pi_s &= w'(s) > 0 \\
\nabla_\pi &= 0 \\
\nabla_\Delta &= \rho + a(m, z(\Delta)) > 0 \\
\nabla_s &= -u'(w(s))w'(s) < 0 \\
S_\pi &= mz_\Delta(\Delta) > 0 \\
S_\Delta &= (\mu + m\pi)z_{\Delta\Delta}(\Delta) < 0 \\
S_s &= -(\rho - r) < 0.
\end{aligned} \tag{20}$$

Notes: (A) The signs given for these partial derivatives are for the region in which $\mu + m\pi > 0$ (which is the relevant region, since that is initially true, and $\mu + m\pi > 0$ is necessary for $\dot{s} > 0$, which in turn is what drives $\mu + m\pi$ down. (B) The computation of the partial derivative ∇_Δ uses the envelope theorem—or equivalently, the first-order condition $v'(z) = m\Delta$. (C) $z_{\Delta\Delta}(\Delta) < 0$ relies on the assumption $v'''(z) > 0$, which amounts to assuming that at high levels of effort the difficulty and therefore disutility of extra effort increases dramatically. Differentiating the identity $v'(z(\Delta)) = m\Delta$ twice with respect to Δ then yields $z_{\Delta\Delta}(\Delta) = \frac{-v'''(z(\Delta))[z_\Delta(\Delta)]^2}{v''(z(\Delta)} < 0$.

To visualize the three-dimensional phase diagram, we show three two-dimensional slices through the stationary point, with the arrows showing the dynamics for key regions in the applicable two dimensions. The signed derivatives above justify the dynamics shown and the slopes shown for the isoclines (the $\dot{\pi} = 0$, $\dot{\Delta} = 0$ and $\dot{s} = 0$ lines in the two-dimensional slices). With two forward-looking variables, π and Δ, and one backward-looking variable, s, there is a well-defined saddle path in three dimensions leading to the stationary point. (The explosive dynamics in the π-Δ slice shows that being off that narrow path would make it impossible to hit the stationary point.) The saddle path gives π and Δ as functions of s.

With the phase diagram in hand, first examine the evolution of worker seniority as described by Proposition 3. In Figure 2a, we illustrate the evolution of seniority s in (s, π) space, taking the path of Δ as given. The thick lines represent the $\dot{s} = 0$ and $\dot{\pi} = 0$ loci respectively. The direction of the dynamics is shown by the small arrows at right angles and the saddle path is as labeled. From the boundary conditions of the Hamiltonian, we know that seniority s is equal to zero when the worker is hired.[14] Also, as illustrated, the contract must begin with $\pi = 0$ in equilibrium as firms would otherwise hire more workers until their expected profit declined to zero. Therefore, the saddle path begins at the origin in (s, π) space and steadily accumulates until it reaches a stationary point $(\bar{s}, \bar{\pi})$ where the $\dot{s} = 0$ and $\dot{\pi} = 0$ loci meet.

Together with the result that Δ increases in s by Proposition 2 and the dynamics of the phase diagram for s, Proposition 3 yields Proposition 4. To illustrate Proposition 4 from another point of view, let's look at the dynamics of the optimal contract in the (s, Δ) space. Figure 2b shows the evolution of seniority s and utility differential Δ. Again, the contract begins at the point where the saddle path intersects with the Δ axis. It becomes clear that

[14] The free end point condition for $\dot{\Delta}$ from the retrograde perspective implies that $\zeta = 0$ and the free end point condition with salvage value for $\dot{\pi}$ implies $\lambda = 1$ at time t_0.

both seniority and the utility differential Δ increase over time until it reaches a stationary point $(\bar{s}, \bar{\Delta})$ where the $\dot{s} = 0$ and $\dot{\Delta} = 0$ loci meet. Lastly, effort increases with time as well since it is a monotonic function of Δ as described by Corollary 1.

Combined, Figures 1 and 2 show the three-dimensional saddle path of s, Δ and π on which the contract lies. As previously described, the phase diagrams reveal that there is a stationary point $(\bar{\pi}, \bar{\Delta}, \bar{s})$ where the contract comes to rest where $\dot{s} = \dot{\Delta} = \dot{\pi} = 0$. Seniority s stops growing in the end because as s increases the term $m\pi$ in the $\mu + m\pi$ of the integrand for s in equation (17) becomes increasingly negative. Beyond a certain point, promising the worker higher wages would make them stay too long from the firm's viewpoint, canceling out the benefit of extra productivity. This limit to how far s can rise is a reflection of diminishing returns to the firm from raising wages. Also preventing seniority s from rising indefinitely is worker impatience. If $\rho > r$, payments far in the future have less effect on effort in the present.[15]

Finally, we describe the relationship between the optimal wage and worker productivity in the following proposition.

Proposition 5. *As worker tenure converges to infinity, the wage is greater than the marginal product of the worker, μz. In other words, for sufficiently senior workers their wage is greater*

[15] Note that combining equation (16) with the differential equation for seniority in (19) yields:

$$\frac{d(\frac{1}{u'(w)})}{dt} = (\mu + m\pi)z_\Delta(\Delta) - (\rho - r)\frac{1}{u'(w)}.$$

This is an "inverse wage Euler equation" for this efficiency wage model that can be compared to the inverse wage Euler equation in search models such as, for example, the wage Euler equation in Acabbi et al.(2022)'s Proposition 3.3. The equation can be further simplified by assuming log utilities $u(w) = log(w)$ and $\rho = r$ to derive,

$$\frac{\dot{w}}{w} = (\mu + m\pi)\frac{z_\Delta(\Delta)}{w},$$

which governs the changes in the path of wages over the life cycle of the contract. The term $m\pi\frac{z_\Delta(\Delta)}{w}$ captures the increase in worker retention to offered utility in dollar units, which is analogous to Accabi et al.(2022)'s expression for wage growth which depends on the elasticity of retention probability to offered utility. We have an additional term for increases in productivity from worker effort $\mu\frac{z_\Delta(\Delta)}{w}$ due to the efficiency wage nature of our framework, which is not present in search and matching frameworks.

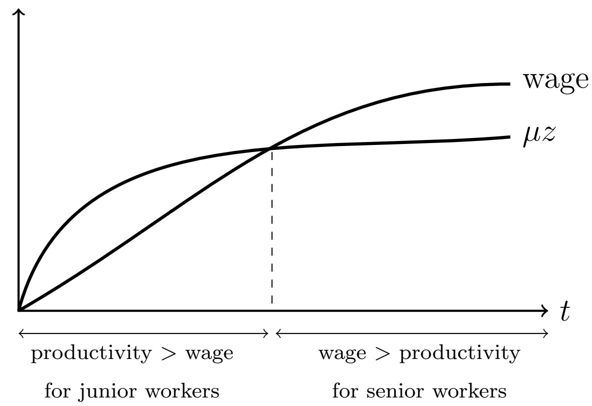


Figure 3: Wage vs productivity

than their productivity.

Consider what happens at the stationary point where $\dot{s} = 0$. We know that the wage w eventually ends up above the marginal product of the worker μz because as $t \to \infty$ the stationary contract satisfies

$$\bar{\pi} = \frac{\mu \bar{z} - \bar{w}}{r + \bar{a}} < 0. \tag{21}$$

Equation (21) can be derived by substituting stationary values into the expression in equation (13). The equation shows that the marginal product at the stationary point $\mu \bar{z}$ must be less than the worker's wage $\bar{w}$. Furthermore, we know that the worker's wage must be below the worker's marginal product at some point earlier on to balance this, or the expected future profits from taking on a worker at t_0 would be negative and the firm would never agree to hire the worker on such terms. Figure 3 shows the simplest possibility: marginal product greater than the worker's wage when the worker is junior and less than marginal product when the worker is senior.[16]

[16]Looking at things from various different angles suggests the generality of the patterns shown by the phase diagrams even with additional complications such as those discussed in the next section. First, any counterparts to the equation $\dot{s} = (\mu + m\pi) z_\Delta(\Delta) - (\rho - r)s$ would imply that $\dot{s} > 0$ when $s = 0$, and given the key "impatience" assumption $\rho > r$, that $\dot{s} < 0$ at a high enough s. Without venturing too deep into the specifics of any particular model, this suggests that s starts at zero, then gradually rises toward an asymptote. This fact, in addition to the argument that s controls the choice on a Δ-π possibility frontier, as shown in Figure 1, guarantees the essentials of the dynamics discussed in this section.

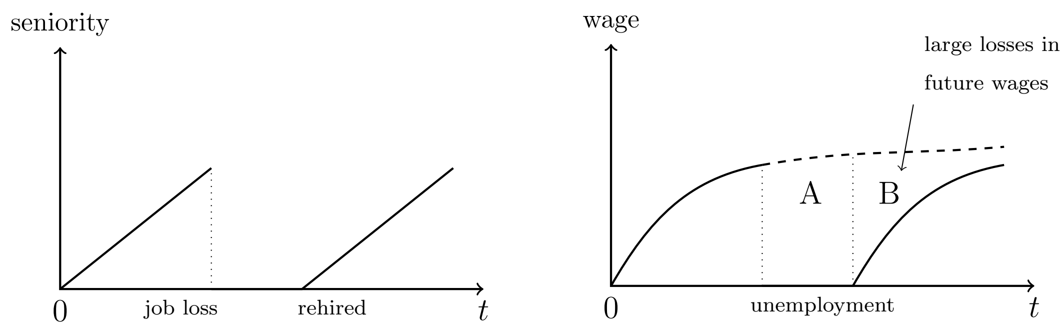
4 Extensions and Applications

In this section, we explore extensions and applications of our framework to show how it sheds light on many aspects of worker-firm relationships. First, we consider the question of large earnings losses from job loss. There is a disparity between empirical results reporting large earnings losses and theoretical models that typically predict a smaller impact from job loss. Also needing explanation is the fact that those senior workers who lose their jobs have higher unemployment rates thereafter, even beyond the time they find the very next job. Second, we consider how optimal monitoring intensity evolves with seniority. Various subtleties can make the direction ambiguous without additional assumptions on the shape of the disutility of effort function $v(z)$. Yet, we show that plausible assumptions guarantee that monitoring intensity will fall with seniority. Third, we consider a firm being able to provide amenities to workers in addition to wages to increase job pleasantness and investigate the optimal strategy for the firm. We find that, in general, firms provide ordinary amenities to workers in a manner similar to wages, but amenities that reduce a worker's marginal disutility of effort are especially attractive when effort elicitation is a problem, as in the type of model presented here. Fourth, we consider a firm's decision of how much on-the-job training to provide. We show that, despite the fact that firm-specific human capital depreciates, it is optimal to provide on-the-job training that makes firm-specific human capital rise throughout a worker's tenure with a firm.

4.1 Earnings Loss

Many empirical studies have found that there is a large loss in life-time earnings when workers lose their jobs: a highly persistent negative shock to future wages and a persistent increase in the likelihood of future unemployment.[17] To date, most research on the topic has

[17] See, for example, Jacobson et al. (1993), Fallick (1996), Kletzer (1998), Couch and Placzek (2010), and Lachowska et al. (2020).



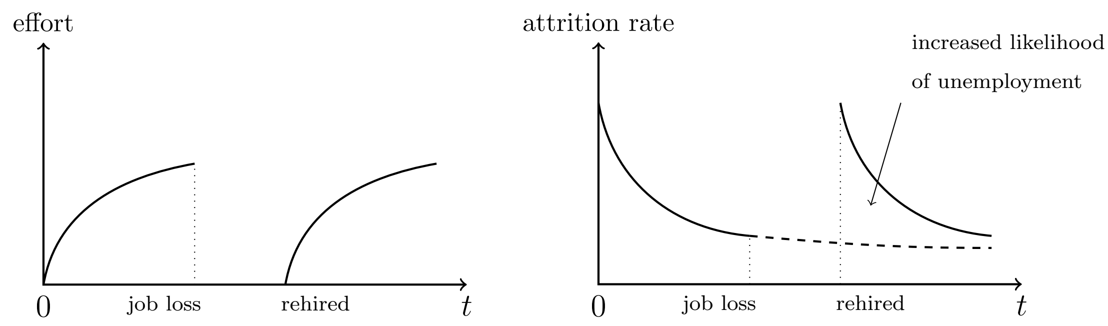
(a) seniority

(b) wage

Figure 4: Wage loss

been conducted within the search and matching framework. For example, Davis and von Wachter (2011) find that leading search models are unable to reconcile these findings with theory. They find that realistically calibrated models predict much smaller earnings losses than found in empirical studies. More recent work tries to reconcile the model to the data by adding various features such as job ladders (Krolikowski, 2017, Jung and Kuhn, 2019), human capital (Burdett et al. 2020), or both (Jarosch, forthcoming). In the model here, large losses in life-time earnings naturally arise from the seniority wage structure. Wages are a reflection of seniority, and when a worker is fired they lose the seniority they have accumulated. Since seniority is not transferred from firm to firm when a worker is fired, workers cannot regain the lost seniority even when reemployed. This force is not represented in any of the aforementioned studies.

Figure 4 illustrates wage losses from the loss of seniority after job loss. Panel (a) shows the seniority a worker accumulates with years at a firm. The solid lines depict the worker's accumulated seniority. After job loss, even if a worker is quickly rehired, the worker must re-accumulate seniority from scratch as would any new hire. In Panel (b), the solid line shows the worker's wage and the dashed line shows the counterfactual wage of the worker had they not been fired. The loss is considerable. In addition to the lost wage during unemployment (the area labeled A), there is a persistent reduction in the wage at the next job compared



(a) effort

(b) attrition rate

Figure 5: Probability of job loss

to the counterfactual case. This is due to the difference in the level of accumulated seniority (the area labeled B). Thus, there is a large and persistent earnings losses from temporary job loss.

The model here also explains why the probability of future unemployment is persistently greater after job loss: it is due to the lower effort level of a worker who has less seniority. The utility differential between being employed and unemployed increases as a worker gains seniority, which leads to effort increasing with seniority. Thus, senior workers are less likely to be fired because they work harder and make fewer mistakes, as shown in Figure 5. When workers lose their accumulated seniority with job loss, the effort level is reset and this leads to a persistent increase in the probability of future job loss, even after they have found their next job.

4.2 Variable Monitoring

In this section we consider how a firm should vary monitoring intensity over the worker's tenure. To capture the trade-offs inherent in the monitoring of workers, we add the cost of monitoring into the firm's problem. We denote the total cost of monitoring to the firm as $c(m) \geq 0$ and assume that the marginal cost is positive and increasing in the intensity of monitoring ($c'(m) > 0, c''(m) > 0$) except that $c'(0) = 0$. That is, we assume that the first bit of monitoring is very easy, with only a second-order cost: $c'(0) = 0$. These functions are defined over the nonnegative reals.

Given the worker's first-order condition $v'(z) = m\Delta$, we define $Z(x) = v'^{-1}(x)$, with $x = m\Delta$, or equivalently, $m = \frac{x}{\Delta}$. Furthermore, we make assumptions on the disutility of effort to derive clear implications for monitoring. Because monitoring intensity determines the attrition rate which in turn affects effort, the shape of the disutility of effort function plays an important role on the path of monitoring intensity. First, note that our assumptions on $v(z)$ ($v'(\cdot) > 0, v''(\cdot) > 0, \lim_{z\to 1} v'(z) = +\infty$) implies that $1 - Z(m\Delta) > 0$). We also assume

$$\frac{(1-z)v''(z)}{v'(z)} \tag{22}$$

is increasing in z and ≥ 1. This is equivalent to assuming

$$\frac{xZ'(x)}{1-Z(x)}$$

is decreasing in x and ≤ 1.[18] Given the assumption that $v(z) \to \infty$ as $z \to 1$ that we use to represent the idea that $z = 1$ —where no big blunders would ever happen—is an

[18] These assumptions, for example, are satisfied by the attractive functional form with $\vartheta > 0, \varkappa > 0$, $\ell \in [0, \vartheta\varkappa]$, and $\varrho \geq -\frac{\ell}{\vartheta}$:

$$v(z) = \varrho + \ell z + \frac{\varkappa}{\vartheta}(1-z)^{-\vartheta}, \qquad v'(z) = \ell + \varkappa(1-z)^{-(1+\vartheta)},$$
$$v''(z) = (1+\vartheta)\varkappa(1-z)^{-(2+\vartheta)}, \qquad v'''(z) = (2+\vartheta)(1+\vartheta)\varkappa(1-z)^{-(3+\vartheta)},$$
$$\frac{(1-z)v''(z)}{v'(z)} = \frac{(1+\vartheta)}{1+\frac{\ell}{\varkappa}(1-z)^{1+\vartheta}} \geq 1, \qquad \frac{v'''(z)v'(z)}{v''(z)^2} = (2+\vartheta)[\frac{1}{1+\vartheta} + \frac{\ell}{\varkappa}(1-z)^{1+\vartheta}] \geq 1.$$

For this functional form, when $x \geq \ell + \varkappa$ (the relevant range, since any $x = m\Delta$ below $\ell + \varkappa$ would yield no effort, so the firm would go there),

$$z(x) = 1 - [\frac{x-\ell}{\varkappa}]^{-\frac{1}{1+\vartheta}}$$
$$z'(x) = \frac{1}{\varkappa(1+\vartheta)}[\frac{x-\ell}{\varkappa}]^{-1-\frac{1}{1+\vartheta}} > 0$$
$$z''(x) = (\frac{-(2+\vartheta)}{\varkappa^2(1+\vartheta)^2}[\frac{x-\ell}{\varkappa}]^{-2-\frac{1}{1+\vartheta}} < 0.$$

unattainable ideal, it is a mild additional assumption to assume that $\frac{(1-z)v''(z)}{v'(z)}$ is weakly increasing in z.

With these assumptions, we can show that the (Δ, π) Efficient Frontier with endogenous monitoring is concave, as was the case for the case when monitoring intensity was exogenously given.[19] This implies that the dynamics of (π, Δ, s) as well as the time-consistent objective $\pi + s\Delta$ will be as analyzed in Section 3. Appendix B.1 also goes most of the way toward showing that monitoring intensity unambiguously declines with seniority. To solve for the optimal path $\{m_t\}$, derive the Hamiltonian equations for the firm's problem (15) and the first-order condition $\frac{\partial H}{\partial m} = 0$:

$$\underbrace{[\mu + m\pi]\Delta Z'(m\Delta)}_{\text{benefit of monitoring}} - \underbrace{[\pi + s\Delta](1 - Z(m\Delta)}_{\text{indirect cost}} - \underbrace{c'(m)}_{\text{direct cost}} = 0. \tag{23}$$

Appendix B.1 shows (a) that for a given Δ, higher m strictly lowers the left-hand side of this first-order condition, ensuring that there is a unique optimum for m and (b) that for a given value of m, an unchanged s and higher Δ reduces the left-hand side of this first-order condition. Of course, given the concavity of the (Δ, π) Efficient Frontier, higher Δ goes along with higher s, which further reduces the left-hand side of this first-order condition for optimal m at any given value of m. That guarantees that as Δ increases, optimal m must fall.

(When $x \leq \ell$, we assume $z(x) = 0$.) Also, note that

$$\frac{z'(x)}{1 - z(x)} = \frac{1}{(1 + \vartheta)(x - \ell)}$$

— which is decreasing in x in the domain $[\ell + \varkappa, +\infty)$. Also, in that same domain,

$$\frac{xz'(x)}{1 - z(x)} = \frac{x}{(1 + \vartheta)(x - \ell)} \leq 1,$$

since $\frac{\ell}{\varkappa} \leq \vartheta$.

[19]See Appendix B.1 for a proof.

The intuition and a key step in the Appendix B.1 proof that m is decreasing with seniority can be seen by dividing the first-order condition for optimal m through by $1-Z(m\Delta)$:

$$\underbrace{[\frac{\mu}{m}+\pi]\left(\frac{m\Delta Z'(m\Delta)}{1-Z(m\Delta)}\right)}_{\text{benefit of monitoring}}-\underbrace{[\pi+s\Delta]}_{\text{indirect cost}}-\underbrace{\left(\frac{c'(m)}{1-Z(m\Delta)}\right)}_{\text{direct cost}}=0. \tag{24}$$

After dividing through by $1-Z$, the marginal benefit of monitoring is declining with higher Δ given fixed m because both the factor $\frac{\mu}{m}+\pi$ and the factor $\frac{xZ'(x)}{1-Z(x)}$ are positive and declining—the former for reasons discussed above, the latter by a key assumption. The first piece of the marginal cost of monitoring divided by $1-Z$ is $\pi+s\Delta$, which increases with seniority for reasons discussed above, given the concavity of the (Δ,π) Efficient Frontier with endogenous monitoring. The second piece of the marginal cost of monitoring divided by $1-Z$ is $\frac{c'(m)}{1-Z(m\Delta)}$, which is obviously decreasing with fixed m and increasing Δ, simply because $1-Z$ declines.

Monitoring intensity declining with seniority when it is endogenous is a key finding. It would be an unusual data set that included a measure of monitoring intensity, but this prediction of monitoring intensity declining with seniority could be taken directly to the data if there were. As it stands, it accords with our sense of how things are within most organizations.

To end this subsection, we note that with endogenous monitoring, we can no longer guarantee that effort $Z(m\Delta)$ increases with seniority. We cannot rule out the possibility that monitoring intensity declines so dramatically with Δ that this decline overwhelms the direct effect of higher Δ on worker effort. If effort ever does decline with seniority, it would be the firm's choice to have that happen by its optimizing choice of monitoring intensity.

4.3 Amenities

Firms may also provide amenities (vacation days, office space, shifts, etc.) to increase worker utility in ways that vary with seniority. We explore amenities in this subsection.

Amenities can be utility-enhancing in an additively separable way, or they can pack more of a punch for the firm by reducing the disutility of effort as well. To allow for this possibility, let amenities A enter the disutility of labor function as a second argument: $v(z, A)$ with $v_A < 0, v_{AA} > 0$. Amenities are measured in dollars. The importance of amenities that reduce the disutility of effort show up through $z_A(\Delta, A)$. To find $z_A(\Delta, A)$, start with the worker's first-order condition (5) modified to take into account amenities: $m\Delta = v_z(z, A)$. Taking the total differential of this equation with respect to variables z and A we get $0 = v_{zz}dz + v_{zA}dA$. Rearranging, this implies: $z_A(\Delta, A) = \frac{dz}{dA} = -\frac{v_{zA}}{v_{zz}}$. This equation implies that if an amenity reduces the disutility of effort, then it increases effort, holding Δ (the value of the job to the worker) fixed.

The optimal path of amenities can be characterized by the first-order condition for the effects of amenities on the modified Hamiltonian (15): $\frac{\partial H}{\partial A} = 0$. The optimal path of $\{A_t\}$ solves:

$$[\mu + m\pi]z_A + s[-v_A] = 1. \tag{25}$$

The term $[\mu + m\pi]z_A$ represents the net benefit the firm reaps from additional effort. $s[-v_A]$ is the firm's indirect benefit from the increase in worker utility. The value of 1 on the right-hand side is the marginal cost of providing an additional unit of amenities (because amenities are measured by their cost in dollars per unit time).

If $v(z, A)$ is additively separable ($v_{zA} = 0$) then an increase in amenities does not affect the marginal disutility of effort and $z_A = 0$. Condition (25) can then be expressed as,

$$s = \frac{1}{-v_A}$$

which is identical to the optimal condition for wages in equation (16) with the marginal utility gain from amenities $-v_A$ replacing the marginal utility gain from wages $u'(w)$. As with wages, amenities increase as the worker gains seniority. These predictions seem to be in line with observations from the real world. Examples of amenities that may not affect the marginal disutility of effort may include vacation days, luxurious office spaces, or tuition subsidies for employees' families. Empirically, these kinds of amenities are often provided more generously to more senior workers.

If $v(z, A)$ is not additively separable, equation (25) becomes instead:

$$s = \frac{-(\mu + m\pi)z_A}{-v_A} + \frac{1}{-v_A}. \tag{26}$$

Consider amenities that decrease the marginal disutility of effort ($v_{zA} < 0$), such as air conditioning or ergonomic chairs that allow workers to work longer hours with less physical fatigue. They will increase effort, and as a result it is optimal to provide them more abundantly. The first term on the right-hand side of equation (26), $\frac{-(\mu+m\pi)z_A}{-v_A}$ is negative in that case. Therefore, all else equal, the second term, $\frac{1}{-v_A}$, which goes up with A, needs a higher value of A to satisfy the equation than if $v(z, A)$ were additively separable. Intuitively, making effort easier is an extra benefit and so it is optimal to provide more of such amenities than those that do not encourage effort. By the corresponding logic, amenities that increase the marginal disutility of effort—such as video game equipment in a law office—will rarely be provided because they discourage effort. Exceptions to the rule that distracting amenities will rarely be provided may arise when workers are so highly motivated that these distractions don't impair their effort very much, or when their direct value to workers is high enough to counterbalance negative effects on effort.

Now, consider how the optimal provision of amenities will depend on seniority when $v_{zA} < 0$, by looking at what would happen to the marginal benefits and marginal costs of amenities if A did not change. The cost of providing amenities is constant at 1 since we measure amenities in dollars. In steady state, the net benefit from additional effort, $[\mu + m\pi]$, decreases with seniority. In the term $s[-v_A]$ representing the utility benefits from amenities, not only s but also $[-v_A]$ rises with seniority since $v_{zA} < 0$ and worker effort increases with seniority. This is the other implication of the complementarity of amenities and effort when $v_{zA} < 0$. The overall effect of seniority on amenities is unclear. On the one hand, providing more amenities that are complementary with effort to more senior workers is especially valuable in creating an attractive career path, increasing effort early on. On the other hand, the relief a firm feels when an expensive high-seniority worker departs (because $\pi < 0$) reduces the benefit from the immediate effect of the current amenity level in encouraging effort. This last effect can be very large if v_{zz} is small, and so can overwhelm the value of amenities going up with seniority for creating an attractive career path.

4.4 On-the-Job Training

Theories of human capital accumulation serve as the main alternative to deferred payment schemes as the cause of upward-sloping wage-tenure profiles. While many papers assume one or the other, it seems likely that both forces are at work simultaneously. The quantitative effects of each should be measured empirically. To help provide a theoretical framework for such investigations, we extend the basic model to consider the effect of on-the-job training on human capital accumulation and firm investment in training. We consider how expenditures on training will be allocated across a worker's tenure and how human capital will be reflected in the wage profile. A simple and intuitive way of modeling human capital is to assume that human capital increases the effectiveness of effort. The productivity of a worker for a given level of effort will increase with greater human capital.

Let h_t be the stock of human capital accumulated by the worker. Then the firm's expected profit from a newly hired worker is as follows:

$$\max_{\{w_t, T_t\}} \pi_{t_0} = \int_{t_0}^{\infty} e^{-\int_{t_0}^{t} [r_{t'} + a(m, \varphi(h_{t'}) z_{t'})] dt'} [\mu_t \varphi(h_t) z(\Delta_t) - w_t - T_t] dt. \tag{27}$$

The worker's productivity is now a function of effective effort $\varphi(h_t)z_t$, the product of the worker's effort and a function of human capital. This specification assumes that workers with higher human capital levels are more productive for a given level of effort.

T_t is the amount of training the firm provides for the worker at time t, in units of cost to the firm. Training improves a worker's human capital according to the function

$$\dot{h}_t = \vartheta(T_t) - \delta h_t,$$

where $\vartheta(\cdot)$ is a function that converts training, measured by the dollar cost of training, into human capital, with $\vartheta'(\cdot) > 0$ and $\vartheta''(\cdot) < 0$. Also, we assume that $\varphi(\cdot)$ is such that $\varphi(h) < 1$ for any h, in line with $z < 1$ so that $z\varphi(h) < 1$. In other words, φ converges to an asymptote at some level below 1. The attrition rate is now $a(m, \varphi(h_t)z_t) = q + m(1 - \varphi(h_t)z_t)$ which is now governed by effective effort level $\varphi(h_t)z_t$. This specification goes back to the idea implicit in shirking models that shirking can only be detected through the random occurrence of big mess-ups. Those with more training make big mistakes less often for a given level of effort.

Adapting the retrograde approach to this case with the ordinary state variable h as well as the retrograde state variables π and Δ is straightforward. While the direction of time flow is meaningful to the interpretation of these variables, it is irrelevant to the solution of the optimal control problem. The firm's Retrograde Hamiltonian can then be adapted from (15) by including the forward law of motion for h:

$$\begin{aligned}\mathcal{H} =& -\lambda\Big([\mu\varphi(h)z(\Delta) - w - T] - [r + a(m, \varphi(h)z(\Delta))]\pi\Big) \\ & - \zeta\Big([u(w) - v(z(\Delta)) - \psi] - [\rho + a(m, \varphi(h)z(\Delta))]\Delta\Big) + \eta[\vartheta(T) - \delta h]\end{aligned} \tag{28}$$

where η is the co-state variable for the human capital accumulation function. The first-order condition for the worker's wage still follows (16) and the expression for seniority can be solved as,

$$s_t = \int_{t_0}^{t} e^{-\int_{t'}^{t}[\rho - r_{\tilde{t}}]d\tilde{t}}[\mu_{t'} + m\pi_{t'}]\varphi(h_{t'})z_{\Delta,t'}dt', \tag{29}$$

which is identical to (17) except for depending on the marginal effect of *effective* effort.

The path of on-the-job training can be derived much like the wage profile. The first-order condition for training is:

$$\frac{1}{\vartheta'(T_t)} = -\frac{\eta_t}{\lambda_t}. \tag{30}$$

As with wages, the left-hand-side of the equation shows the cost of increasing a unit of human capital h_t and the right-hand-side shows the benefits. Defining $i_t \equiv -\frac{\eta_t}{\lambda_t}$, which we call "the incentive for the firm to invest in training,"

$$i_t = \int_{t}^{\infty} e^{-\int_{t}^{t'}[r_{t''}+\delta+a(m,\varphi(h_{t''})z_{t''})]dt''}[\mu_{t'} + m(\pi_{t'} + s_{t'}\Delta_{t'})]z_{t'}\varphi'(h_{t'})dt'. \tag{31}$$

Greater human capital contributes to every term according to the proportional factor $z\varphi'(h)$: direct production proportional to μ, and a reduction in the attrition rate proportional to m and valued at $\pi + s\Delta$. In steady-state, $\dot{i}_t$ would be zero at values satisfying

$$\bar{i} = \bar{z}\varphi'(\bar{h})\frac{\mu + m(\bar{\pi} + \bar{s}\bar{\Delta})}{r + \delta + m(1 - \bar{z})}.$$

Then, how does optimal training T_t and human capital h_t depend on seniority? The following proposition addresses this question.

Proposition 6. *Assume that the evolution of Δ, π and $\pi + s\Delta$ with worker tenure are monotonic in the direction established by Proposition 2. Then, in steady state, on-the-job training T_t is provided such that, once human capital h_t begins to increase, it continually increases with seniority.*

The proof for Proposition 6 is provided in appendix B.2. To gain intuition behind this result, it helps to first consider the case with the future path of human capital h fixed. Given a fixed level of h, training senior workers more is a better deal for firms because (a) they work harder, (b) because they are less likely to become separated from the firm with the attendant loss of the human capital they gain from training and (c) because, in accordance with the "time-consistent objective," the reduction in attrition caused by the training itself (as those workers make big blunders less often), is treated as more valuable the more senior a worker is. Of course, the integral expression for i_t in (31) also shows that, given a concave φ, the incentive to provide training goes down with the amount of human capital a worker already has. So the actual rate of training can be higher for younger workers.

However, we can show that in steady state, *despite human capital being subject to depreciation*, once human capital begins to increase, it must continue to increase thereafter (to an asymptote). Intuitively, this is because firms will provide more training to more senior workers if the level of human capital is the same, meaning that it is impossible for the junior worker to overtake the senior worker once he falls behind. The possibility of an initial decrease in human capital is hard to avoid if a worker is hired with a large amount of human capital and training is not cost effective. If human capital is zero at hiring, then this condition is satisfied (at least weakly) at the moment of hiring.

5 Conclusion

In this paper, we find that when firms cannot monitor worker effort perfectly, and can't ask for performance bonds, they will want to pay wages that reflect worker seniority. Seniority is accumulated in a manner that reflects the value of the worker to the firm and can have an important effect on the wage-tenure profile. When seniority matters like this, the loss in seniority from job loss leads to large and persistent decreases in life-time earnings.

Extending the model, we find that when monitoring intensity can vary, firms have many reasons to monitor junior workers more intensively than senior workers. When amenities can be provided, the effort elicitation problem makes amenities that reduce the marginal disutility of effort become especially attractive. For comparison, when effort is observable, in which case only the effect of amenities on the level of disutility at the level of effort required by the firm factors into their provision. Finally, when on-the-job-training is brought into the model, firms will invest in on-the-job training in a way that guarantees that firm-specific human capital increases with seniority, despite that human capital being subject to depreciation.

The theoretical analysis in this paper points to many qualitative predictions that, with real ingenuity in finding the right data, can be checked out empirically. Moreover, the theoretical framework here suggests how to set up a structural model of workers' tenure with firms. Both of these efforts are beyond the scope of this paper. We hope we have set the stage for such work, and have provided insights that aid understanding of the real world even now.

References

Abraham, Katharine G., and Henry S. Farber, 1987. "Job Duration, Seniority, and Earnings." American Economic Review, 77(3): 278-297

Abeler, Johannes, Armin Falk, Lorenz Goette, and David Huffman, 2011. "Reference Points and Effort Provision," American Economic Review, 101: 470-492

Acabbi, Edoardo Maria, Andrea Alati, and Luca Mazzone, 2022. "A Labor Market Sorting Model of Hysteresis and Scarring," Working paper

Akerlof, George A., and Lawrence F. Katz, 1989. "Workers' Trust Funds and the Logic of Wage Profiles," Quarterly Journal of Economics, 104(3): 525-536

Altonji, Joseph G. and Robert A. Shakotko, 1987. "Do Wages Rise with Job Seniority?," The Review of Economic Studies, 54(3): 437-459

Altonji, Joseph G., and Nicolas Williams, 2005. "Do Wages Rise with Job Seniority? A Reassessment," Industrial and Labor Relations Review, 58(3): 370-397

Athey, Susan, 2002. "Monotone Comparative Statics under Uncertainty," Quarterly Journal of Economics, 117(1): 187-223

Barth, Erling, 1997. "Firm-Specific Seniority and Wages," Journal of Labor Economics, 15(3): 459-506

Basu, Susanto, and Miles S. Kimball, 1997. "Cyclical productivity with unobserved input variation," NBER Working Paper no. 5915.

Basu, Susanto, John Fernald and Miles S. Kimball, 2006. "Are Technology Improvements Contractionary?" American Economic Review, 96: 1418-1448.

Blackburn, McKinle, and David Neumark, 1992. "Unobserved Ability, Efficiency Wages, and Interindustry Wage Differentials," Quarterly Journal of Economics, Vol.107(4): 1421-

1436

Buchinsky, Moshe, Denis Fougere, Francis Kramarz and Rusty Tchernis, 2010. “Interfirm Mobility, Wages and the Returns to Seniority and Experience in the United States,” Review of Economic Studies, 77: 972-1001

Buhai, I. Sebastian, Miguel A. Portela, Coen N. Teulings, and Aico Van Vuuren, 2014. “Returns to Tenure or Seniority?” Econometrica, Vol. 82(2): 705-730

Bulow, Jeremy I., and Lawrence H. Summers, 1986. “A Theory of Dual Labor Markets with Application to Industrial Policy, Discrimination, and Keynesian Unemployment,” Journal of Labor Economics, 4(3): 376-414

Burdett, Kenneth, Carlos Carrillo-Tudela, and Melvyn Coles, 2020. “The Cost of Job Loss,” Review of Economic Studies, 87(4): 1757-1798

Burdett, Kenneth, and Melvyn Coles, 2010. “Tenure and Experience Effects on Wages: A Theory,” CESifo Working Paper No.3203

Cappelli, Peter, and Keith Chauvin, 1991. “An Interplant Test of the Efficiency Wage Hypothesis,” Quarterly Journal of Economics, Vol.106(3): 769-787

Couch, Kenneth A. and Dana W. Placzek, 2010. “Earnings Losses of Displaced Workers Revisited,” American Economic Review, 100(1): 572-589

Davis, Steven J., and Till von Wachter, 2011. “Recessions and the Costs of Job Loss,” Brookings Papers on Economic Activity, 43(2): 1-72

Daniel, Kirsten, and John S. Heywood, 2007. “The Determinants of Hiring Older Workers: UK Evidence,” Labour Economics, 14: 35-51

Dellavigna, Stefano, and Devin Pope, 2018. “What Motivates Effort? Evidence and Expert Forecasts,” Review of Economic Studies, 85: 1029-1069

Dustmann, Christian, and Costas Meghir, 2005. "Wages, Experience and Seniority," Review of Economic Studies, 72: 77-108

Epstein, Brendan, and Miles S. Kimball, 2014. "The Decline of Drudgery and the Paradox of Hard Work," International Finance Discussion Papers no. 1106. Board of Governors of the Federal Reserve System

Fallick, Bruce C., 1996. "A Review of the Recent Empirical Literature on Displaced Workers," Industrial and Labor Relations Review, 50(1): 5-16

Heywood, John, Uwe Jirjahn, and Georgi Tsertsvardze, 2010. "Hiring Older Workers and Employing Older Workers: German Evidence," Journal of Population Economics, 23(2): 595-615

Huck, Steffen, Andrew J. Seltzer, and Brian Wallace, 2011. "Deferred Compensation in Multiperiod Labor Contracts: An Experimental Test of Lazear's Model," American Economic Review, Vol.101: 819-843

Hutchens, Robert M., 1989. "Seniority, Wages and Productivity: A Turbulent Decade," Journal of Economic Perspectives, 3(4): 49-64

Jacobson, Louis, Robert LaLonde, and Daniel Sullivan. 1993. "Earnings Losses of Displaced Workers," American Economic Review 83(4): 685-709

Jarosch, Gregor, forthcoming. "Searching for Job Security and the Consequences of Job Loss," Econometrica

Jung, Philip, and Moritz Kuhn. 2019. "Earnings Losses and Labor Mobility over the Life Cycle," Journal of the European Economic Association, 17(3): 678-724

Kamien, Morton I., and Nancy L. Schwartz, 1991. "Dynamic Optimization: The Calculus of Variations and Optimal Control in Economics and Management," Amsterdam, North-

Holland: Elsevier Science; 2nd edition

Kimball, Miles S., 1994. “Labor Market Dynamics when Unemployment is a Worker Discipline Device,” American Economic Review, 84(4): 1045-1059

Kletzer, Lori G. 1998. “Job Displacement,” Journal of Economic Perspectives, 12(1): 271-304

Krolikowski, Pawel, 2017. “Job Ladders and Earnings of Displaced Workers,” American Economic Journal: Macroeconomics, 9(2): 1-31

Krueger, Alan B., and Lawrence H. Summers, “Efficiency Wages and the Inter-Industry Wage Structure,” Econometrica, Vol.56(2): 259-293

Kugler, Adriana D., 2003. “Employee Referrals and Efficiency Wages,” Labour Economics, Vol.10: 531-556

Kwon, Illoong, and Eva M. Meyersson Milgrom, 2014. “The Significance of Firm and Occupation Specific Human Capital in Hiring and Promotion,” Labour Economics, 31: 162-173

Lachowska, Marta, Alexandre Mas, and Stephen A. Woodbury, 2020. “Sources of Displaced Worker’s Long-Term Earnings Losses,” American Economic Review, 110(10): 3231-3266

Lazear, Edward P., 1979. “Why is there Mandatory Retirement?” Journal of Political Economy, 87(6): 1261-1284

Lazear, Edward P., 1981. “Agency, Earnings Profiles, Productivity, and Hours Restrictions,” American Economic Review, 71(4): 606-620

Lazear, Edward P., 2009. “Firm-Specific Human Capital: A Skill-Weights Approach,” Journal of Political Economy, 117(5): 914-940

Milgrom, Paul and Chris Shannon, 1994. "Monotone Comparative Statics," Econometrica, 62(1): 157-180

Neal, Derek, 1995. "Industry-Specific Human Capital: Evidence from Displaced Workers," Journal of Labor Economics, 13(4): 653-677

Pavan, Ronni, 2011. "Career Choice and Wage Growth," Journal of Labor Economics, 29(3): 549-587

Shapiro, Carl, and Joseph E.Stiglitz, 1984. "Equilibrium Unemployment as a Worker Discipline Device," American Economic Review, 74(3): 433-444

Shimer, Robert, 2012. "Reassessing the Ins and Outs of Unemployment," Review of Economic Dynamics, 15(2): 127-148

Stevens, Margaret, 2003. "Earnings Functions, Specific Human Capital, and Job Matching: Tenure Bias is Negative," Journal of Labor Economics, 21(4): 783-805

Strobl, Eric, and Frank Walsh, 2007. "Estimating the Shirking Model with Variable Effort," Labour Economics, 14: 623-637

Topel, Robert, 1991. "Specific Capital, Mobility, and Wages: Wages Rise with Job Seniority," Journal of Political Economy, 99(1): 145-175

Vives, Xavier, 2000. "Oligopoly Pricing: Old Ideas and New Tools," Cambridge, MA: MIT Press.

Zivin, Joshua, and Matthew Neidell, 2012. "The Impact of Pollution on Worker Productivity," American Economic Review, 102(7): 3652-3673

Zwick, Thomas, 2011. "Seniority Wages and Establishment Characteristics," Labour Economics, 18: 853-861

Appendix A The Steady State Equilibrium

A.1 Equilibrium

The equilibrium of this model consists of the path of endogenous variables

$$\Big\{\{w_t(\tau)\}_\tau, \{z_t(\tau)\}_\tau, \{\Delta_t(\tau)\}_\tau, \{s_t(\tau)\}_\tau, \{\pi_t(\tau)\}_\tau, \mu_t, \psi_t\Big\}_t$$

where $w_t(\tau), z_t(\tau), \Delta_t(\tau), s_t(\tau), \pi_t(\tau)$ denote wage, effort, utility differential, seniority, and expected profit at time t from a worker hired at time τ. The variables must satisfy equations (4), (12), (16), (17) at each time t for all cohorts τ, as well as conditions,

$$\pi_t(\tau) = \int_t^\infty e^{-\int_t^{t'} [r_{\tilde{t}} + a(m, z_{\tilde{t}})] d\tilde{t}} [\mu_{t'} z(\Delta_{t'}(\tau)) - w_{t'}(\tau)] dt' \tag{A.1}$$

$$\mu_t = p_t f'(Z_t) \tag{A.2}$$

$$\psi_t = \bar{u} + \frac{\theta_t}{N_t - L_t} \Delta_{0,t} \tag{A.3}$$

where total effective labor is defined as $Z_t \equiv \int_{-\infty}^t z_t(\tau) l_t(\tau) d\tau$, total labor supply $L_t \equiv \int_{-\infty}^t l_t(\tau) d\tau$, and $l_t(\tau)$ denotes the number of workers hired at time τ that are employed by the firm at time t. The number of workers a firm hires $l_t(t)$ is determined by the labor market clearing condition $\pi_{0,t} = 0$. As long as the marginal profit of a newly hired worker is positive, the firm will hire more workers. If the marginal profit is negative, there will be no hiring. Thus for markets to clear in equilibrium, the marginal profit of a newly hired worker must be zero as long as there is positive hiring.

A.2 The Steady State

The equilibrium dynamics of this model can be quite complex and beyond the scope of this paper.[20] Instead we focus on the steady state equilibrium and comparative statics to gain some insight into the general equilibrium properties of this model.

In steady state, aggregate variables $(\psi, \mu, r, \theta, \pi_0, \Delta_0, Z, L)$ must be constant. In addition, the steady state involves a constant flow of new hirings balancing separations due to firings and exogenous quits. The employment rate must equal the attrition rate,

$$\theta_t = \int_{-\infty}^{t} a(m, z_t(\tau)) l_t(\tau) d\tau = \bar{a}_t L_t$$

where $\bar{a}_t \equiv \frac{\int a(m,z_t(\tau))l_t(\tau)d\tau}{\int l_t(\tau)d\tau}$.

The steady state equilibrium does not imply that the contract is at the stationary point. The solution to $\dot{\Delta} = \dot{s} = \dot{\pi} = 0$ visible in Figures 1, 2a and 2b does not describe the steady state, but only the eventual condition of employees who by good fortune remain with the firm for a long time. Since new workers are constantly being hired, much of the employed labor force will always be far from this stationary point. Rather, the steady state of the system involves constant motion out of unemployment and motion toward this stationary point coupled with attrition into unemployment.

Two key variables ψ and μ determine the steady state path of wages, effort, and utility. The flow benefit of unemployment ψ, as the opportunity cost of working, is counterposed to the wage. Higher ψ increases the attractiveness of being unemployed and lowers the cost of job loss to the worker, and all else equal, decreases labor supply. The marginal product of effort μ, on the other hand, is critical in determining the firm's demand for labor as it governs the firm's expected profit from workers. Furthermore, μ can be expressed as a function of

[20] See Kimball(1994) for an exposition of the dynamics of a basic efficiency wage model.

total employment L as $\mu = pf'(\bar{z}L)$ where $\bar{z}$ is the average effort of the firm's employees $\bar{z}_t \equiv \frac{\int z_t(\tau) l_t(\tau) d\tau}{\int l_t(\tau) d\tau}$.

We can then express the steady state equilibrium on the (ψ, L) plane in terms of appropriate counterparts of supply and demand curves in the labor market. The labor supply curve of workers is governed by the pseudo-utility of unemployment ψ. The supply of labor to the firm is horizontal, since ψ is exogenous to the firm. Thus, ψ is analogous to the market wage in a competitive labor market. However, the market supply curve for labor is not horizontal. In steady state,

$$\psi = \bar{u} + \frac{\bar{a}L}{N-L}\Delta_0$$

where $\bar{a}_t \equiv \frac{\int a(m, z_t(\tau)) l_t(\tau) d\tau}{\int l_t(\tau) d\tau}$ is the weighted average attrition rate of labor as previously described. This curve is upward sloping in general, the more so if the attrition rate $\bar{a}$ increases with ψ. For the supply curve to be downward sloping, it would require $\bar{a}$ to be a strong inverse function of ψ, but other things equal, ψ raises attrition as workers exert less effort due to less fear of job loss.[21] Barring such perversities, the supply curve will be upward sloping. The curve will become very steeply sloped as total labor employed approaches N.

Labor demand by firms is governed by the equilibrium condition for the expected profit from newly hired workers $\pi_0 = 0$ (there must be positive hiring in steady state). From the firm's perspective, a higher level of ψ lowers the cost of job loss for the worker and the firm must either accept lower effort from workers or raise wages. Either way the profits expected from a worker suffer and the expected present value of hiring a new worker falls unless the marginal revenue product of effort μ rises to compensate for the reduced effort or increased wages of a worker.

The relationship between the total number of employed workers L and μ is ambiguous.

[21] Shimer(2012) finds that employment exit probability is largely acyclical.

At a cursory glance it would seem that the total number of workers L should decrease in order for μ – a function of marginal productivity of total effective effort $f'(Z)$ – to increase. However, since Z is the product of the average effort of firm's employees $\bar{z}_t = \frac{\int z_t(\tau) l_t(\tau) d\tau}{\int l_t(\tau) d\tau}$ multiplied by L, it is possible that a fall in average effort might increase μ by enough to force the firm to increase the number of workers in order to make up for less effort per worker.

For example, consider the following. Suppose that the interest rate r is zero so that the cross-section of workers gives the same relative weights to different seniorities as does the integral giving the expected present value of the career of a newly hired worker. Substituting in steady state values in equation (13) and imposing the condition $\pi_0 = 0$ we arrive at,

$$\bar{w} = \bar{z}\mu = \bar{z}pf'(\bar{z}L) \tag{A.4}$$

where $\bar{w}$ and $\bar{z}$ are both present value averages and cross-sectional averages since $r = 0$. Writing (A.4) in elasticity terms,

$$\%\Delta\bar{w} = \%\Delta\bar{z} + \%\Delta\bar{p} - \gamma[\%\Delta\bar{z} + \%\Delta L] \tag{A.5}$$

where $\gamma = -\frac{f''(Z)Z}{f'(Z)}$, measures the degree of diminishing returns to labor input. Solving (A.5) for $\%\Delta L$, we obtain

$$\%\Delta L = \frac{1}{\gamma}[(1-\gamma)\%\Delta\bar{z} + \%\Delta\bar{p} - \%\Delta\bar{w}] \tag{A.6}$$

Given (A.6), consider the following. Suppose that $f(Z) = Z^{\alpha}$ where $\alpha = 0.7$ implying $\gamma = 0.3$. Now, suppose that an exogenous increase in ψ leads to a 3% decrease in effort $\bar{z}$. Then a 3% increase in wage $\bar{w}$ and a corresponding 4% increase in L will be consistent with the steady state labor demand condition (A.4).

As can be seen from the foregoing example, while the graph of π_0 in terms of μ against ψ

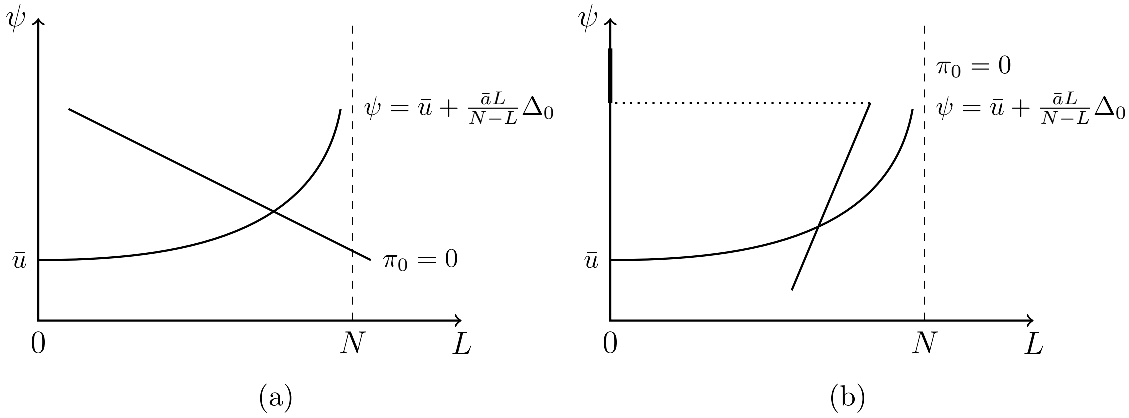


Figure 6: Steady state equilibrium

is upward sloping and the graph of ψ against Z is downward sloping, the graph of $\pi_0 = 0$ in $L-\psi$ space can be either upward or downward sloping. If effort is only moderately responsive to changes in ψ, $\bar{z}$ is relatively constant and L will decrease, resulting in a downward-sloping curve. However, if effort is highly responsive to changes in ψ, $\bar{z}$ will fall sharply and L must increase, resulting in an upward-sloping curve.

A downward-sloping curve corresponds to the usual intuition, since this curve is in effect the firm's demand curve for labor. If the number of workers and the amount of effective labor move in opposite direction, as in the example above, the demand curve will be upward sloping. One aspect worth noting is that the demand curve in $\psi-L$ space must be downward sloping for high levels of ψ, since sufficiently high levels of ψ would make it unprofitable to employ any workers. This could be a continuous function of L but there may be a discontinuous jump from positive to zero L as ψ rises. L can continue to decrease indefinitely as ψ decreases since a firm may be sated with a certain amount of effective labor so that any increases in the effort per worker may only reduce the number of workers needed.

The market demand and supply of labor can now be combined. Note that the supply curve is the analog of Shapiro and Stiglitz'(1984) "No Shirking Condition," while the demand curve is the analog to their "Marginal Product of Labor" schedule. Panel (a) of Figure 6 shows the standard case in which the demand is downward sloping. Panel (b) illustrates the upward sloping demand curve in the case where at high levels of ψ the labor demand suddenly drops to zero.

The comparative statics in the steady state can be analyzed from supply and demand. An increase in N will move the $\psi = \bar{u} + \frac{\bar{a}L}{N-L}\Delta_0$ curve out, while a reduction in $\bar{u}$ will move it down. Therefore, both changes increase employment in Panel (a) of Figure 6 and decrease employment in Panel (b), reducing ψ in both cases. An improvement in productive technology that increases the marginal product of labor will move the demand curve outward. A reduction in the exogenous quit rate, an improvement in monitoring technology or exogenous worker motivation as reflected in worker utility functions will shift the $\pi_0 = 0$ curve out in $\psi - \mu$ space and thereby cause an increase in demand in $\psi - L$ space. Any increase in demand causes both L and ψ to rise.

In the intuitive case of a downward-sloping labor demand curve, the supply curve will be the main determinant of the equilibrium unemployment rate due to the steep slope of the supply curve for reasonable levels of unemployment.[22] Therefore, the main effect of shifts in the demand curve will be on equilibrium ψ, with very little effect on equilibrium L. As a result, the effects of changes in the exogenous quit rate, monitoring technology or worker utility functions on unemployment through the supply curve overshadow their effects through the demand curve.

For example, if two distinct worker groups have different exogenous quit rates but the same utility functions, the implied lower work effort for one group will have only a small effect on the unemployment rate. The primary factor leading to a higher unemployment rate for this group would be that labor discipline requires about the same average spell of unemployment for both groups, but the flow into unemployment would be greater for the

[22]For example, the slope of the supply curve will equal $a\Delta_0 \frac{N}{(N-L)^2} = 100\bar{a}\Delta_0 N$ for 10% unemployment.

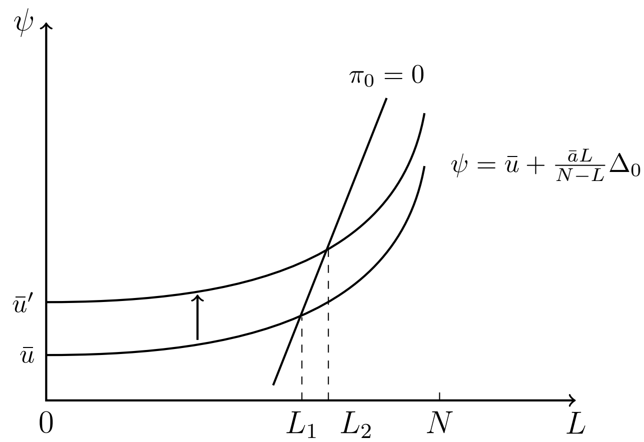


Figure 7: Increase in unemployment benefits

group with the higher attrition rate, resulting in the percentage of persons unemployed at any time to be greater.

Finally, as described above, the shape of the labor demand curve can potentially have important implications on the comparative statics of the labor market. This can in turn lead to interesting policy implications. For example, if effort is very responsive to changes in ψ, the slope of the demand curve will become steeper and steeper and may even imply that the demand curve is upward sloping. A steep demand curve can imply that increased unemployment benefits will have very little effect on the unemployment rate and may even lead to a decrease in the unemployment rate if the curve is upward sloping. This is because an increase in unemployment benefits will increase $\bar{u}$ which would shift the supply curve upwards. The equilibrium values of ψ and L will move along the demand curve as depicted in Figure 7 for the upward-sloping case. Nevertheless, the analysis shows that it is unclear whether this is a desirable outcome as the increase in employment is driven by a large decrease in effort and therefore productivity.

Appendix B Proofs

B.1 Proof of the concavity of the (Δ, π) Efficient Frontier

Because the case with exogenous monitoring intensity can be shown as a relatively straightforward extension of the endogenous monitoring case, we first prove the case in which firms choose the optimal level of monitoring intensity every instant in time.

B.1.1 The Case with Endogenous Monitoring Intensity

We prove the convexity of the feasible set by recursively going back in time in a discrete-time model that limits into the continuous-time firm problem. Note that any discrete-time model that limits into the continuous-time firm problem is suitable for this proof. (If there were any extra solutions in the continuous-time model, we impose the selection criterion of selecting among solutions that are limits of the solutions to the discrete-time model.) Thus, we have some flexibility in exactly how we set up the functional forms and the timing in the discrete-time model. Denote dt as the length of a period and take the limit as the finite horizon becomes infinite: $T \to \infty$. Our argument will be based on what happens when dt gets very small, showing that there is a uniform discrete-time grid size dt small enough the (Δ, π) Efficient Frontier can't be at least weakly concave at time t and fail to be even weakly concave at time $t - dt$.

Definitions

Denote the (Δ_t, π_t) Feasible Set at time t, given period length dt, and terminal time T, by: $\Omega_{t;dt,T}$. Use the notation $x = m\Delta$ (or equivalently, $m = \frac{x}{\Delta}$) and define

$$Z(x) = v'^{-1}(x)$$

so that

$$z_t = Z(m_t \Delta_t).$$

Then

$$Z'(x) = \frac{1}{v''(v'^{-1}(x))} = \frac{1}{v''(Z(x))} > 0;$$

$$Z''(x) = \frac{d}{dx}\Big(\frac{1}{v''(v'^{-1}(x))}\Big) = \frac{-v'''(v'^{-1}(x))}{[v''(v'^{-1}(x))]^3} < 0.$$

Next, define

$$\begin{aligned}
\mathcal{F}(\Delta_t, x_t, w_t; dt, T) &= \{1 + [\frac{x_t}{\Delta_t} Z(x_t) - \frac{x_t}{\Delta_t} - \rho - q]dt\}\Delta_t + [u(w_t) - v(Z(x_t)) - \psi]dt \\
&= [1 - (\rho + q)dt]\Delta_t - x_t[1 - Z(x_t)]dt + [u(w_t) - v(Z(x_t)) - \psi]dt \\
\mathcal{G}(\Delta_t, \pi_t, x_t, w_t; dt, T) &= \{1 + [\frac{x_t}{\Delta_t} Z(x_t) - \frac{x_t}{\Delta_t} - r - q]dt\}\pi_t + [\mu Z(x_t) - w_t - c(\frac{x_t}{\Delta_t})]dt \\
&= [1 - (r + q)dt]\pi_t - \frac{\pi_t}{\Delta_t} x_t[1 - Z(x_t)]dt + [\mu Z(x_t) - w_t - c(\frac{x_t}{\Delta_t})]dt
\end{aligned}$$

so that

$$\Delta_{t-dt} = \mathcal{F}(\Delta_t, x_t, w_t; dt, T)$$

and

$$\pi_{t-dt} = \mathcal{G}(\Delta_t, \pi_t, x_t, w_t; dt, T).$$

Define

$$\begin{aligned}
\hat{\Pi}_{t-dt}(\Delta_{t-dt}) = &\max_{\Delta_t, \pi_t, x_t} \mathcal{G}(\Delta_t, \pi_t, x_t, \hat{w}_t; dt, T) \qquad \text{(B.1)} \\
s.t. \qquad &(\Delta_t, \pi_t) \in \Omega_{t;dt,T} \\
&\mathcal{F}(\Delta_t, x_t, \hat{w}_t; dt, T) \geq \Delta_{t-dt}.
\end{aligned}$$

From a given point (Δ_t, π_t) next period, and given $\hat{w}_t$, the optimization problem for x can be written out as:

$$\max_x [1-(r+q)dt]\pi_t - \frac{\pi_t}{\Delta_t} x[1-Z(x)]dt + [\mu Z(x) - \hat{w}_t - c(\frac{x}{\Delta_t})]dt$$
$$s.t. \qquad [1-(\rho+q)dt]\Delta_t - x[1-Z(x)]dt + [u(\hat{w}_t) - v(Z(x)) - \psi]dt \geq \Delta_{t-dt}.$$

The Lagrangian is:

$$L = [1-(r+q)dt]\pi_t - \frac{\pi_t}{\Delta_t} x[1-Z(x)]dt + [\mu Z(x) - \hat{w}_t - c(\frac{x}{\Delta_t})]dt$$
$$+ S\big[[1-(\rho+q)dt]\Delta_t + x[Z(x)-1] - (\rho+q)\Delta_t + u(\hat{w}) - v(Z(x)) - \psi dt - \Delta_{t-dt}\big].$$

By the envelope theorem,

$$S = -\hat{\Pi}'_{t-dt}(\Delta_{t-dt}).$$

We choose capital "S" for the Lagrange multiplier because due to the additive separability of the effect of the wage, this slope of the possibility frontier for a given $\hat{w}_t$ must be equated to the overall slope of the possibility frontier, which in turn must be equated to seniority s in the firm's overall optimization problem.

The first-order condition for x, dividing through by dt, and using the worker's first-order condition, $v'(Z(x)) = x$, is:

$$\mu Z'(x) - \frac{\pi_t}{\Delta_t}[1 - Z(x) - xZ'(x)] - S[1-Z(x)] - \frac{1}{\Delta_t} c'(\frac{x}{\Delta_t}) = 0 \qquad \text{(FOCx)}$$

Assumptions

On the shape of $u(w)$, on the domain of non-negative reals, we assume $u'(w) > 0, u''(w) < 0$, and the Inada conditions $\lim_{w\to 0} u'(w) = +\infty$ and $\lim_{w\to\infty} u'(w) = 0$. We assume $u(0) > -\infty$ and for convenience normalize $u(0) = 0$. We assume $\bar{u} \geq u(0) = 0$. As for c, on the domain of the non-negative reals, we assume only that $c(\cdot) \geq 0$, $c'(\cdot) > 0$ and $c''(\cdot) > 0$.

On the shape of $v(z)$, on the domain of the non-negative reals, we assume $v(0) \geq 0$, $v'(\cdot) > 0$, $v''(\cdot) > 0$ (including $v'(0) > 0, v''(0) > 0$), and $\lim_{z\to 1} v'(z) = +\infty$ (which implies $1 - Z(x) > 0$). We also assume

$$(1 - z)\frac{v''(z)}{v'(z)}$$

is increasing in z, or equivalently, that

$$\frac{xZ'(x))}{1 - Z(x)} \tag{B.2}$$

is decreasing in x. This implies that:

$$-\frac{xZ''(x)}{Z'(x)} = \frac{v'''(Z(x))v'(Z(x))}{v''(Z(x))^2} \geq 1 + \frac{xZ'(x)}{1 - Z(x)} > 1.$$

Defining

$$\mathbb{H}(x) = x(1 - Z(x)) > 0,$$

we assume

$$\mathbb{H}'(x) = 1 - Z(x) - xZ'(x) \geq 0, \tag{B.3}$$

or equivalently,

$$\frac{xZ'(x)}{1 - Z(x)} \leq 1.$$

Given $x = v'(Z(x))$ and the derivatives of $Z = v'^{-1}(x)$ computed above, this assumption is equivalent to:

$$\frac{d}{dx}v'(Z(x))(1 - Z(x)) = Z'(x)[v''(Z(x))(1 - Z(x)) - v'(Z(x))] \geq 0,$$

or as an assumption directly on the shape of $v(z)$,

$$(1-z)v''(z) - v'(z) \geq 0, \quad \forall z \in [0,1) \qquad \Longleftrightarrow \qquad \frac{(1-z)v''(z)}{v'(z)} \geq 1, \quad \forall z \in [0,1).$$

Endogenizing the wage. The wage choice will be handled in the following way: prove concavity for a fixed wage at time $t, w_t = \hat{w}$, and use the argument that the set sum of two convex sets is convex to get concavity of the the (Δ, π) Efficient Frontier given an optimal wage choice. This works because the effect of $(\Delta_t, \pi_t, m_t, w_t)$ on $(\Delta_{t-dt}, \pi_{t-dt})$ is additively separable between w_t and (Δ_t, π_t, m_t).

That is, to establish the concavity of the (Δ, π) Efficient Frontier at $T - dt$ simply note that the set sum of two convex sets is convex. (The sum of two sets is the set of all points that can result from vector addition of a point from one set and a point from the other.) The two relevant convex sets are (A) the set of points vector-dominated by $\hat{\pi}_{t-dt} = \hat{\Pi}_{t-dt}(\Delta_{t-dt}; dt, T)$ and (B) the set of points vector-dominated by the set of points of the form $dt(u(w) - u(\hat{w}), -[w - \hat{w}])$. The part of the northeast boundary of the convex set A+B satisfying $\Delta_{t-dt} \geq 0$ and $\pi_{t-dt} \geq -\frac{\mu\Delta}{x}$ is the (Δ, π) Efficient Frontier at $T - dt$, which therefore is concave.

Backward induction. The strategy of considering a finite-horizon and discrete-time and then taking the limit into continuous time and finally the limit into an infinite horizon enables us to make a recursive proof of the concavity (and downward slope) of the (Δ, π) Efficient Frontier, starting at the end of time. First, consider what happens in the last two periods. If $t = T$, effort at time T is zero since there is no future to motivate the worker, with $\Delta_T = 0$, and also $\pi_T = 0$, optimal monitoring is zero. Then the curve $(\Delta_{T-dt}, \hat{\Pi}_{T-dt}(\Delta_{t-dt}))$

is determined parametrically by:

$$\Delta_{T-dt} = [u(\hat{w}) - v(0) - \psi]dt$$
$$\pi_{T-dt} = -[\hat{w} + c(0)]dt.$$

That is, it is a single point. Set addition of this single point with the subgraph of $(u(w) - u(\hat{w}), -(w - \hat{w}))$ implies—at the northeast boundary of that set sum—a very well-behaved (Δ, π) Efficient Frontier at time $T - dt$. All proof steps below assume that we are talking about periods strictly earlier than $T - dt$.

Next, consider the following lemma.

Lemma 1. *Recall that $u'' < 0$ over the positive reals. This implies that, for any time t, the slope of the (Δ, π) Efficient Frontier at time $t - dt$, at any point that is locally concave, is equal to $-\frac{1}{u'(w)} < 0$.*

Proof. Let $(\Delta^*_{t-dt}, \pi^*_{t-dt})$ denote any point one can attain from some given (Δ_t, π_t). Then, at time $t - dt$, one can attain at least the profit $\pi^*_{t-dt} + \epsilon \cdot dt$ at the cost of reducing the utility differential to $\Delta^*_{t-dt} - u'(w)\epsilon \cdot dt + o(\epsilon \cdot dt)$ by reducing the current wage by ϵ. By similar reasoning, one can attain at least the utility differential $\Delta^*_{t-dt} + u'(w)\epsilon \cdot dt + o(\epsilon \cdot dt)$ at the cost of reducing profit by $\pi^*_{t-dt} - \epsilon \cdot dt$ by increasing the current wage by ϵ. Taking the limit as $\epsilon \to 0$ and using the local concavity at the point being considered on the (Δ, π) Efficient Frontier at $t - dt$, implies that the slope is $-\frac{1}{u'(w)}$. Since u' is positive the result follows. □

Extension of Lemma 1. *$\hat{\Pi}_{t-dt}$ is differentiable wherever it is locally concave.*

Proof. The optimal contract must have a strictly positive wage at some future date. Unless $t = T$ (the case handled above), then even holding w_t at $\hat{w}$, there is then a future wage we can consider varying in each direction following the model of the proof of Lemma 1, which, combined with local concavity, implies differentiability.

Lemma 2. *For any time t and period length dt,* $\Delta_{t-dt} \leq \left(\frac{1-(1-rdt)^{\frac{T+dt-t}{dt}}}{r}\right) u\left(\frac{-r\pi_t}{1-(1-rdt)^{\frac{T+dt-t}{dt}}}\right)$.

Proof. Note the timing convention in our equations: $u(w_t)$ and w_t are not discounted in how they affect Δ_{t-dt} and π_{t-dt} respectively. Restricting ourselves to period lengths that divide the remaining time equally, the constrained maximum,

$$\begin{aligned} \max_{\{w_{t+ndt}\}_n} \quad & \sum_{n=0}^{\frac{T-t}{dt}} [(1-rdt)^n u(w_{t+ndt})] dt \\ s.t. \quad & \sum_{n=0}^{\frac{T-t}{dt}} [(1-rdt)^n w_{t+ndt}] dt \leq -\pi_t, \end{aligned}$$

is achieved with a constant wage w. The maximum for this simplified problem is the quantity given in Lemma 2. Now consider the actual situation in the model. Since $\rho > r$ and there is attrition, the discounting is more severe, which reduces the maximum, given the non-negativity of $u(w)$ and of w. And there are subtractions from at least $\bar{u} \geq 0$ and from $v(z) \geq 0$ that also reduce the maximum. Finally, the non-negative costs of monitoring make the constraint tighter and so would also reduce the maximum. $\square$

Corollary to Lemma 2: Given any downward-sloping line, the function $\hat{\Pi}_{t-dt}(\Delta_{t-dt})$ must go below that line as $\Delta_{t-dt} \rightarrow +\infty$.

Proof: Since the option to pay a wage different from $\hat{w}$ has to be valuable, the inequality $\left(\frac{1-(1-rdt)^{\frac{T+dt-t}{dt}}}{r}\right) u\left(\frac{-r\pi_t}{1-(1-rdt)^{\frac{T+dt-t}{dt}}}\right)$ for the (Δ, π) Efficient Frontier at time $t - dt$ has to imply at least as tight an inequality for $\hat{\Pi}_{t-dt}(\Delta_{t-dt})$:

$$\hat{\Pi}_{t-dt}^{-1}(\pi_{t-dt}) \leq \Big(\frac{1-(1-rdt)^{\frac{T+dt-t}{dt}}}{r}\Big)u\Big(\frac{-r\pi_t}{1-(1-rdt)^{\frac{T+dt-t}{dt}}}\Big). \tag{B.4}$$

Thus, $\hat{\Pi}_{t-dt}(\Delta_{t-dt})$ is bounded on the right by a curve whose slope tends to $-\infty$ as $\Delta_{t-dt} \rightarrow +\infty$.

The Corollary to Lemma 2 is important because it means that geometrically, if $\hat{\Pi}_{t-dt}(\Delta_{t-dt})$ fails to be weakly concave globally, then there must be at least one pair of points $(\Delta^{\dagger}, \hat{\Pi}_{t-dt}(\Delta^{\dagger}))$ and $(\Delta^{\ddagger}, \hat{\Pi}_{t-dt}(\Delta^{\ddagger}))$ with $\Delta^{\ddagger} > \Delta^{\dagger}$ satisfying the following properties. Both points must be at a locally concave part of the curve, share a tangent line, and points at all the values of $\Delta_{t-dt} \in (\Delta^{\dagger}, \Delta^{\ddagger})$ must be strictly below that shared tangent line (see the example in Figure 8). The inequality $\hat{\Pi}'_{t-dt}(\Delta_{t-dt}) \leq 0$, which follows from the principle that relaxing a constraint of the form on $\Delta \geq \hat{\Delta}$ makes at least as high a π possible in the maximization problem in (B.1), guarantees that the slope of this common tangent cannot be zero, when combined with the existence of a point on the efficient frontier below the line segment between $(\Delta^{\dagger}, \hat{\Pi}_{t-dt}(\Delta^{\dagger}))$ and $(\Delta^{\ddagger}, \hat{\Pi}_{t-dt}(\Delta^{\ddagger}))$.

Proof by Contradiction. To demonstrate the concavity of $\hat{\Pi}_{t-dt}(\Delta_{t-dt})$ we proceed by contradiction. If, in the limit as $dt \rightarrow 0$, the (Δ, π) Efficient Frontier ever fails to be weakly concave, then there must be some t for which $\hat{\Pi}_{t-dt}$ fails to be weakly concave even for very small dt.[23]

Let's focus on the pair of this type that has the lowest value of $\Delta^{\dagger}$. In the proof by

[23]Note that because weak concavity is defined by weak inequalities, in an interval for τ that in the limit as $dt \rightarrow 0$, has weakly concave $\hat{\Pi}_{\tau}$ for some values of time τ and $\hat{\Pi}_{\tau}$ not weakly concave for some other values of τ, it is possible to have a smallest value of τ for which $\hat{\Pi}_{\tau}$ is weakly concave but no largest value of τ for which $\hat{\Pi}_{\tau}$ fails to be weakly concave. That is, the set of τ for which $\hat{\Pi}_{\tau}$ is weakly concave in the limit as $dt \rightarrow 0$ will be closed, but the set of τ for which $\hat{\Pi}_{\tau}$ fails to be weakly concave in the limit as $dt \rightarrow 0$ will be open.

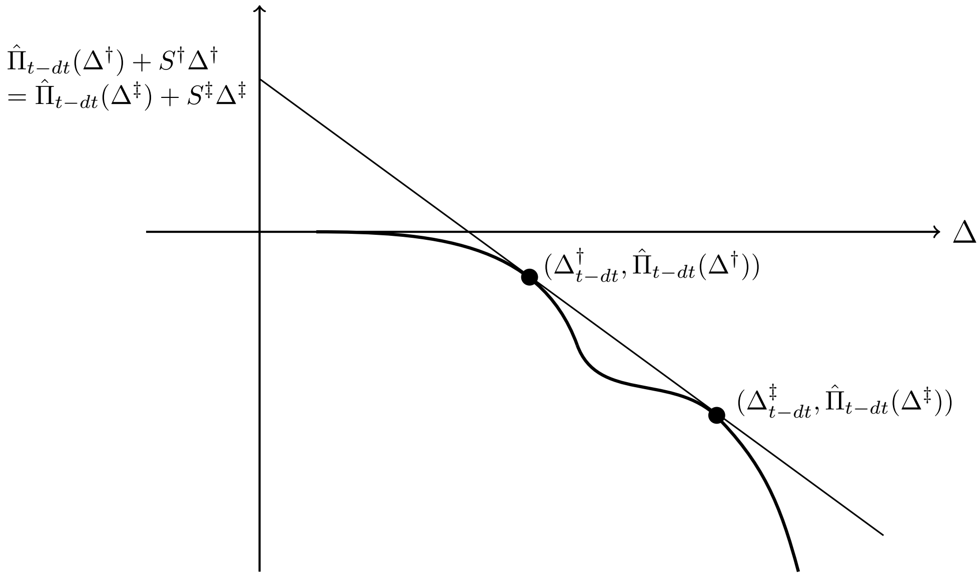


Figure 8: An example of a non-concave (Δ, π) Efficient Frontier

contradiction, we use the fact that these two points have exactly the same slope $-S^\ddagger = -S^\dagger < 0$. We also leverage the fact that $\hat{\Pi}_{t-dt}(\Delta_{t-dt})$ is concave to the left of $\Delta^\dagger$, by focusing on the lowest value of $\Delta^\dagger$ that violates weak concavity. In addition, $\Delta > 0$ is needed to get effort, meaning there will be positive Δ_t at $\pi_t = 0$ on the (Δ, π) Efficient Frontier at time t. Therefore, there will be a strictly positive Δ_{t-dt}, before the (Δ, π) Efficient Frontier at time $t - dt$ begins to drop below $\hat{\Pi}_{t-dt}(\Delta_{t-dt}) = 0$. This guarantees that the intercept of the common tangent line is non-negative:

$$\hat{\Pi}_{t-dt}(\Delta^\dagger) + S^\dagger\Delta^\dagger = \hat{\Pi}_{t-dt}(\Delta^\ddagger) + S^\ddagger\Delta^\ddagger > 0.$$

Now, choose a very small positive value of dt—small enough to satisfy all of the conditions that will become apparent below—as well as the divisibility condition that there exists an integer N such that $n\,dt = T - t$. Then let (Δ^I, π^I) be a point on the (Δ, π) Efficient Frontier at time t from which $(\Delta^\dagger, \hat{\Pi}_{t-dt}(\Delta^\dagger))$ is reachable with $w_t = \hat{w}$ and an optimized value of x. Likewise, let (Δ^{II}, π^{II}) be a point on the (Δ, π) Efficient Frontier at time t from which

$(\Delta^{\ddagger}, \hat{\Pi}_{t-dt}(\Delta^{\ddagger}))$ is reachable with $w_t = \hat{w}$ and an optimized value of x. Note that with small enough dt this implies $\Delta^{II} > \Delta^{I}$. As notation, for given $\hat{w}_t$, let x^I be an optimal value of x that makes $(\Delta^{\dagger}_{t-dt}, \hat{\Pi}_{t-dt}(\Delta^{\dagger}_{t-dt}))$ reachable from (Δ^I, π^I) and let x^{II} be an optimal value of x that makes $(\Delta^{\ddagger}_{t-dt}, \hat{\Pi}_{t-dt}(\Delta^{\ddagger}_{t-dt}))$ reachable from (Δ^{II}, π^{II}).

By the recursion hypothesis, $\Omega_{t;dt,T}$ has a concave Efficient Frontier. That ensures that for $\theta \in (0,1)$,

$$((1-\theta)\Delta^I + \theta\Delta^{II}, (1-\theta)\pi^I + \theta\pi^{II}) \in \Omega_{t;dt,T}.$$

Thus, it is useful to define:

$$\mathbb{F}(\theta) = \mathcal{F}((1-\theta)\Delta^I + \theta\Delta^{II}, (1-\theta)x^I + \theta x^{II}, \hat{w}; dt, T)$$
$$\mathbb{G}(\theta) = \mathcal{G}((1-\theta)\Delta^I + \theta\Delta^{II}, (1-\theta)\pi^I + \theta\pi^{II}, (1-\theta)x^I + \theta x^{II}, \hat{w}; dt, T).$$

Lemma 3. *Suppose* $\exists(\Delta^i_{t-dt}, \hat{\Pi}_{t-dt}(\Delta^i_{t-dt}))$ *pairs where* $i = \{I, II\}$*, such that* $\hat{\Pi}_{t-dt}((1-\theta)\Delta^I_{t-dt} + \theta\Delta^{II}_{t-dt}) < (1-\theta)\hat{\Pi}_{t-dt}(\Delta^I_{t-dt}) + \theta\hat{\Pi}_{t-dt}(\Delta^{II}_{t-dt}),\ \forall\theta \in (0,1)$*. Then, there exists some* $(\Delta^{\dagger}_{t-dt}, \hat{\Pi}_{t-dt}(\Delta^{\dagger}_{t-dt}))$ *and* $(\Delta^{\ddagger}_{t-dt}, \hat{\Pi}_{t-dt}(\Delta^{\ddagger}_{t-dt}))$ *such that* $\hat{\Pi}_{t-dt}((1-\theta)\Delta^{\dagger}_{t-dt} + \theta\Delta^{\ddagger}_{t-dt}) < (1-\theta)\hat{\Pi}_{t-dt}(\Delta^{\dagger}_{t-dt}) + \theta\hat{\Pi}_{t-dt}(\Delta^{\ddagger}_{t-dt}),\ \forall\theta \in (0,1)$ *and* $S^{\ddagger} = -\hat{\Pi}'_{t-dt}(\Delta^{\dagger}_{t-dt}) = -\hat{\Pi}'_{t-dt}(\Delta^{\ddagger}_{t-dt}) = S^{\dagger}$*. Then, if* $(\mathbb{F}(\theta), \mathbb{G}(\theta))$ *parametrically traces out a downward-sloping and concave curve for* $\theta \in [0,1]$ *in* $(\Delta_{t-dt}, \pi_{t-dt})$ *space then* $\hat{\Pi}_{t-dt}((1-\theta)\Delta^{\dagger}_{t-dt} + \theta\Delta^{\ddagger}_{t-dt}) \geq (1-\theta)\hat{\Pi}_{t-dt}(\Delta^{\dagger}_{t-dt}) + \theta\hat{\Pi}_{t-dt}(\Delta^{\ddagger}_{t-dt}),\ \forall\theta \in (0,1)$*. This is a contradiction.*

Proof. For $\theta \in (0,1)$, choosing the optimal x in the maximization problem defining $\hat{\Pi}_{t-dt}(\Delta_{t-dt})$ has to yield at least as high a value of π_{t-dt} as choosing $(1-\theta)x^I + \theta x^{II}$, while at $\theta = 0$ and at $\theta = 1$, this is an exactly optimal value of θ. Therefore, if $(\mathbb{F}(\theta), \mathbb{G}(\theta))$ is above the secant line between $(\Delta^{\dagger}_{t-dt}, \hat{\Pi}_{t-dt}(\Delta^{\dagger}_{t-dt}))$ and $(\Delta^{\ddagger}_{t-dt}, \hat{\Pi}_{t-dt}(\Delta^{\ddagger}_{t-dt}))$ for all $\theta \in (0,1)$, then $\hat{\Pi}_{t-dt}((1-\theta)\Delta^{\dagger}_{t-dt} + \theta\Delta^{\ddagger}_{t-dt}) \geq \mathbb{G}(\theta) \geq (1-\theta)\hat{\Pi}_{t-dt}(\Delta^{\dagger}_{t-dt}) + \theta\hat{\Pi}_{t-dt}(\Delta^{\ddagger}_{t-dt})$ for all $\theta \in (0,1)$.
□

Main Proof in the Endogenous Monitoring Intensity Case

What remains to be shown is that $(\mathbb{F}(\theta), \mathbb{G}(\theta))$ parametrically traces out a curve above the secant line between $(\Delta^{\dagger}_{t-dt}, \hat{\Pi}_{t-dt}(\Delta^{\dagger}_{t-dt}))$ and $(\Delta^{\ddagger}_{t-dt}, \hat{\Pi}_{t-dt}(\Delta^{\ddagger}_{t-dt}))$. Note that as part of the recursion hypothesis, we can take the strict downward slope (more negative than $-\frac{1}{u'(w)}$) of the (Δ_t, π_t) Efficient Frontier as given. The slope of the curve parametrically traced out by $(\mathbb{F}(\theta), \mathbb{G}(\theta))$ is:

$$\frac{\mathbb{G}'(\theta)}{\mathbb{F}'(\theta)}.$$

We need to show that this ratio is negative and that it becomes more negative as θ increases. To compute and analyze this, expand out $\mathbb{F}(\theta)$, then take its derivatives:

$$\begin{aligned}
\mathbb{F}(\theta) &= ((1-\theta)\Delta^I + \theta\Delta^{II})[1-(\rho+q)dt] + [u(\hat{w}) - \psi]dt \\
&\quad - dt\{((1-\theta)x^I + \theta x^{II})[1 - Z\big((1-\theta)x^I + \theta x^{II}\big)] - v(Z\big((1-\theta)x^I + \theta x^{II}\big)\} \\
\mathbb{F}'(\theta) &= (\Delta^{II} - \Delta^I)[1-(\rho+q)dt] - (x^{II} - x^I)[1 - Z\big((1-\theta)x^I + \theta x^{II}\big)]dt \\
\mathbb{F}''(\theta) &= (x^{II} - x^I)^2 Z'\big((1-\theta)x^I + \theta x^{II}\big)dt.
\end{aligned}$$

As $dt \to 0$, the first derivative is positive in the limit, as none of the factors multiplying dt are unbounded. The second derivative is also positive and converges to zero as $dt \to 0$.

Now, expand out $\mathbb{G}(\theta)$ using the definition

$$\alpha = \pi^I - \Big(\frac{\pi^{II} - \pi^I}{\Delta^{II} - \Delta^I}\Big)\Delta^I.$$

Geometrically, this is the intercept on the π axis of the secant line between (Δ^I, π^I) and (Δ^{II}, π^{II}), which is positive because the (Δ, π) Efficient Frontier at time t starts on the left with non-negative Δ and zero π, and is concave by the recursion hypothesis. This definition allows us to write:

$$
\begin{aligned}
\pi^I &= \alpha + \Big(\frac{\pi^{II} - \pi^I}{\Delta^{II} - \Delta^I}\Big)\Delta^I \\
\pi^{II} &= \alpha + \Big(\frac{\pi^{II} - \pi^I}{\Delta^{II} - \Delta^I}\Big)\Delta^{II} \\
(1-\theta)\pi^I + \theta\pi^{II} &= \alpha + \Big(\frac{\pi^{II} - \pi^I}{\Delta^{II} - \Delta^I}\Big)((1-\theta)\Delta^I + \theta\Delta^{II}) \\
\frac{(1-\theta)\pi^I + \theta\pi^{II}}{(1-\theta)\Delta^I + \theta\Delta^{II}} &= \frac{\alpha}{(1-\theta)\Delta^I + \theta\Delta^{II}} + \Big(\frac{\pi^{II} - \pi^I}{\Delta^{II} - \Delta^I}\Big).
\end{aligned}
$$

Recall our definition

$$\mathbb{H}(x) = x(1 - Z(x)) > 0,$$

and our assumption that

$$\mathbb{H}'(x) = 1 - Z(x) - xZ'(x) > 0.$$

Then,

$$\begin{aligned}
\mathbb{G}(\theta) &= [1-(r+q)dt]((1-\theta)\pi^I + \theta\pi^{II}) + \mu Z((1-\theta)x^I + \theta x^{II})dt \\
&\quad - [\hat{w} + c\Big(\frac{(1-\theta)x^I + \theta x^{II}}{(1-\theta)\Delta^I + \theta\Delta^{II}}\Big)]dt - \Big[\frac{\alpha}{(1-\theta)\Delta^I + \theta\Delta^{II}} + \Big(\frac{\pi^{II}-\pi^I}{\Delta^{II}-\Delta^I}\Big)\Big]\mathbb{H}\big((1-\theta)x^I + \theta x^{II}\big)dt \\
\mathbb{G}'(\theta) &= (\pi^{II}-\pi^I)[1-(r+q)dt] + (x^{II}-x^I)\mu Z'((1-\theta)x^I + \theta x^{II})dt \\
&\quad - \Big(\frac{x^{II}}{\Delta^{II}} - \frac{x^I}{\Delta^I}\Big)\frac{\Delta^I\Delta^{II}}{[(1-\theta)\Delta^I + \theta\Delta^{II}]^2}c'\Big(\frac{(1-\theta)x^I + \theta x^{II}}{(1-\theta)\Delta^I + \theta\Delta^{II}}\Big)dt \\
&\quad - (x^{II}-x^I)\mathbb{H}'\big((1-\theta)x^I + \theta x^{II}\big)\Big[\frac{\alpha}{(1-\theta)\Delta^I + \theta\Delta^{II}} + \Big(\frac{\pi^{II}-\pi^I}{\Delta^{II}-\Delta^I}\Big)\Big]dt \\
&\quad + \frac{\alpha(\Delta^{II}-\Delta^I)}{[(1-\theta)\Delta^I + \theta\Delta^{II}]^2}\mathbb{H}\big((1-\theta)x^I + \theta x^{II}\big)dt \\
\mathbb{G}''(\theta) &= (x^{II}-x^I)^2\mu Z''((1-\theta)x^I + \theta x^{II})dt \\
&\quad + 2\Big(\frac{x^{II}}{\Delta^{II}} - \frac{x^I}{\Delta^I}\Big)\frac{\Delta^I\Delta^{II}(\Delta^{II}-\Delta^I)}{[(1-\theta)\Delta^I + \theta\Delta^{II}]^3}c'\Big(\frac{(1-\theta)x^I + \theta x^{II}}{(1-\theta)\Delta^I + \theta\Delta^{II}}\Big)dt \\
&\quad - \Big(\frac{x^{II}}{\Delta^{II}} - \frac{x^I}{\Delta^I}\Big)^2\Big(\frac{\Delta^I\Delta^{II}}{[(1-\theta)\Delta^I + \theta\Delta^{II}]^2}\Big)^2 c''\Big(\frac{(1-\theta)x^I + \theta x^{II}}{(1-\theta)\Delta^I + \theta\Delta^{II}}\Big)dt \\
&\quad - (x^{II}-x^I)^2\mathbb{H}''\big((1-\theta)x^I + \theta x^{II}\big)\Big[\frac{\alpha}{(1-\theta)\Delta^I + \theta\Delta^{II}} + \Big(\frac{\pi^{II}-\pi^I}{\Delta^{II}-\Delta^I}\Big)\Big]dt \\
&\quad + 2(x^{II}-x^I)\frac{\alpha(\Delta^{II}-\Delta^I)}{[(1-\theta)\Delta^I + \theta\Delta^{II}]^2}\mathbb{H}'\big((1-\theta)x^I + \theta x^{II}\big)dt \\
&\quad - \frac{2\alpha(\Delta^{II}-\Delta^I)^2}{[(1-\theta)\Delta^I + \theta\Delta^{II}]^3}\mathbb{H}\big((1-\theta)x^I + \theta x^{II}\big)dt.
\end{aligned}$$

In the limit as $dt \to 0$, the first derivative is clearly negative.

For $\frac{\mathbb{G}'(\theta)}{\mathbb{F}'(\theta)}$ to become more negative as θ increases, what we need is:

$$\frac{d}{d\theta}\ln\Big|\frac{\mathbb{G}'(\theta)}{\mathbb{F}'(\theta)}\Big| = \frac{\mathbb{G}''(\theta)}{\mathbb{G}'(\theta)} - \frac{\mathbb{F}''(\theta)}{\mathbb{F}'(\theta)} > 0,$$

or equivalently,

$$\mathbb{G}''(\theta) - \frac{\mathbb{G}'(\theta)}{\mathbb{F}'(\theta)}\mathbb{F}''(\theta) < 0.$$

Note that as $dt \to 0$,

$$\frac{\mathbb{F}'(\theta)}{\mathbb{G}'(\theta)} = \frac{\pi^{II} - \pi^{I}}{\Delta^{II} - \Delta^{I}} + O(dt).$$

Hence, since the second derivatives are $O(dt)$, as $dt \to 0$,

$$\begin{aligned}
\mathbb{G}''(\theta) - \frac{\mathbb{G}'(\theta)}{\mathbb{F}'(\theta)}\mathbb{F}''(\theta) &= \mathbb{G}''(\theta) - \Big[\Big(\frac{\pi^{II} - \pi^{I}}{\Delta^{II} - \Delta^{I}}\Big) + O(dt)\Big]\mathbb{F}''(\theta) \\
&= \mathbb{G}''(\theta) - \Big(\frac{\pi^{II} - \pi^{I}}{\Delta^{II} - \Delta^{I}}\Big)\mathbb{F}''(\theta) + O(dt^2).
\end{aligned}$$

Computing, using $\mathbb{H}'(x) = 1 - Z(x) - xZ'(x)$ and $\mathbb{H}''(x) = -2Z'(x) - xZ''(x)$ and

$$\Delta^{I}\Delta^{II}\Big(\frac{x^{II}}{\Delta^{II}} - \frac{x^{I}}{\Delta^{I}}\Big) = (x^{II} - x^{I})((1-\theta)\Delta^{I} + \theta\Delta^{II}) - (\Delta^{II} - \Delta^{I})((1-\theta)x^{I} + \theta x^{II}),$$

yields:

$$\begin{aligned}
dt^{-1}\Big[\mathbb{G}''(\theta) - \Big(\frac{\pi^{II} - \pi^{I}}{\Delta^{II} - \Delta^{I}}\Big)\mathbb{F}''(\theta)\Big] &= (x^{II} - x^{I})^2\mu Z''((1-\theta)x^{I} + \theta x^{II}) \\
&+ 2\Big(\frac{x^{II}}{\Delta^{II}} - \frac{x^{I}}{\Delta^{I}}\Big)\frac{\Delta^{I}\Delta^{II}(\Delta^{II} - \Delta^{I})}{[(1-\theta)\Delta^{I} + \theta\Delta^{II}]^3}c'\Big(\frac{(1-\theta)x^{I} + \theta x^{II}}{(1-\theta)\Delta^{I} + \theta\Delta^{II}}\Big) \\
&- \Big(\frac{x^{II}}{\Delta^{II}} - \frac{x^{I}}{\Delta^{I}}\Big)^2\Big(\frac{\Delta^{I}\Delta^{II}}{[(1-\theta)\Delta^{I} + \theta\Delta^{II}]^2}\Big)^2 c''\Big(\frac{(1-\theta)x^{I} + \theta x^{II}}{(1-\theta)\Delta^{I} + \theta\Delta^{II}}\Big) \\
&+ \Big[2Z'((1-\theta)x^{I} + \theta x^{II}) + ((1-\theta)x^{I} + \theta x^{II})Z''((1-\theta)x^{I} + \theta x^{II})\Big] \\
&\times (x^{II} - x^{I})^2\Big[\frac{\alpha}{(1-\theta)\Delta^{I} + \theta\Delta^{II}} + \Big(\frac{\pi^{II} - \pi^{I}}{\Delta^{II} - \Delta^{I}}\Big)\Big] \\
&+ 2(x^{II} - x^{I})\frac{\alpha(\Delta^{II} - \Delta^{I})}{[(1-\theta)\Delta^{I} + \theta\Delta^{II}]^2} \\
&\times \Big[1 - Z((1-\theta)x^{I} + \theta x^{II}) - ((1-\theta)x^{I} + \theta x^{II})Z'((1-\theta)x^{I} + \theta x^{II})\Big] \\
&- \frac{2\alpha(\Delta^{II} - \Delta^{I})^2}{[(1-\theta)\Delta^{I} + \theta\Delta^{II}]^3}((1-\theta)x^{I} + \theta x^{II})[1 - Z((1-\theta)x^{I} + \theta x^{II})] \\
&- \Big(\frac{\pi^{II} - \pi^{I}}{\Delta^{II} - \Delta^{I}}\Big)(x^{II} - x^{I})^2 Z'((1-\theta)x^{I} + \theta x^{II}).
\end{aligned}$$

These terms can be grouped in a way that allows us to show that each group is negative:

$$
\begin{aligned}
dt^{-1}\Big[\mathbb{G}''(\theta) - \Big(\frac{\pi^{II}-\pi^{I}}{\Delta^{II}-\Delta^{I}}\Big)\mathbb{F}''(\theta)\Big] &= \Big(\frac{\pi^{II}-\pi^{I}}{\Delta^{II}-\Delta^{I}}\Big)(x^{II}-x^{I})^2 Z'\big((1-\theta)x^{I}+\theta x^{II}\big) \\
&- \Big(\frac{x^{II}}{\Delta^{II}} - \frac{x^{I}}{\Delta^{I}}\Big)^2 \Big(\frac{\Delta^{I}\Delta^{II}}{[(1-\theta)\Delta^{I}+\theta\Delta^{II}]^2}\Big)^2 c''\Big(\frac{(1-\theta)x^{I}+\theta x^{II}}{(1-\theta)\Delta^{I}+\theta\Delta^{II}}\Big) \\
&+ (x^{II}-x^{I})^2 Z''\big((1-\theta)x^{I}+\theta x^{II}\big)\Big[\mu + \frac{((1-\theta)x^{I}+\theta x^{II})((1-\theta)\pi^{I}+\theta\pi^{II})}{(1-\theta)\Delta^{I}+\theta\Delta^{II}}\Big] \\
&+ \frac{2\Delta^{I}\Delta^{II}}{[(1-\theta)\Delta^{I}+\theta\Delta^{II}]^3}\Big(\frac{x^{II}}{\Delta^{II}} - \frac{x^{I}}{\Delta^{I}}\Big) \\
&\times \Big\{(\Delta^{II}-\Delta^{I})c'\Big(\frac{(1-\theta)x^{I}+\theta x^{II}}{(1-\theta)\Delta^{I}+\theta\Delta^{II}}\Big) + \alpha(\Delta^{II}-\Delta^{I})[1 - Z\big((1-\theta)x^{I}+\theta x^{II}\big)] \\
&\quad + \alpha(x^{II}-x^{I})[(1-\theta)\Delta^{I}+\theta\Delta^{II}]Z'\big((1-\theta)x^{I}+\theta x^{II}\big)\Big\}.
\end{aligned}
$$

The first-line term on the right-hand side is negative simply because the downward slope of the (Δ, π) Efficient Frontier at time t makes $\frac{\pi^{II}-\pi^{I}}{\Delta^{II}-\Delta^{I}}$ negative. The second-line term is negative because $c''(\cdot) > 0$ and because its other factors are squared. To show that the third-line term is negative, note both that $Z'' < 0$ and that concavity of the (Δ, π) Efficient Frontier at time t and positive Δ at the top of the (Δ, π) Efficient Frontier at time t implies that the secant line from (Δ^{I}, π^{I}) to (Δ^{II}, π^{II}) has a steeper downward slope than ray from the origin to (Δ^{I}, π^{I}) and therefore that the far end of that secant line (Δ^{II}, π^{II}) is on a steeper ray from the origin than $((1-\theta)\Delta^{I} + \Delta^{II}, (1-\theta)\pi^{I} + \pi^{II})$. Thus:

$$\frac{\pi^{II}}{\Delta^{II}} \le \frac{(1-\theta)\pi^{I}+\theta\pi^{II}}{(1-\theta)\Delta^{I}+\theta\Delta^{II}},$$

and since $x^{II} \ge x^{I}$ (to be shown below),

$$\mu + \frac{(1-\theta)x^{I}+\theta x^{II})((1-\theta)\pi^{I}+\theta\pi^{II})}{(1-\theta)\Delta^{I}+\theta\Delta^{II}} \ge \mu + \frac{x^{II}\pi^{II}}{\Delta^{II}} = \mu + m^{II}\pi^{II} > 0.$$

The last inequality follows from the first-order condition for optimal x_t, given $\hat{w}_t$, which rearranged is:

$$\mu + \frac{x^{II}\pi^{II}}{\Delta^{II}} = \Delta_t^{-1}\Big\{(\pi_t + S^{\ddagger}\Delta_t)[1 - Z(x)] + c'(\frac{x}{\Delta_t})\Big\} > 0.$$

$\pi_t + S^{\ddagger}\Delta_t$ is non-negative as $dt \to 0$, trivially when $\pi_t = 0$ and otherwise because then

$$\hat{\Pi}_{t-dt}(\Delta^{\dagger}) + S^{\dagger}\Delta^{\dagger} = \hat{\Pi}_{t-dt}(\Delta^{\ddagger}) + S^{\ddagger}\Delta^{\ddagger} > 0,$$

and

$$\pi_t + S^{\ddagger}\Delta_t = [\hat{\Pi}_{t-dt}(\Delta^{\ddagger}) + O(dt)] + S^{\ddagger}[\Delta^{\ddagger} + O(dt)] = \hat{\Pi}_{t-dt}(\Delta^{\ddagger}) + S^{\ddagger}\Delta^{\ddagger} + O(dt).$$

The remainder of the right-hand side is all one group. Because $c', 1 - Z$ and Z' are all strictly positive, and $(\Delta^{II} - \Delta^{I}) > 0$ simply by the orientation we have chosen (without loss of generality), to show that this large group of terms is negative, we only need to show that $(\frac{x^{II}}{\Delta^{II}} - \frac{x^{I}}{\Delta^{I}}) < 0$ and $(x^{II} - x^{I}) > 0$. The lemmas below do that work.

Lemma 4. *If $\mu + \frac{x}{\Delta}\pi > 0$, and $\pi + S\Delta \geq 0$,then for fixed values of Δ, π, μ,and S, the expression $\mu Z'(x) - \frac{\pi}{\Delta}[1 - Z(x) - xZ'(x)] - S[1 - Z(x)] - \frac{1}{\Delta}c'(\frac{x}{\Delta})$, which we call FOCx, is strictly decreasing in x, implying that, within the region in which $\mu + \frac{x}{\Delta}\pi > 0$, there is at most one x that satisfies the first-order condition for optimal x given Δ, π, μ, and S and that changes in the combination of Δ, π, μ,and S that raise this expression raise optimal x,if any such new optimum exists, while changes in the combination of Δ, π, μ and S that lower this expression lower optimal x, if any such new optimum exists.*

Proof. Compute the derivative of this expression FOCx with respect to x:

$$(\mu + \frac{x}{\Delta}\pi)Z''(x) - \Big(\frac{\pi + S\Delta}{\Delta}\Big)Z'(x) - \frac{1}{\Delta^2}c''(\frac{x}{\Delta}) < 0.$$

Remark: Note that only the region where $\mu + \frac{x}{\Delta}\pi > 0$ and $\pi + S\Delta \geq 0$ is relevant to the economic problem of the firm, for reasons discussed above in connection with the sign of the third-line term.

Lemma 5. *If $\mu + \frac{x}{\Delta}\pi > 0$, $\pi + S\Delta \geq 0$, and $S^{\ddagger} = S^{\dagger}$, then $x^{II} > x^{I}$.*

Proof. Consider how going from I to II affects FOCx. For reasons discussed above in connection with the sign of the third-line term, $\frac{\pi^{II}}{\Delta^{II}} < \frac{\pi^{I}}{\Delta^{I}}$ and $-\frac{\pi^{II}}{\Delta^{II}} > -\frac{\pi^{I}}{\Delta^{I}}$. That pulls FOCx up, since $1 - Z(x) - xZ'(x) > 0$. μ and S are unchanged in going from I to II.

Remark. The proof strategy of proof by contradiction was necessary in order to have Lemma 5 do the work it needs to do, since Lemma 5 depends on $S^{\ddagger} = S^{\dagger}$ (or actually, on $S^{\ddagger} \leq S^{\dagger}$).

Lemma 6. *If $\mu + \frac{x}{\Delta}\pi > 0$, $\pi + S\Delta \geq 0$, and $S^{\ddagger} = S^{\dagger}$, then $\frac{x^{II}}{\Delta^{II}} < \frac{x^{I}}{\Delta^{I}}$.*

Proof. Consider how going from I to II affects FOCx when $\frac{x}{\Delta}$ is held constant. To facilitate that, rewrite FOCx with $m = \frac{x}{\Delta}$. The objective here is to prove that monitoring intensity m satisfies $m^{II} < m^{I}$. Note that the first-order condition for optimal x given Δ, π, μ, and S is one hundred percent equivalent to the first-order condition for optimal m given Δ, π, μ, and S. And having at most one optimal x given Δ, π, μ, and S implies having at most one optimal m given Δ, π, μ, and S. Writing that first-order condition in terms of m rather than x, rearranging and dividing through by $\frac{1-Z(m\Delta)}{\Delta}$ yields:

$$[\frac{\mu}{m} + \pi]\frac{m\Delta Z'(m\Delta)}{(1 - Z(m\Delta))} - (\pi + S\Delta) - \frac{c'(m)}{[1 - Z(m\Delta)]} = 0.$$

In going from I to II, the common tangent means that $\pi + S\Delta$, which is the intercept of that tangent, is unchanged. At a given value of m a higher value of Δ makes $-\frac{c'(m)}{[1-Z(m\Delta)]}$ a bigger negative. As for the first term, the factor $\frac{\mu}{m} + \pi$ becomes a smaller positive number for given m when π goes down in going from I to II, while for given m the factor $\frac{m\Delta Z'(m\Delta)}{1-Z(m\Delta)}$ goes down with the higher Δ, by our assumption that $\frac{xZ'(x)}{1-Z(x)}$ is decreasing in $x = m\Delta$. Thus, the optimal m must be higher at II than at I:

$$m^{II} = \frac{x^{II}}{\Delta^{II}} < \frac{x^I}{\Delta^I} = m^I. \qquad \square$$

B.1.2 The case with exogenous monitoring intensity

The case with exogenous monitoring intensity can be easily proved using the argument for the endogenous case. Furthermore, for the exogenous case, we only need the basic assumptions on the disutility function v that we make in Section 2, and no longer need to make any extra assumptions on the on the shape of function $Z(\cdot)$ that we needed for the endogenous monitoring case, in particular assumptions (B.2) and (B.3). We show how the exogenous case can be proved below.

As before denote the (Δ_t, π_t) Feasible Set at time t, given period length dt, and terminal time T, by: $\Omega_{t;dt,T}$. Define

$$Z(\Delta_t) = v'^{-1}(m\Delta_t)$$

so that

$$z_t = Z(\Delta_t).$$

Then

$$Z'(\Delta_t) = \frac{1}{v''(v'^{-1}(\Delta_t))} = \frac{1}{v''(Z(\Delta_t))} > 0;$$

$$Z''(\Delta_t) = \frac{d}{d\Delta_t}\Big(\frac{1}{v''(v'^{-1}(\Delta_t))}\Big) = \frac{-v'''(v'^{-1}(\Delta_t))}{[v''(v'^{-1}(\Delta_t))]^3} < 0.$$

Next, define

$$\begin{aligned}
\mathcal{F}(\Delta_t, w_t; dt, T) &= \{1 + [mZ(\Delta_t) - m - \rho - q]dt\}\Delta_t + [u(w_t) - v(Z(\Delta_t)) - \psi]dt \\
&= [1 - (\rho + q)dt]\Delta_t - m\Delta_t[1 - Z(\Delta_t)]dt + [u(w_t) - v(Z(\Delta_t)) - \psi]dt \\
\mathcal{G}(\Delta_t, \pi_t, w_t; dt, T) &= \{1 + [mZ(\Delta_t) - m - r - q]dt\}\pi_t + [\mu Z(\Delta_t) - w_t]dt \\
&= [1 - (r + q)dt]\pi_t - m\pi_t[1 - Z(\Delta_t)]dt + [\mu Z(\Delta_t) - w_t]dt
\end{aligned}$$

so that

$$\Delta_{t-dt} = \mathcal{F}(\Delta_t, w_t; dt, T)$$

and

$$\pi_{t-dt} = \mathcal{G}(\Delta_t, \pi_t, w_t; dt, T).$$

Define

$$\begin{aligned}
\hat{\Pi}_{t-dt}(\Delta_{t-dt}) = \max_{\Delta_t, \pi_t} & \, \mathcal{G}(\Delta_t, \pi_t, \hat{w}_t; dt, T) \\
s.t. \qquad & (\Delta_t, \pi_t) \in \Omega_{t;dt,T} \\
& \mathcal{F}(\Delta_t, \hat{w}_t; dt, T) \geq \Delta_{t-dt}.
\end{aligned} \tag{B.5}$$

Note that Lemma 1 and Lemma 2 remains completely unchanged with exogenous monitoring. As before we proceed by contradiction to prove the concavity of $\hat{\Pi}_{t-dt}(\Delta_{t-dt})$. Equation (B.4) also remains unchanged. Thus, as before, if $\hat{\Pi}_{t-dt}(\Delta_{t-dt})$ fails to be weakly concave globally, then there must be at least one pair of points on $(\Delta^\dagger, \hat{\Pi}_{t-dt}(\Delta^\dagger))$ and $(\Delta^\ddagger, \hat{\Pi}_{t-dt}(\Delta^\ddagger))$, with $\Delta^\ddagger > \Delta^\dagger$ and both at a locally concave part of the curve, which share a tangent line, with points at all the values of $\Delta_{t-dt} \in (\Delta^\dagger, \Delta^\ddagger)$ strictly below that shared tangent line

Focusing on the pair of this type that has the lowest value of $\Delta^\dagger$, let (Δ^I, π^I) be a point on the (Δ, π) Efficient Frontier at time t from which $(\Delta^\dagger, \hat{\Pi}_{t-dt}(\Delta^\dagger))$ is reachable with $w_t = \hat{w}$. Likewise, let (Δ^{II}, π^{II}) be a point on the (Δ, π) Efficient Frontier at time t from which $(\Delta^\ddagger, \hat{\Pi}_{t-dt}(\Delta^\ddagger))$ is reachable with $w_t = \hat{w}$.

Define:

$$\mathbb{F}(\theta) = \mathcal{F}((1-\theta)\Delta^I + \theta\Delta^{II}, \hat{w}; dt, T)$$
$$\mathbb{G}(\theta) = \mathcal{G}((1-\theta)\Delta^I + \theta\Delta^{II}, (1-\theta)\pi^I + \theta\pi^{II}, \hat{w}; dt, T).$$

The statement of Lemma 3 remains unchanged. The proof also directly follows from the definition of of $\hat{\Pi}_{t-dt}$ in the maximization problem (B.5).

All that remains is to show that $(\mathbb{F}(\theta), \mathbb{G}(\theta))$ parametrically traces out a curve above the secant line between $(\Delta^\dagger_{t-dt}, \hat{\Pi}_{t-dt}(\Delta^\dagger_{t-dt}))$ and $(\Delta^\ddagger_{t-dt}, \hat{\Pi}_{t-dt}(\Delta^\ddagger_{t-dt}))$, which can be done much more easily than in the endogenous case. As the slope of the curve parametrically traced out by $(\mathbb{F}(\theta), \mathbb{G}(\theta))$ is:

$$\frac{\mathbb{G}'(\theta)}{\mathbb{F}'(\theta)}.$$

we show that this ratio is negative and that it becomes more negative as θ increases. Taking derivatives:

$$\begin{aligned}
\mathbb{F}(\theta) &= ((1-\theta)\Delta^I + \theta\Delta^{II})[1-(\rho+q)dt] + [u(\hat{w}) - \psi]dt \\
&\quad - mdt\{((1-\theta)\Delta^I + \theta\Delta^{II})[1 - Z\big((1-\theta)\Delta^I + \theta\Delta^{II}\big)] - v(Z\big((1-\theta)\Delta^I + \theta\Delta^{II}\big)\} \\
\mathbb{F}'(\theta) &= (\Delta^{II} - \Delta^I)[1-(\rho+q)dt] - m(\Delta^{II} - \Delta^I)[1 - Z\big((1-\theta)x^I + \theta\Delta^{II}\big)]dt \\
\mathbb{F}''(\theta) &= m(\Delta^{II} - \Delta^I)^2 Z'\big((1-\theta)\Delta^I + \theta x^{II}\big)dt,
\end{aligned}$$

As $dt \to 0$, the first derivative is positive in the limit, as none of the factors multiplying dt are unbounded. The second derivative is also positive and converges to zero as $dt \to 0$.

Also,

$$
\begin{aligned}
\mathbb{G}(\theta) &= [1-(r+q)dt]((1-\theta)\pi^I + \theta\pi^{II}) + \mu Z((1-\theta)\Delta^I + \theta\Delta^{II})dt \\
&\quad - \hat{w}dt - m((1-\theta)\pi^I + \theta\pi^{II})\big[1 - Z\big((1-\theta)\Delta^I + \theta\Delta^{II}\big)\big]dt \\
\mathbb{G}'(\theta) &= (\pi^{II} - \pi^I)[1-(r+q)dt] + (\Delta^{II} - \Delta^I)\mu Z'((1-\theta)\Delta^I + \theta\Delta^{II})dt \\
&\quad - m(\pi^{II} - \pi^I)\big[1 - Z\big((1-\theta)\Delta^I + \theta\Delta^{II}\big)\big]dt \\
&\quad + m(\Delta^{II} - \Delta^I)((1-\theta)\pi^I + \theta\pi^{II})Z'\big((1-\theta)\Delta^I + \theta\Delta^{II}\big)dt \\
\mathbb{G}''(\theta) &= (\Delta^{II} - \Delta^I)^2\mu Z''((1-\theta)\Delta^I + \theta\Delta^{II})dt \\
&\quad + m(\Delta^{II} - \Delta^I)(\pi^{II} - \pi^I)Z''((1-\theta)\Delta^I + \theta\Delta^{II})dt \\
&\quad + m(\Delta^{II} - \Delta^I)(\pi^{II} - \pi^I)Z'((1-\theta)\Delta^I + \theta\Delta^{II})dt \\
&\quad + m(\Delta^{II} - \Delta^I)^2((1-\theta)\pi^I + \theta\pi^{II})Z''\big((1-\theta)\Delta^I + \theta\Delta^{II}\big)dt
\end{aligned}
$$

In the limit as $dt \to 0$, the first derivative is clearly negative.

As before we need to show:

$$\mathbb{G}''(\theta) - \frac{\mathbb{G}'(\theta)}{\mathbb{F}'(\theta)}\mathbb{F}''(\theta) < 0.$$

Since

$$\frac{\mathbb{F}'(\theta)}{\mathbb{G}'(\theta)} = \frac{\pi^{II} - \pi^I}{\Delta^{II} - \Delta^I} + O(dt).$$

as $dt \to 0$,

$$
\begin{aligned}
\mathbb{G}''(\theta) - \frac{\mathbb{G}'(\theta)}{\mathbb{F}'(\theta)}\mathbb{F}''(\theta) &= \mathbb{G}''(\theta) - \Big[\Big(\frac{\pi^{II} - \pi^I}{\Delta^{II} - \Delta^I}\Big) + O(dt)\Big]\mathbb{F}''(\theta) \\
&= \mathbb{G}''(\theta) - \Big(\frac{\pi^{II} - \pi^I}{\Delta^{II} - \Delta^I}\Big)\mathbb{F}''(\theta) + O(dt^2)
\end{aligned}
$$

Then,

$$dt^{-1}\Big[\mathbb{G}''(\theta)-\Big(\frac{\pi^{II}-\pi^{I}}{\Delta^{II}-\Delta^{I}}\Big)\mathbb{F}''(\theta)\Big]=(\Delta^{II}-\Delta^{I})^2\mu Z''((1-\theta)\Delta^{I}+\theta\Delta^{II})dt$$
$$+m(\Delta^{II}-\Delta^{I})(\pi^{II}-\pi^{I})Z''((1-\theta)\Delta^{I}+\theta\Delta^{II})dt$$
$$+m(\Delta^{II}-\Delta^{I})^2((1-\theta)\pi^{I}+\theta\pi^{II})Z''\big((1-\theta)\Delta^{I}+\theta\Delta^{II}\big)dt$$

This is negative as $Z''<0$. $\square$

B.2 Proof of Proposition 6

Proposition 6. *Assume that the evolution of Δ, π and $\pi+s\Delta$ with worker tenure are monotonic in the direction established by Proposition 2. Then, in steady state, on-the-job training T_t is provided such that, once human capital h_t begins to increase, it continually increases with seniority.*

(Proof.) First, consider the case when human capital h is fixed. From the integral equation (31), we can see that higher Δ with greater seniority leads to higher effort z, which in turn raises the factor z in the integrand in the equation for i_t and reduces the attrition rate in the discounting. All of these forces make the incentive to invest in training—and therefore make training itself rise with seniority s, if the future path of subsequent human capital were (counterfactually) held fixed.

Now, to see that, in steady state, once human capital begins to increase it must continue to increase thereafter, consider two workers who were identical at hiring, but one was hired shortly before the other. When the less senior worker is at the level of seniority where human capital first begins to weakly increase (at hiring if human capital starts at zero), the slightly more senior worker must have a higher level of human capital. So the less senior worker could only have a higher level of human capital thereafter by overtaking the more senior worker in human capital. But the less senior worker can't overtake the more senior worker (which would mean human capital was declining at that point) and stay forever ahead, because then higher human capital thereafter combined with all the forces discussed above guarantees that investment in training is *lower* at the moment the level of human capital is the same, contradicting the claim of overtaking by the less senior worker. And the less senior worker can't overtake the more senior worker at some point and then fall behind later, because viewing the differential equation version of the equation for i from a retrograde perspective,

$$-\dot{i} = [\mu_{t'} + m(\pi_{t'} + s_{t'}\Delta_{t'})]z_{t'}\varphi'(h_{t'}) - (r + \delta + a(m, z))i,$$

if there is a moment where i is lower for the more junior worker (as is needed for the more junior worker's human capital to be overtaken), and prior to that, there is a period of time over which the more junior worker always has higher human capital, then i must be lower for the more junior worker throughout that period and at its boundary, contradicting the claim that the more junior worker pulled ahead at some point after having been behind. Therefore, once the more junior worker is behind, the more junior worker will stay behind in human capital. That in turn implies that the optimal contract has human capital increasing to an asymptote, once it begins to first increase. And if human capital at hiring is zero (say, because it is entirely firm-specific), then despite human capital being subject to depreciation, the optimal path of training guarantees that human capital is (at least weakly) increasing with seniority from hiring on. □