Discussion Paper No. 488

A THEORY OF WAGE DYNAMICS

by

Milton Harris and Bengt Holmstrom *

April 1981
Revised December 1981

J.L. Kellogg Graduate School of Management
Northwestern University
Evanston, Illinois 60201

* We would like to thank Bruce Greenwald, Edward Lazear, Edward Prescott, Sherwin Rosen, Steve Ross and participants at the conference on Information and Transaction Cost Economics (University of Pennsylvania, May 7-8, 1981), the Labor Workshop, Graduate School of Business, University of Chicago, and the IMSSS Seminar, Stanford University (June 30, 1981) for helpful suggestions. We also thank David Kreps for significant technical advice. Partial support from NSF grants SES-8105876 and SES-7826262 is gratefully acknowledged.

ABSTRACT

A dynamic, equilibrium model of long term (implicit) labor contracts under incomplete but symmetric information is developed. Workers are assumed to be risk averse and of unknown ability or productivity. Risk neutral firms learn, as do workers, about each worker's productivity by observing the worker's output over time. It is shown that equilibrium contracts provide for wages which never decline with age and increase only when the worker's market value increases above his current wage. In addition to characterizing the equilibrium wage contract, we also derive some of its implications for the behavior of aggregate wages across various groups of workers. These implications explain some findings in the recent empirical literature on age-earnings profiles. In particular our model can explain why earnings may be positively related to experience even after controlling for productivity, as some empirical studies have indicated.

1. Introduction

This paper presents and analyzes a model of wage dynamics based on a process in which both firm and worker learn about the worker's ability. Our primary purpose is to discover to what extent such a model can account for some of the major stylized facts regarding the relationship among experience, earnings and ability. A secondary purpose is to develop a basis and framework for analyzing dynamic agency models in which market reputation and learning are key ingredients.

A great deal of empirical research on earnings and experience has been undertaken in connection with the human capital model [see Becker (1962) for an early statement of this model]. One of the most comprehensive studies is that of Mincer (1974) which sets out a fairly general human capital model along with extensive empirical analysis [see also Becker (1975)]. The most fundamental and universal fact to emerge from this research is simply that, on average, earnings increase with work experience. Human capital theory explains this increase as a return on investment in productivity enhancing skills accumulated while working.[1]

An interesting variant of traditional human capital theory is contained in the job matching models of Mortensen (1978), Ross, Taubman and Wachter (1980), Jovanovic (1979), Prescott and Visscher (1980) and others. Although primarily meant to explain turnover phenomena, these models can also account for the observed upward sloping experience - earnings profile. This is accomplished by postulating that firms learn about worker productivities in various jobs (the workers' skills actually remain constant) instead of workers

---

[1] A number of other empirical results are also interpreted using this same, productivity based approach. Some of these will be discussed later in Section 5.

developing skills through experience. This learning allows more senior workers to be matched better to tasks than less senior workers. The result is that more senior workers exhibit higher productivity on average, and this accounts for their higher average earnings.

Some recent empirical evidence suggests, however, that there may be factors other than acquisition of productivity enhancing human capital which produce upward sloping experience - earnings profiles. Pascal and Rapping (1972) find that experience and earnings of major league baseball players are positively related <u>even controlling for measured productivity</u>. Similarly, Medoff and Abraham (1980) find that more experienced managerial employees earn more on average even though their performance is not as highly rated as less experienced workers in the same job category.[2] Without denying the obvious importance of productivity increases in explaining upward sloping experience-earnings profiles, it appears to us that development of complementary, non-productivity-based models is of interest. In this paper we present a model which can account, to a considerable extent, for the empirical findings of Mincer, Pascal and Rapping, Medoff and Abraham, and others, but which is not based on the acquisition of productivity enhancing human capital.[3]

Our model of wage dynamics is one in which both workers and firms are imperfectly informed about worker abilities. Both learn gradually about ability by observing the worker's output over time. Since output and ability are not perfectly correlated, this learning process results in random

---

2 Obviously measuring and controlling for productivity is not easily done. Not surprisingly the results of these studies are subject to some reservations which are discussed in Section 5, below.

3 Other models of this general type include Lazear (1979, 1981), Becker and Stigler (1974), and Freeman (1977). These are discussed in the concluding section.

fluctuations of perceived ability over time. Long-term implicit contracts will therefore emerge in order to protect risk averse workers from wage changes induced by fluctuations in perceived productivity.[4] The analysis consists of characterizing the optimal long-term contract and deriving its aggregate implications.[5]

We have chosen our learning process to be normal as in Jovanovic (1979). This leads to a simple characterization of the optimal contract. It also provides a fairly rich set of predictions for the aggregate variables, since we have three worker characteristics -- age, perceived ability and precision of beliefs about ability -- with which to work. Moreover, initial beliefs and their precision can be allowed to depend on observable characteristics, such as education.

Regarding the optimal long-term contract, we show that if firms are risk neutral, such a contract will entail a downward rigid wage, i.e. under this contract one's wage will never fall. The wage is not fully rigid, since we assume that workers can quit to accept higher offers from the market. The threat of quitting will force the wage to be bid up whenever the market wage is higher than the current wage. A bid up wage may be viewed as a renegotiated contract so that contracts can simply be interpreted as a minimum wage guarantee which is equal to the market value of the worker at the time of

---

4 Implicit contract theory was originated by Azariadis (1975), Baily (1974), and Gordon (1974). Our model is most closely related to those of Freeman (1977) and Holmstrom (1980).

5 One might justifiably view our model as being, in one way, a specific version of the traditional human capital model in which the capital being acquired is information about productivity [Mincer (1974) expresses a similar view - see his footnote 7, p. 33]. In another way, our model extends the traditional human capital approach to include long-term contracts and aspects of insurance.

contracting.

Using the normal learning process, we show that the market wage of a worker is his current mean perceived productivity minus a term which depends only on his age and the precision of beliefs about his productivity. This second term may be thought of as an insurance premium for the downward rigidity of wages. We show that this premium decreases both with age and precision.[6]

In the aggregate, these results imply that senior workers earn more on average, because they have had more time to have their wages bid up by the market. The well-established positively sloped experience-earnings profile is here generated purely by the insurance effect. Moreover, as we show, senior workers earn more on average even when holding perceived productivity constant. The economic rationale for this result comes in two parts. First, senior workers have had more chances for their wages to be bid up. Second, they pay lower insurance premia both because their ability can be more precisely assessed and they have fewer periods to remain in the market.

The result that, controlling for productivity, more experienced workers earn more is consistent with Pascal and Rapping (1972) and Medoff and Abraham (1980). We also derive a number of other implications of the model for aggregate relationships between experience and earnings and compare them to empirical

---

6 Freeman (1977) develops a two-period model based on learning about productivity, which yields a downward rigid second period wage. His learning process (a Bernoulli process) is different from ours and does not result in optimal contracts with the specific properties of ours. In particular, Freeman's market wage cannot be separated into mean perceived productivity and an insurance premium which depends only on age and precision. This makes it difficult to develop aggregate implications within his model. Freeman does not consider any aggregate implications, which is the main concern of our paper.

findings. Noteworthy is the implied positive relationship between wage variance and experience, a relationship which has been empirically established by Mincer and others. This positive correlation is a direct consequence of learning about productivity and lends strong support to the learning part of the model. The paper proceeds as follows. In the next section we present the model. Section 3 is devoted to proving the downward rigidity of individual contracts. In Section 4 we develop more detailed properties of optimal contracts, and in Section 5 we derive their aggregate empirical implications. In Section 6 we present a summary, make some comparisons with other models, and explain how our model could provide a basis for a dynamic theory of agency with moral hazard.

2. The Model

We consider a market with a large number of identical firms producing a single consumption good with labor as the only factor of production. Output of a worker depends on the worker's ability η, and a random disturbance term ε. Firm output is simply the sum of individual worker outputs; in other words, production is constant returns to scale.

Firms are infinitely lived, while workers live for T periods only. Generations overlap and are of equal size. Time is indexed by t. A worker of ability η will produce, in period t,

$$y_t = f(\eta, \varepsilon_t), \tag{1}$$

where ε_t is the realization of the random disturbance in period t.

The only decision variable for workers is the choice of firm for which to work. This choice will depend on the contracts firms offer. In particular, workers are assumed to derive no utility from leisure (and no disutility from working). They simply supply their labor inelastically to the firm offering the most attractive contract at the time. This assumption eliminates all moral hazard considerations, a point to which we shall return in the concluding section. Finally, we assume workers can change firms in any period at no cost.

When a worker enters the labor market his ability is not known with certainty. Prior beliefs regarding the ability η of a worker with observable characteristic e (e.g. education level) are assumed to be normal with mean $m_1(e)$ and precision $h_1(e)$. This observable characteristic, which we shall henceforth call education, has the property that workers with more education have higher mean productivity while prior beliefs about them are equally precise, i.e., m_1 is increasing in e and h_1 is constant. The distribution of education levels is assumed to be exogenous and the same for each generation.[7]

All information is assumed to be common knowledge. Firms as well as workers share common beliefs about η at all stages. Beliefs about η for any particular worker change as a function of his observed output y_t. How beliefs are updated will depend on the specific structure of the production function, to which we now turn.

In order to keep the analysis as simple as possible, we assume that current output is given by

---

[7] The observable education level will play no significant role until the aggregate properties of the model are considered in Section 5. Note that, since h_1 is independent of e, schooling in our model is purely an investment in productivity enhancing human capital. There are no screening or informational aspects. This will result in some implications which are at variance with observations as discussed in Section 5.

(2) $$y_t = \eta + \varepsilon_t$$

with ε_t's distributed independently and normally with mean 0 and precision 1. Therefore, posterior beliefs about η in period t (i.e. after observing $y_1,\ldots,y_t$) will also be normal. We denote by m_t and h_t the mean and precision, respectively, of this distribution. Note that expected output in period t and mean perceived ability in period t are the same, namely m_t.

The rules for updating beliefs about η are given by (see deGroot):

(3) $$m_{t+1} = \frac{h_t m_t + y_{t+1}}{h_t + 1},$$

(4) $$h_{t+1} = h_t + 1.$$

The precision variable h_t develops deterministically, whereas the mean follows a random walk with declining variance. Clearly, the history $y^t = (y_1,\ldots,y_t)$ is equivalent to the history $m^t = (m_1,\ldots,m_t)$. It can be shown that m_{t+1} and $[m_{t+1}|m^t]$ are normally distributed with means and variances as follows:

(5) $$E[m_{t+1}|m^t] = m_t,$$

(6) $$Var[m_{t+1}|m^t] = (h_t h_{t+1})^{-1},$$

(7) $$E(m_t) = m_1,$$

(8) $$Var(m_t) = \left(h_1(h_1+t-1)/t-1\right)^{-1}.$$

Equation (6) gives the conditional variance of m_{t+1}, which goes to zero as $t \to \infty$. Equation (8) gives the unconditional variance of m_t, which goes to $1/h_1$ as $t \to \infty$. Thus, unconditionally, the distribution of m_t converges to the prior distribution of η. That is the distribution of estimated abilities in the population of the oldest workers is approximately the same as the initial prior distribution of abilities for the youngest workers. The distribution of estimated (or expected) abilities in the population starts as a point mass on $m_1(e)$ for new workers with education e and gradually spreads out to coincide (approximately) with the prior distribution on η.

To complete the description of the model we state preferences. Workers are assumed risk averse with preferences over the consumption stream $c^T = (c_1,\ldots,c_T)$ given by

$$(9)\quad U(c^T) = \sum_{t=1}^{T} \beta^{t-1} u(c_t),\ \beta < 1.$$

As mentioned, workers have no utility for leisure. It is also assumed that workers may neither save nor borrow, i.e. their consumption stream must coincide with their earning stream. That workers may not save is not restrictive in that, in equilibrium they will not wish to do so. That they may not borrow is an important assumption of the model, which we will comment on below. The success of the model can be taken as an indication that the ability of workers to borrow is, in reality, severely limited (perhaps by lack of collateral) and not of great importance to the determination of wage contracts. Finally, we assume that risks associated with random fluctuations in perceived productivity cannot be insured except through the firm via the wage contract. There are several justifications for this assumption; the most compelling is based on a moral hazard argument. One may interpret our model as one in which workers do prefer less effort to more but effort is perfectly observable by the firm. This would result in the non-shirking behavior postulated for the workers of our model even if the firm provides insurance [see Harris-Raviv (1979)]. Assuming that no outside insurer can observe effort would imply that outside insurance is much less efficient than insurance provided by the firm.

Firms are risk neutral. They value a profit stream c^T by the index:

$$(10) \qquad \Pi(c^T) = \sum_{t=1}^{T} \beta^{t-1} c_t .$$

The discount factor β is common to all firms and workers.

The analysis to follow is aimed at characterizing competitive equilibrium wage contracts in this environment.

3. Equilibrium Contracts and Wage Rigidity

Firms are assumed to be able to make binding contingent contracts with workers. Workers, on the other hand, may not make binding commitments either to work for a particular firm at a particular wage in the future or to make payments to a firm in lieu of fulfilling such a commitment. That a worker may leave a firm to accept a better job elsewhere without financial obligation to the original firm we take as an exogenous feature of the environment, reflecting prohibitions against involuntary servitude. The fact that firms may make binding promises is usually defended on grounds of reputation [see Holmstrom (1981)], but as we will see, in our model optimal promises will really be non-contingent, so the reputation argument need not necessarily be invoked.

A state-contingent contract δ is defined as a collection of wage payments $\delta = \{w_t(m^t)\}_{t=1}^{T}$, where $w_t(m^t)$ is the wage to be paid in period t given the history m^t. An optimal contract can be found by solving the following program:

$$(11) \qquad \max_{\delta} \quad E_1\Big[\sum_{t=1}^{T} \beta^{t-1} u(w_t)\Big]$$

$$(12) \qquad \text{s.t.} \quad E_1\Big[\sum_{t=1}^{T} \beta^{t-1}(m_t - w_t)\Big] = 0,$$

$$(13) \qquad E_\tau\Big[\sum_{t=\tau}^{T} \beta^{t-\tau}(m_t - w_t)\Big] \leq 0, \ \forall\tau = 1,\ldots,T.$$

In this problem, E_s for $s = 1,\ldots,T$, denotes the expectation operator conditional on information up to time s, i.e. m^s. The constraints in (13) reflect the existence of a market for workers services: if expected profits from a particular worker as of time τ were positive, then that worker would be bid away by another firm offering a slightly better contract. Note that, for each τ, constraint (13) depends on the realized history m^τ although this has been suppressed in the E_τ notation.

Equation (12) requires that expected profits are equal to zero when the contract is signed. Subject to (12) and (13) an efficient contract maximizes the worker's expected utility.

It should be noted that (12) and (13) do not imply that (13) holds as an equality. Firms will have $m_1 - w_1 > 0$ to offset expected losses in later periods. Also, (13) does not imply that $m_\tau - w_\tau \leq 0$ for $\tau \geq 2$, only that the discounted expected profits are nonpositive. In general firms will make positive profits in early periods and losses in later periods, though it is, of course, perfectly possible that realized output is such that the firm makes profit (or losses) in all periods.

An optimal solution to the program (11)-(13) is actually very simple to state. Let us introduce the following definitions.

Definition. A wage policy $\{w_t(m^t)\}_{t=1}^{T}$ is said to be downward rigid if

$$w_{t+1}(m^{t+1}) \geqslant w_t(m^t) \text{ for all } m^T = (m_1,\ldots,m_T).$$

Wages never fall along any realized history if the wage policy is downward rigid.

Definition. A wage policy $\{w_t(m^t)\}_{t=1}^{T}$ is said to be upward rigid if

$$w_{t+1}(m^{t+1}) > w_t(m^t) \text{ for some } t,\ m^T,$$

implies that (13) is binding for $\tau = t+1$ given m^{t+1}.

In an upward rigid wage policy, wages are bid up only when the market constraint (13) requires it.

Definition. A wage policy which is downward and upward rigid is said to be rigid.

Theorem 1. A rigid policy is an optimal policy for the program (11)-(13).

The proof will be provided subsequent to showing how a best rigid wage policy is constructed. Notice that a rigid wage policy is simply a promise by the firm to pay the worker x units per period until the worker receives a better offer from the market. If such an offer is received, the old contract is cancelled and a new one is made which matches the market offer. The new contract guarantees a wage which is chosen so as to give zero expected profits (this follows from the fact that (13) is binding if the wage is bid up).

We will construct a best rigid wage policy by backward dynamic programming. Let $v_t(m,h,x)$ be the net value to a firm from offering a contract (wage guarantee) x to a worker with characteristics (m,h,t), where m is the perceived average ability of the worker, h is the precision of beliefs and t is the age of the worker. A worker of age t has worked t-1 periods and has T-t+1 periods left as a participant in the labor market. Thus v_T, for example, is the net value of a contract with a worker beginning his last productive period.

The net value v_t accounts for the fact that the worker will quit (or equivalently have his wage bid up) if his market value at some stage after t exceeds x. By market value we mean the highest wage guarantee he can receive from a firm making non-negative profits. Let $x_t(m,h)$ be the market value of a worker with characteristics (m,h,t). Formally, $x_t(m,h)$ is defined by

$$(14) \qquad v_t\big(m,\ h,\ x_t(m,\ h)\big) = 0.$$

We will see below that (14) gives a unique value for $x_t(m,h)$. Define $m_t(h,\ x)$ as the inverse function of $x_t(m,\ h)$ or equivalently through

$$(15) \qquad v_t\big(m_t(h,\ x),\ h,\ x\big) = 0.$$

A worker (m,h,t) with an m higher than $m_t(h,\ x)$ will not find x an acceptable contract. The reverse is true if m is less than $m_t(h,\ x)$. Therefore, $m_t(h,\ x)$ is the critical value for determining whether the wage will be bid up or not as a function of the position of the random walk $\{m_t\}$.

Starting from t=T we have:

$$
\begin{aligned}
& v_T(m,\ h,\ x) = m-x, \\
(16) \qquad & x_T(m,\ h) = m \quad , \\
& m_T(h,\ x) = x \quad .
\end{aligned}
$$

With one period to work, a worker's net expected value is simply his expected output m net of his wage payment x. His market wage at this point is simply his expected output m. Thus if his current expected output m is above x, then x will not be a sufficiently high wage guarantee to keep this worker from being bid away, and vice-versa.

Given v_t, x_t, m_t, we have:

$$(17) \qquad v_{t-1}(m,\ h,\ x) = m-x + \beta \int_{m' \leq m_t(h+1,x)} v_t(m',\ h+1,\ x)\ dP(m'|m,\ h),$$

Where $P(m'|m,h)$ is the distribution of next period's perceived mean ability m' for a worker whose current ability is normal with mean m, precision h, i.e., P is normal with mean m, precision h(h+1). The integration is over the region $m' \leq m_t(h+1,\ x)$ since if $m' > m_t(h+1,\ x)$ the worker's wage will have to be bid up implying zero expected profits for the firm.

Starting from (16) we can use (17) and then (14) and (15) to arrive at functions $v_t(m,\ h,\ x)$, $x_t(m,\ h)$, $m_t(h,\ x)$ for t=T,...,1. That this construction is well-defined follows from:

Lemma 1: For all t=1,...,T, $v_t(m,\ h,\ x)$ is strictly increasing in m and strictly decreasing in x. Therefore $x_t(m,\ h)$ is strictly increasing in m and $m_t(h,\ x)$ is strictly increasing in x.

Proof: The statements are true for $t=T$. Assume they are true for $t<T$. Consider $t-1$.

Let $m > \hat{m}$. Then, it follows from (17) by first-order stochastic dominance and the monotonicity of v_t that $v_{t-1}(m, h, x) > v_{t-1}(\hat{m}, h, x)$.

Let $x > \hat{x}$. Then, from (17),

$$v_{t-1}(m, h, x) - v_{t-1}(m, h, \hat{x})$$

$$\leq \hat{x}-x + \beta \int_{m' \leq m_t(h+1,\hat{x})} [v_t(m', h+1, x) - v_t(m', h+1, \hat{x})]dP(m'|m,h) \leq \hat{x}-x < 0.$$

The first step follows from the facts that $v_t(m', h+1, x) \leq 0$ for $m' \leq m_t(h+1, x)$ and $m_t(h+1, x) > m_t(h+1, \hat{x})$ by the induction assumption. The second step follows from the induction hypothesis on v_t.

The function $v_{t-1}(m, h, x)$ takes both positive and negative values. First, $v_{t-1}(m, h, m) < 0$ as the integral in (17) is negative. Secondly, v_t is decreasing in x so that v_{t-1} decreases at a higher rate than -1. Clearly then, there is an x satisfying (14) and similarly an m satisfying (15). The claimed properties of x_t and m_t follow trivially.

Q.E.D.

Given the functions $v_t(m, h, x)$, we can construct a rigid wage policy $\{w_t(m^t)\}_{t=1}^{T}$, which is feasible for the program (11)-(13) as follows:

(18) $$w_t(m^t) = \max_{1 \leq s \leq t} \{x_1(m_1, h_1), \ldots, x_t(m_t, h_t)\}, \; t=1,\ldots,T.$$

Lemma 2: The policy defined by (18) is the only rigid policy that is feasible for (11)-(13).

Proof. We can write (18) as

$$w_t(m^t) = \max\{w_{t-1}(m^{t-1}),\ x_t(m_t,\ h_t)\}. \tag{19}$$

Downward rigidity follows immediately. Upward rigidity follows, since if $w_t(m^t) > w_{t-1}(m^{t-1})$, then $w_t(m^t) = x_t(m_t,\ h_t)$ implying that $v_t(m_t,\ h_t,\ x_t(m_t,\ h_t)) = 0$, i.e. that (13) is binding.

Feasibility of (18) follows by (14), Lemma 1 and (19). Uniqueness follows since $v_1(m,\ h,\ x)$ is strictly decreasing in x by Lemma 1 so that there can only be one starting wage given $(m_1,\ h_1)$.

Q.E.D.

We are now ready to prove Theorem 1.

Proof of Theorem 1.

We show, using duality, that the only feasible, rigid policy in Lemma 2 is optimal from which the result follows.

Let $\{w_t^*(m^t)\}$ be any feasible rigid policy. Define the Lagrange multipliers $\lambda_t(m^t)$, t=1,...,T using (18) as follows:

$$\lambda_1 = u'(w_1^*)$$

(20)

$$\lambda_t(m^t) = u'(w_t^*(m^t)) - \lambda^{t-1}(m^{t-1}) = u'(w_t^*(m^t)) - u'(w_{t-1}^*(m^{t-1})),$$

where

$$\lambda^t(m^t) = \sum_{s=1}^{t} \lambda_s(m^s). \tag{21}$$

Note that $\lambda_1 > 0$ and $\lambda_t(m^t) \leqslant 0$ for t=2,...,T. Moreover $\lambda_t(m^t) = 0$ if and only if $w_t^*(m^t) = w_{t-1}^*(m^{t-1})$. Thus, by upward rigidity of

w_t^*, $\lambda_t(m^t) = 0$ unless (13) is binding. Therefore we have complementary slackness.

Now, consider the following program:

$$\text{(22)} \qquad \max_{\delta} E_1 \sum_{t=1}^{T} \beta^{t-1}\left(u(w_t) + \lambda^t(y_t - w_t)\right)$$

This is the Lagrangian to the program (11)-(13) (after multiplying constraints (13) by the unconditional probabilities of the respective m^t's), with dependencies on m^t suppressed. The first order conditions for (22) are, by concavity, sufficient for a solution of (22). But these conditions are simply (20)-(21) which $\{w_t^*\}$ satisfies by construction. Therefore $\{w_t^*\}$ solves (22). The solution is feasible for (11)-(13) by assumption. We have also complementary slackness as argued above. By duality, $\{w_t^*\}$ is therefore a solution to (11)-(13).

Q.E.D.

Remarks

1. Notice that the proof does not assume that a solution exists, proceeding from there on to use first-order conditions. Indeed, the use of Kuhn-Tucker conditions with a continuum of constraints is not generally permissible. What we use is weak duality; if one can find a feasible solution that satisfies complementary slackness by maximizing the unconstrained program, one has found a solution to the constrained program as well. This is valid with a continuum of constraints (provided expectations are finite) as is elementary to prove.

2. Once we have shown that an optimal solution exists, it can be proved that all optimal solutions have to be almost surely equal to (18). We omit the proof which is a straightforward variational exercise.

3. The proof that any feasible, rigid wage policy is optimal does not require normality. This proof would still be valid with the general output specification (1), provided only that the expectations (11)-(13) exist.

4. In setting up the program (11)-(13) it was assumed that the worker consumes his wage instantaneously. We could have allowed saving. Even if the worker could save, he would not like to do so given (18), since his marginal utility of consumption keeps declining as his wage increases. More generally, any saving the worker would wish to do could be done by the firm without violating the constraints in (13). Thus, the savings option is actually covered implicitly in our formulation.

5. Borrowing is not implicitly covered in program (11)-(13). Moving payments forward in time will violate constraints (13). The effect of borrowing is to allow consumption and wages to be separated. In this case a first-best risk sharing arrangement (in which workers' consumption is independent both of time and output realizations) could be achieved as follows. In the first period each worker would post a bond with his employer. This bond would equal the present value of the worker's maximum expected productivity over his lifetime (i.e. the present value of the sequence of expected productivities if he produced maximum output in each period) minus his current expected productivity, m_1.[8] The worker would borrow

---

[8] Obviously, with output normally distributed, there is no maximum expected productivity in each period. For this solution to be valid, one would have to rule out distributions with unbounded support. The normal output process could, of course, be approximated arbitrarily closely with one having bounded support. Alternatively, even with the normal output process, the

this bond at t = 1. He would then receive wage payments equal to initial expected productivity, m_1, in every period $t=1,\ldots,T$, and, in the last period of employment, his bond would be refunded with interest, provided he had not left the firm to accept a better paying job with another firm. The proceeds of the bond refund would be used to pay off the worker's loan. In this equilibrium, workers would be perfectly insured, consuming m_1 every period for sure, no worker would ever be bid away by another employer, even if his estimated productivity increased drastically, and there would be no risk that workers would be unable to repay their loans at age T. The fact that this arrangement is not observed in practice is due, no doubt, to the presence of extreme moral hazard problems. Workers could, presumably, quit to accept better paying jobs if they were willing to default on their loans. Indeed, although long-term borrowing by workers is often observed, such loans are always guaranteed by putting up some real, durable property, such as a house or car, as collateral.

4. Characteristics of the Optimal Contract

Let us characterize further the optimal wage contract.

Lemma 3: $v_t(m, h, x) = \hat{v}_t(m-x, h)$ and is concave in m and x.

Proof: The statement is true for t=T. Assume it is true for t<T. Then by (14) and (15) and the induction assumption,

solution described could be approximately first best in the sense that the probability that a worker could be bid away is arbitrarily small.

(23) $$x_t(m, h) = m - z_t(h),$$

(24) $$m_t(h, x) = x + z_t(h),$$

where $z_t(h) > 0$ for $t=1,\ldots,T-1$. Consider $t-1$. By (17) and the inductive assumption,

(25)
$$\begin{aligned} v_{t-1}(m, h, x) &= m-x + \beta \int_{m' \leq x + z_t(h+1)} \hat{v}_t(m'-x, h+1)dP(m'|m, h) \\ &= m-x + \beta \int_{m'' \leq x-m + z_t(h+1)} \hat{v}_t(m''+m-x, h+1)dP(m''|0, h), \end{aligned}$$

where the last equality follows upon a change of variable $m'' = m' - m$. Clearly $v_{t-1}(m, h, x) = \hat{v}_{t-1}(m-x, h)$. Therefore, $x_{t-1}(\cdot)$ and $m_{t-1}(\cdot)$, defined through (14) and (15), can be written as in (23), (24).

Concavity of v_{t-1} in m and x follows if $\hat{v}_{t-1}$ is concave in its first argument, which will be true if

(26) $$\int_{m' \leq x+z_t(h+1)} \hat{v}_t(m'-x, h+1)dP(m'|m, h)$$

can be shown concave in m. Define

(27) $$\tilde{v}_t(y, h) = \begin{cases} \hat{v}_t(y, h), & \text{if } y \leq z_t(h), \\ 0 & \text{, if } y > z_t(h). \end{cases}$$

$\tilde{v}_t$ is concave in y since $\tilde{v}_t\big(z_t(h), h\big) = 0$ and $\hat{v}_t$ is increasing and concave in y by the induction hypothesis. We can write (26) as

$$\int \tilde{v}_t(m'-x, h+1)dP(m'|m, h)$$

$$= \int \tilde{v}_t(m''+m-x,\ h+1)dP(m''|0,h)$$

which is concave in m since it is the integral of a function concave in m.

Q.E.D.

We are now able to give a fairly complete characterization of v_t, x_t and m_t:

Theorem 2: For every $t=1,\ldots,T$

(i) $\hat{v}_t(m-x,\ h)$ is strictly increasing in m, strictly decreasing in x, increasing in h and strictly so for $t \neq T$.

(ii) $x_t(m,\ h) = m-z_t(h)$, with $z_T(h) = 0$, $z_t(h) > 0$ and $z_t(h)$ strictly decreasing in h for $t \neq T$.

(iii) $m_t(h,\ x) = x + z_t(h)$.

Proof: Except for the claims concerning changes in h, the theorem has been proved in the previous lemmata.

The statement that $z_T(h) = 0$ is a consequence of (16). The fact that $z_t(h) > 0$ follows by monotonicity of $\hat{v}_t$ and the definition of x_t.

Assume now that $z_t(h)$ is non-decreasing. If we can show $z_{t-1}(h)$ is strictly decreasing in h, we have proved the theorem. This is equivalent to showing that $\hat{v}_{t-1}(m-x,\ h)$ is strictly increasing in h given that $\hat{v}_t$ is non-decreasing in h. To prove that, let $h' > h$. Then,

$$\hat{v}_{t-1}(m-x,\ h) \leq m-x + \beta \int_{m' \leq m_t(h'+1,x)} \hat{v}_t(m'-x,\ h+1)dP(m'|m,\ h)$$

(since $m_t(h'+1,\ x) \leq m_t(h+1,\ x)$ by the induction hypothesis and $\hat{v}_t \leq 0$ in the integral),

$$\leqslant m-x + \beta \int_{m' \leqslant m_t(h'+1,\ x)} \hat{v}_t(m'-x,\ h'+1) dP(m' | m,\ h),$$

(since $\hat{v}_t$ is increasing in h by the induction hypothesis),

$$< m-x + \beta \int_{m' \leqslant m_t(h'+1,\ x)} \hat{v}_t(m'-x, h'+1) dP(m' | m,\ h'),$$

(since $P(m' | m,\ h)$ is a mean preserving spread of $P(m' | m, h')$ and $\hat{v}_t$ is concave in m' by Lemma 3),

$$= \hat{v}_{t-1}(m-x,\ h'). \qquad\qquad \text{Q.E.D.}$$

<u>Theorem 3</u>: For $t' > t$,

(i) $v_t(m,\ h,\ x) < v_{t'}(m,\ h,\ x)$,

(ii) $z_t(h) > z_{t'}(h)$.

<u>Proof</u>: $v_T(m,\ h,\ x)$ is given by (16). From (17) it follows that $v_{T-1}(m,\ h,\ x) < v_T(m,\ h,\ x)$, since the integral is negative. Thus, $z_{T-1}(h) > z_T(h) = 0$. Assume $v_t(m,\ h,\ x) < v_{t+1}(m,\ h,\ x)$ and $z_t(h) > z_{t+1}(h)$ for $t < T$. From (17)

$$v_{t-1}(m,\ h,\ x) = m-x + \beta \int_{m' \leqslant m_t(h+1,\ x)} v_t(m',\ h+1,\ x) dP(m' | m,\ h)$$

$$< m-x + \beta \int_{m' \leqslant m_{t+1}(h+1,\ x)} v_{t+1}(m',\ h+1,\ x) dP(m' | m,\ h)$$

$$= v_t(m,\ h,\ x).$$

The inequality follows, since $m_{t+1}(h+1, x) < m_t(h+1, x)$ as $z_t(h+1) > z_{t+1}(h)$, the integrand is negative, and, from the induction hypothesis, $v_t < v_{t+1}$. This proves $v_{t-1} < v_t$ and hence (i) for all t, t'. Part (ii) follows immediately from (i).

Q.E.D.

Theorems 1-3 characterize completely the wage dynamics in the labor market we are modeling. We find that contract wages are downward rigid and are bid up only as the market value of the worker exceeds the wage guarantee in his current contract. The market value as given by $x_t(m, h)$ is furthermore linearly increasing in perceived ability m. A discount $z_t(h) > 0$ is deducted from the worker's marginal product m in computing the market value (the contract wage). This discount is part of an insurance premium paid by the worker up front in order to insure against low realizations of ability later on. The up side cannot be insured because the worker is not permitted to sell himself to the firm. The premium is smaller the lower the worker's variance (the higher h) since with lower variance there is less risk on the downside to be insured. In the limit as $h \to \infty$, $x_t(m, h) \to m$, and the worker is paid his marginal product. This is depicted in Figure 1 (in the figure, $t' > t$).

Looking at the random walk $\{m_t\}$ we can illustrate the same qualitative features using a somewhat different format in Figure 2. As long as m_t stays below the curve in Figure 2, the contract wage x is feasible and will not be bid up. If m_t crosses into the region "NEW CONTRACT", the wage x is no longer viable because it implies positive profits for the firm. A new contract is written with a higher guaranteed wage. Notice that the critical boundary is sloping downward since $m_{t+1}(h_{t+1}, x) < m_t(h_{t+1}, x) < m_t(h_t, x)$ as $h_{t+1} > h_t$, using both Theorems 2 and 3.

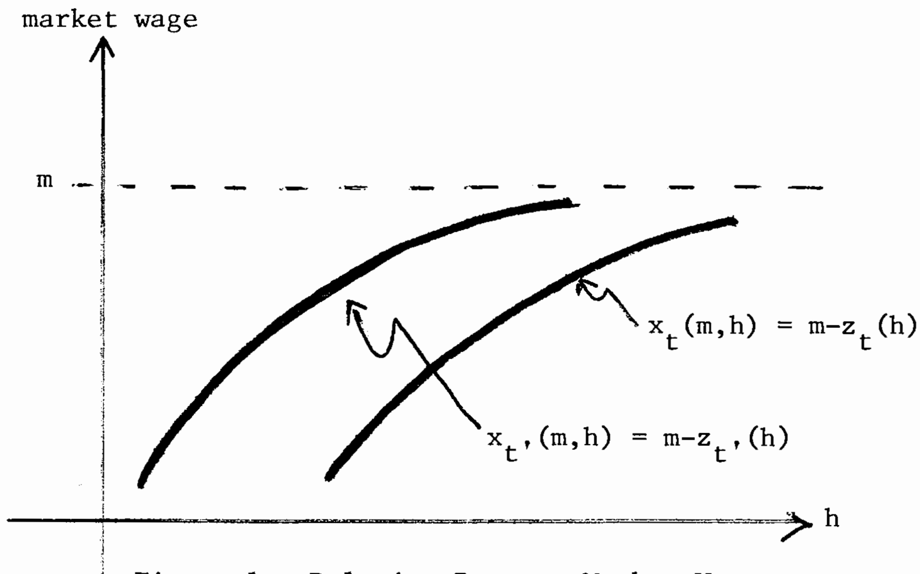

Figure 1: Relation Between Market Wage and Precision for Given Age

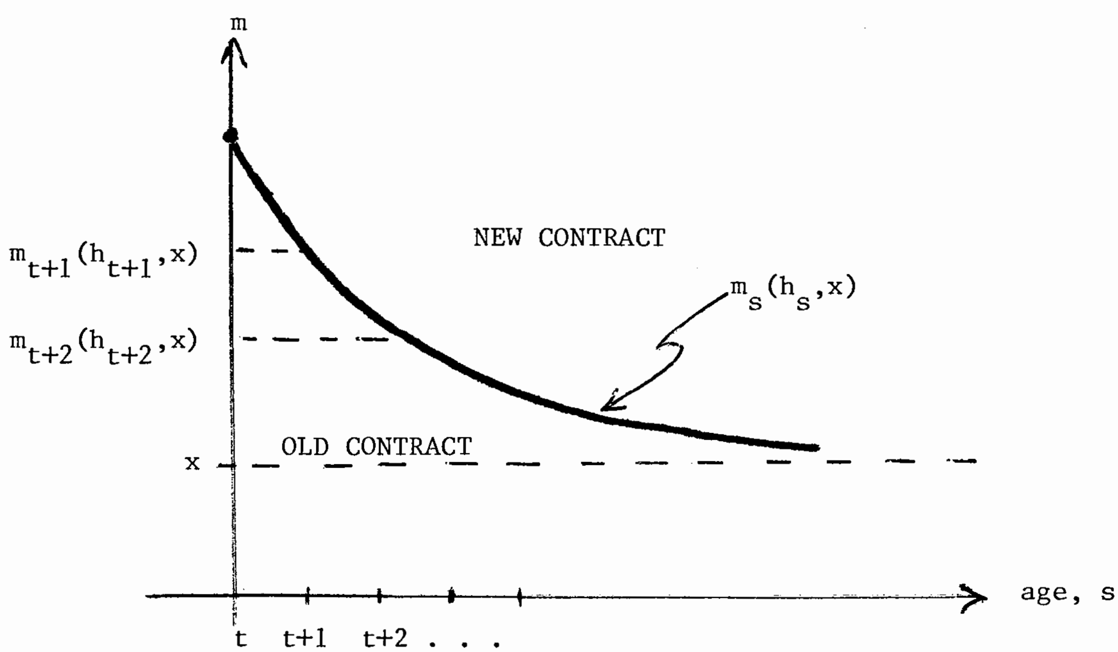

Figure 2: Critical Value of Perceived Productivity as a Function of Age

5. Aggregate Behavior

Having described the features of optimal individual contracts we turn now to predictions about aggregates; how do the three main variables of the model, age, ability and wage, relate to each other on the economy level? In particular, can our model account for some of the more important stylized facts concerning these relationships? We begin by listing some empirical observations:[9]

1. Experience and Earnings. Experience and earnings are positively related, but the rate of growth of earnings decreases with experience. [Mincer (1974, Chapter 4) and Becker (1975, Chapter 7)]. Moreover, in-company experience is more strongly related to earnings than pre-company experience. [Medoff and Abraham (1980)].

2. The Variance of Earnings. The variance of earnings increases with experience and with the level of schooling. [Mincer (1974, pp. 99-101)].

3. Earnings, Experience and Productivity. Earnings increase more with experience than can be accounted for by productivity increases. [Medoff and Abraham (1980), Pascal and Rapping (1972, Table 4-3).

4. Schooling and Earnings. Schooling and earnings are positively related and the absolute increment to earnings from additional schooling

---

[9] The list is not meant to be exhaustive. In particular, we have omitted observations about which our model either has no implications or such implications could not be derived.

tends to increase with experience [Mincer (1974, p. 75)]. Workers with more schooling achieve higher earnings mainly because they are assigned to higher grade levels. [Medoff and Abraham (1980)].

5. Skewness. The distribution of earnings is positively skewed (the mode is below the mean) overall as well as within any schooling-experience group. [Mincer (1974, Chart 6.4 and Section 6.2)].

Except for point 3, these results are not particularly controversial. Some comment on point 3 is, however, in order. The Medoff and Abraham study measures productivity using relative performance ratings submitted by supervisors for workers within a given job category. Controlling for job category is essential in order to use performance ratings as a productivity measure. Their results are subject to criticism on the grounds that performance evaluations may reflect the opinion of the evaluator as to the performance of the worker relative to an expected performance based on the worker's level of experience (there is some evidence of this in the M-A data). Thus, younger workers may be rated higher than older ones in a given job not because they performed better absolutely, but because they performed better for their age. Also one might criticize the Pascal and Rapping study for not controlling for such non-measurable aspects of productivity as "star quality" (a defect which they recognize but are unable to correct). Both of these criticisms reflect the difficulties one faces in measuring productivity directly, and much more empirical work appears necessary before definite conclusions can be drawn. Nevertheless, it seems to us that some increase in earnings in excess of increases in productivity is evidenced by these empirical studies.

Experience, Earnings, and the Variance of Earnings.

Our first aggregate implication addresses points 1 and 2. We show that, for any schooling group, the experience-earnings profile is upward sloping. We also show that, for any schooling group, the variance of earnings increases with experience. Naturally these results also hold when averaged over schooling groups. These results are consistent with the evidence cited in 1 and 2 above.

Theorem 4: (i) $E(w_{t'}|e) > E(w_t|e)$, for $t' > t$.

(ii) $Var(w_{t'}|e) > Var(w_t|e)$, for $t' > t$.

Proof: Part (i) is an immediate consequence of the fact that $w_{t+1} = \max(w_t, x_{t+1})$.

Part (ii) we prove by induction. Comparing t=1 with t+1 = 2 is trivial. Assume (ii) is true for t < T and consider t+1. Let $\tilde{w}_t = \max(x_2-x_1, x_3-x_1, \ldots, x_t-x_1)$. By the independent increments property, the induction hypothesis implies $Var(\tilde{w}_t) > Var(\tilde{w}_{t-1})$. It is also clear that rigid wages imply that the distribution of $\tilde{w}_t$ dominates the distribution of $\tilde{w}_{t-1}$ in a first-order stochastic dominance sense. From this follows that $Var[\max(0,\tilde{w}_t)] > Var[\max(0,\tilde{w}_{t-1})]$. Note that

$$w_t = x_1 + \max(0,\tilde{w}_{t-1}),$$

$$w_{t+1} = x_1 + \max(0,\tilde{w}_t).$$

Since x_1 and $\tilde{w}_{t-1}$ and x_1 and $\tilde{w}_t$ are independent we have:

$$Var(\tilde{w}_{t+1}) = Var\ x_1 + Var[\max(0,\tilde{w}_t)]$$

$$> Var\ x_1 + Var[\max(0,\tilde{w}_{t-1})] = Var(w_t).$$

Note that Var x_1 is zero conditional on e.

Q.E.D.

With respect to concavity of the earnings profile mentioned in point 1 (i.e. average earnings grow with experience at a decreasing rate), the model does exhibit this behavior as a limiting feature since the random walk $\{m_t\}$ has an incremental variance that goes to zero [by equation (6)]. Therefore, both expected wage and wage variance will converge with age (for large T). We do not currently have results that compare rates of growth between consecutive periods, however.

Regarding the more specific evidence concerning in-company vs. pre-company experience, in our model, workers and firms are indifferent as to whether a worker changes firms or simply receives a higher wage at his current firm when his market wage exceeds his current wage guarantee. If a worker does change firms at such times, then he will generally earn less than his contemporaries in the same job who have been with the firm longer. This is simply because the worker who has just switched firms will be earning exactly the market wage for his job and age level. The other people in that job and of the same age will be earning at least this market wage and some will be earning more. Thus, in-company experience is worth more than pre-company experience. Furthermore, pre-company experience does have a positive value on its own, because, according to Theorems 2 and 3, experience (i.e. age) increases the worker's market value as the insurance premium $z_t(h)$ becomes smaller. In this sense, even without being specific about mobility, we can state that our model is consistent with both the general and the specific part of point 1.[10]

With respect to the relationship between variance of earnings and schooling (holding experience constant) in point 2, these are independent in our model. This, however, is an artifact of our assumption that schooling is purely an investment in productivity enhancing human capital and has no sorting role. A more explicit model of schooling would almost certainly include its role in providing information about native ability. In such a model, not all workers with the same level of schooling would be viewed as identical in their first year of work experience (as they are in the present model). Instead, education would tend to increase the precision of the productivity estimate for individual workers while increasing the dispersion of estimates across workers with the same value of e. When calculating the variance of earnings for a particular schooling group in this version, one would be aggregating over workers who started with different estimated productivities, and, indeed those groups with more schooling would have more widely dispersed initial estimates. In such a model, the variance of earnings for a given level of experience would increase with schooling.

Earnings, Experience and Productivity.

In connection with point 3 above, we wish to show that, controlling for estimated productivity m_t, older workers earn more on average than younger ones, i.e.

---

10 The indifference between changing firm or staying could formally be broken by viewing the production function in (2) as the envelope of all firms' production functions. At any particular moment, the worker may actually have to move in order to capitalize on an improved perception of his ability, because a task matching this productivity is not available in his present firm. The implications on in-company vs. pre-company experience drawn above are best interpreted in this light.

(28) $$E(w_{t'} | m_{t'} = m) > E(w_t | m_t = m), \text{ for } t' > t.$$

At first blush (28) may appear as obvious as part (i) of Theorem 4. On one hand senior workers have a higher market wage by Theorem 3 and on the other hand they should have had more chances to have had their wage bid up than junior workers. But a closer look at the problem reveals that the statement is less than obvious (because of the conditioning). In fact, we have failed to prove (28) as stated (though we still think it is true). What we are currently able to prove is that if one goes to the continuous time case where the random walk $\{m_t\}$ becomes (essentially) Brownian motion, then (28) holds. Therefore, with sufficiently short time intervals between recontracting we find that age and earnings are positively related even when one controls for the grade level (and hence perceived productivity).

In the continuous time version, workers are productively active in the time interval $(1,T)$. The continuous time case can be arrived at by a limiting process in which the time interval $(1,T)$ is divided into increasingly finer partitions. For each partition n (the number of subintervals), we will get a premium function $z_n(t)$ (suppressing h for notational convenience as we only consider a single worker) by solving the discrete case as described earlier and extending the discrete function $z_n(\cdot)$ over the whole range $(1,T)$ in the obvious fashion to a step function. Since $z_n(t)$ is decreasing in t (Theorem 3) and since an upper bound can easily be given so that $0 < z_n(t) < M$ for all n and t, Helly's extraction principle will assure that a subsequence n' will converge to a limit function $z(t)$, which also is decreasing (actually strictly). From this $z(t)$, one may recover the functions v, x, and m. In particular,

$$x(m,t) = m-z(t).$$

On the other hand, as the partitions become finer the random walk $\{m_t\}$ becomes a continuous time stochastic process $M(t)$, which is Brownian motion with dying variance. In analogy with formulas (7) and (8) we have

$$E\left(M(t)\right) = M(1), \ t\varepsilon[1,T],$$
$$Var\left(M(t)\right) \equiv \sigma(t) = \left(h_1(h_1+t-1)/(t-1)\right)^{-1}, t\varepsilon[1,T].$$

Thus, $\sigma(1)=0$, $\sigma(t)$ is increasing and converges to h_1 as $t \to \infty$. Without loss of generality, we may take $M(1)=0$. We define the process $W(t)$ as

$$W(t) = \max_{1\leq s\leq t} x\left(M(s),t\right) = \max_{1\leq s\leq t} \left\{M(s) - z(s)\right\}.$$

Note that $M(s)$ is stochastic, whereas $z(s)$ is deterministic.

We may now state and prove the main result of this section:[11]

Theorem 5: For $t' > t$,

(29) $$E\left[W(t')\,|\,M(t') = m\right] > E\left[W(t)\,|\,M(t) = m\right].$$

What is claimed in the theorem is that if we look at the average of the maximum of $M(s) - z(s)$ taken over all paths $M(s)$ that start at $M(1) = 0$ and end up at $M(t') = m$, then this average is bigger than the corresponding one

---

11 The proof of Theorem 5 was suggested by David Kreps.

taken over the paths starting at $M(1) = 0$ and ending up at $M(t) = m$.

Proof: Considering $M(s)$ over the interval $(1,t)$ we may transform the process to an equivalent one by redefining time twice; first so that time is read as $\hat{\tau} = \sigma(s)$ and secondly so that it is read as $\tau = \hat{\tau}/\sigma(t)$. Since $\sigma(1) = 0$, this will map the time interval $(1,t)$ into $(0,1)$, i.e. as s goes from 1 to t, τ goes from 0 to 1. Through this transformation of time, we have that $M(s)$ behaves between $(1,t)$ equivalent to the process
$Y(\tau\sigma(t))$ between $\tau = 0$ and $\tau = 1$, where $Y(\cdot)$ is standard Brownian motion. Moreover, by the property of Brownian motion we have

$$Y(\tau\sigma(t)) \sim \sqrt{\sigma(t)}\ Y(\tau),$$

where $\sim$ refers to probabilistic equivalence.

We can do the same sequence of transformations for $M(s)$ over $(1,t')$ to the equivalent process $\sqrt{\sigma(t')}\ Y(\tau)$ where τ runs from 0 to 1. As we make these time changes we also have to transform the function $z(s)$. Let

$$g(\tau) = -z(\sigma^{-1}(\tau\sigma(t))),$$
$$g'(\tau) = -z(\sigma^{-1}(\tau\sigma(t'))).$$

Note that $g(\cdot) < g'(\cdot)$ since $z(\cdot)$ is decreasing. We have then the two representations:

$$M(s) - z(s) \sim aY(\tau) + g(\tau),$$
$$M(s) - z(s) \sim a'Y(\tau) + g'(\tau),$$

where $a = \sqrt{\sigma(t)}$, and $a' = \sqrt{\sigma(t')}$. Note that $a' > a$ since $\sigma(\cdot)$ is increasing. For the case where $s \varepsilon (1,t)$ we consider the first equivalence, for the case where $s \varepsilon (1,t')$ we consider the second one.

We need one further transformation to get rid of the end point conditions $aY(1) = m$ and $a'Y(1) = m$ (corresponding to $M(t)=m$ and $M(t')=m$). What we have is a Brownian bridge from 0 to m/a (and 0 to m/a' respectively). It is well known [see Billingsley (1968)] that a Brownian bridge $Y(0) = 0$ to $Y(1) = 0$ behaves equivalently to the process $Y(\tau) - \tau Y(1)$ with $Y(0) = 0$ and no end point condition for $\tau = 1$. Thus $aY(\tau) + g(\tau)$ with conditions $Y(0) = 0$, $Y(1) = m/a$ is equivalent to

$$(30) \qquad a\left(Y(\tau) - \tau Y(1)\right) + \tau m + g(\tau)$$

and similarily for the primed case:

$$(31) \qquad a'\left(Y(\tau) - \tau Y(1)\right) + \tau m + g'(\tau).$$

After these transformations our claim (30) can be rephrased as:

$$(32) \qquad E\left[\max_{0 \leq \tau \leq 1} \left(a'(Y(\tau) - \tau Y(1)) + \tau m + g'(\tau)\right)\right] >$$

$$E\left[\max_{0 \leq \tau \leq 1} \left(a(Y(\tau) - \tau Y(1)) + \tau m + g(\tau)\right)\right].$$

To prove this we make a path by path comparison. Since we are only concerned with expectations we may indeed think of $Y(\tau)$ as the same process both for the case $(1,t)$ and $(1,t')$.

Let $y_+(\tau)$ be any fixed path of $Y(\tau)$. By symmetry of Brownian motion the path $y_-(\tau) = -y_+(\tau)$ is an equally likely path. Since the paths defined by y_+ and y_- are equally likely, the expected maxima in (32) will equal the expectation of the sum of the maxima of such pairs of paths for each process. It suffices therefore to show that the sum of the maxima over the paths defined by y_+ and y_- for the primed process (31) exceed the corresponding sum for the unprimed process (30).

To prove this let τ_+ and τ_- be the times at which the paths defined by y_+ and y_- achieve their maxima for the process (30) and τ_+' and τ_-' be the corresponding times for (31). Also let $k_+(\tau) = y_+(\tau) - \tau y_+(1)$ and similarly for $k_-(\tau)$. We must prove that

$$(33) \qquad \begin{aligned} &a'k_+(\tau_+') + m\tau_+' + g'(\tau_+') + a'k_-(\tau_-') + m\tau_-' + g'(\tau_-') \\ &\geq ak_+(\tau_+) + m\tau_+ + g(\tau_+) + ak_-(\tau_-) + m\tau_- + g(\tau_-). \end{aligned}$$

But

$$\begin{aligned} &a'k_+(\tau_+') + m\tau_+' + g'(\tau_+') + a'k_-(\tau_-') + m\tau_-' + g'(\tau_-') \\ &\geq a'k_+(\tau_+) + m\tau_+ + g'(\tau_+) + a'k_-(\tau_-) + m\tau_- + g'(\tau_-), \end{aligned}$$

so, after some manipulation, we see that it suffices to show that

$$(34) \qquad \begin{aligned} &(a'-a)[k_+(\tau_+) + k_-(\tau_-)] + g'(\tau_+) - g(\tau_+) \\ &\qquad + g'(\tau_-) - g(\tau_-) \geq 0. \end{aligned}$$

As noted above, $g' > g$ and $a' > a$, so it suffices to show that $k_+(\tau_+) + k_-(\tau_-) \geq 0$. By definition of τ_+ and τ_-, we have

$$ak_+(\tau_+) + m\tau_+ + g(\tau_+) \geq ak_+(\tau_-) + m\tau_- + g(\tau_-),$$

$$ak_-(\tau_-) + m\tau_- + g(\tau_-) \geq ak_-(\tau_+) + m\tau_+ + g(\tau_+).$$

Adding these two and noting that $k_{-}(\tau) = - k_{+}(\tau)$, we get the desired result.

Q.E.D.

The intuition behind this proof is relatively simple. The process which lasts for a longer time interval (from 1 to t') can be viewed as a process on the shorter time interval (from 1 to t) with an increased variance. That the process with higher variance will have a higher expected maximum appears intuitive. Yet we have to note that the end points are tied down and that there is some upward drift. These are the two aspects that complicate the proof.

Economically, the basic idea here is that older workers of the same perceived average productivity as younger workers are paid more because they pay less for insurance against decreases in their perceived ability. The reason for this is that the estimates of their abilities are more precise.

<u>Schooling and Earnings.</u>

That schooling and earnings are positively related in our model is shown in

<u>Theorem 6.</u> $E(w_t | m_1 = m)$ is increasing in m for all $t=1,\ldots,T$. Therefore $E(w_t | e)$ is increasing in e for all t.

<u>Proof.</u> We first show that $E(w_t | m_t)$ is increasing in m_t for all t. The statement is obvious for t=1 since $w_1 = m_1 - z_1(h_1)$. Assume it is true for $t<T$ and consider t+1. The distribution of m_t conditional on m_{t+1} is given by a normal distribution, which is shifted in a first-order stochastic dominance

sense as m_{t+1} is increased. Since $E(w_t|m_t)$ is assumed by the induction hypothesis to be increasing in m_t, so is $E(w_t|m_{t+1})$. The claim follows from the fact that $E(w_{t+1}|m_{t+1}) = E\big(\max(w_t, x_{t+1}(m_{t+1}, h_{t+1}))|m_{t+1}\big)$.

Now $E(w_t|m_1 = m) = E[E(w_t|m_t)|m_1 = m]$. But conditional on $m_1 = m$, m_t is normally distributed with mean m. Thus, an increase in m shifts the distribution of m_t to the right. The theorem now follows since $E(w_t|m_t)$ is increasing in m_t. Q.E.D.

It can easily be shown, using an argument similar to part (ii) of Theorem 4, that the increment to earnings due to additional schooling is independent of experience. This result is inconsistent with part of point 4. Again, the problem might be remediable by adopting a model which accounts for the sorting role of schooling.

In order to address the finding of Medoff and Abraham that very little of the within-grade-level variation in earnings is accounted for by education differences, we must introduce grade levels into our model. One simple way of doing this is to identify a grade level with mean estimated productivity m.[12] It can then be shown that, controlling for current mean estimated productivity and, therefore, current job level, average earnings are higher for more educated workers (see the proof of Theorem 6) but the difference is smaller than if one does not control for current job level. By how much the average earnings premium due to education is reduced when one controls for the job level we are unable to say. We only note that the wage differential due to education is fully explained by the job level in the initial period and later on less than that.

---

[12] It is possible to construct a simple model of the job assignment procedure in which the optimal assignment is uniquely determined in a one-to-one fashion by the worker's current expected productivity.

Skewness.

Concerning the observed skewness of the earnings distribution (see point 5 above), our model does deliver such a distribution for each education - experience group (provided experience is at least two periods). This is due to the fact that downward rigidity of the wage contract will truncate the lower tail of the distribution but not the upper one. For instance, in a two-period model, the wage distribution for the cohort in their second year would be a normal distribution truncated at a point below its mean with a mass at that point. Holding experience constant, skewness in our model is independent of schooling.

6. Conclusion

We have presented a model of wage dynamics based on learning about worker ability. Due to worker risk aversion and the idiosyncratic nature of ability uncertainty, optimal wage contracts will entail wage guarantees which insure the workers against adverse changes in perceived abilities. On the other hand, wage will be bid up with improved perceptions of ability to prevent the worker from accepting offers from competing firms. We have shown that the resulting downward rigid wage contracts, when aggregated over individuals, can accomodate recent empirical evidence concerning the earnings profile of skilled labor. In particular, long-term contracting as developed in the paper is consistent with findings that experience and earnings are positively related even when controlled for by productivity.

Some other theories that can explain this main point are those of Becker and Stigler (1974), Lazear (1979), Grossman (1978), and Salop and Salop (1976). All are based on an idea of deferred payments. In Lazear's work the reason for deferring wages is to provide the worker with an incentive not to shirk. In Grossman's model the issue is worker reliability. A junior worker is on average more likely to quit and therefore is forced to pay a higher premium for his wage guarantee. In Salop and Salop's model the concern is also quitting, but the model is one of adverse selection in which wage profiles screen quitters from non-quitters.

As such these models appear too simple to accomodate some of the other empirical evidence. For example, in these models all workers of a given age receive the same wage if they started at the same position. If job changes take place, no premium should be paid for pre-company experience at least when controlled for job level, which is in conflict with the specifics of point 1 above. Regarding point 2 on wage variance, no variance at all is implied. It is also unclear whether earnings in these models will grow at a decreasing rate as should be the case with ours.

On the other hand, our model is inconsistent with mandatory retirement, a phenomenon that Lazear (1979) cites as support for his incentive model of wage profiles. There are no Pareto gains to the institution of mandatory retirement in our model, though this conclusion would change by including a value for leisure. Furthermore, with a retirement option, firms in our model would like to retire workers with lower than expected realizations of productivity, while Lazear cites evidence that there is a positive correlation between unanticipated wage growth and mandatory retirement.

Of course, one may mix some of the theories mentioned above to arrive at predictions more fully consistent with evidence. Indeed, all theories may have partial bearing on the evidence of earnings profiles. In that regard we have to await more specific parameterized models to test the relative strength of the various causes as well as derivations of more diverse predictions of the models that can be refuted by the available evidence.

A main reason for building our model has been the goal of developing a dynamic market analysis of agencies. It seems intuitively plausible and has been suggested by Fama (1980) that the problem of moral hazard, particularly in managerial positions, is largely alleviated by the concern for repuation in the labor market. Indeed, Fama argues that ex post settling up in the market place will police managerial behavior adequately when managers are risk neutral and the discount rate is close to zero.

Fama's conclusions are not of general validity. They rely heavily on no discounting and on an implicit assumption that managerial decisions in equilibrium do not affect the statistical properties of the learning process.[13] His analysis is also confined to a long run stationary state, where learning about ability is offset by shocks in ability. This leaves open an important question about convergence to such a stationary state. The model we have presented is a first and natural step in the direction of investigating these complex issues. The obvious extension is to model worker effort explicitly. We do not imply that the task is easy, but the relative success of our simpler model with regard to the empirical evidence makes us hopeful that we are on the right track.

---

13 Fama is not explicit about the manager's actions. His conclusions can be arrived at by taking $y_t = a_t + \eta_t + \varepsilon_t$ and $\eta_{t+1} = \eta_t + \delta_t$, where y_t is output, a_t is the manager's action, η_t is managerial ability, ε_t is the random component of noise in observing output and δ_t is a disturbance term in the manager's ability. In the stationary state of this model, learning about η_t through y_t will be just offset by the shocks δ_t and the manager will have a constant incentive to take the optimal action (when there is no discounting).

References

Azariadis, C., "Implicit Contracts and Underemployment Equilibria," JPE, 83, 6, Dec. 1975, 1183-1202.

Baily, M.N., "Wages and Employment Under Uncertain Demand," R. Econ. Studies, 41, 1, Jan. 1974, 37-50.

Becker, G.S., "Investment in Human Capital: A Theoretical Analysis," JPE, 70, 5, pt. 2, Oct. 1962, 9-49.

Becker, G., Human Capital, second edition, Chicago: University of Chicago Press, 1975.

Becker, G. and G. Stigler, "Law Enforcement, Malfeasance, and Compensation of Enforcers," J of Legal Studies, 1974, 1-18.

Billingsley, P., Convergence of Probability Measures, New York: John Wiley & Sons, 1968.

DeGroot, M., Optimal Statistical Decisions, New York: McGraw-Hill, 1970.

Fama, E., "Agency Problems and the Theory of the Firm," JPE, 88, 2, April 1980, 288-307.

Freeman, S., "Wage Trends as Performance Displays Productive Potential: A Model and Application to Academic Early Retirement", Bell J, 8, Autumn 1977, 419-443.

Grossman, H., "Risk Shifting and Reliability in Labor Markets", Scandinavian J. Econ., 79, 2, 1977, 187-209.

Harris, M. and A. Raviv, "Optimal Incentive Contracts With Imperfect Information," JET, 20, 2, April, 1979, 231-259.

Holmstrom, B., "Equilibrium Long-Term Labor Contracts," Northwestern University working paper, January, 1980; (revised Janury, 1981).

Holmstrom, B., "Contractual Models of the Labor Market," American Economic Review, Papers and Proceedings, May 1981.

Jovanovic, B., "Job Matching and the Theory of Turnover," JPE, 87, 5, pt. 1, Oct. 1979, 972-989.

Lazear, E., "Agency, Earnings Profiles, Productivity, and Hours Restrictions," University of Chicago working paper, December 1980, forthcoming AER, 1981.

Lazear, E., "Why Is There Mandatory Retirement?" JPE, 87, 6, Dec. 1979, 1261-1284.

Medoff, J.L. and K.G. Abraham, "Experience, Performance and Earnings," QJE, Dec. 1980, 703-736.

Mincer, J., Schooling, Experience, and Earnings, New York: NBER, 1974.

Mortensen, D., "Specific Capital and Labor Turnover," Bell Journal of Economics 9, 2, Autumn 1978, 572-586.

Pascal, A.H. and L.A. Rapping, "The Economics of Racial Discrimination in Organized Baseball," in Racial Discrimination in Economic Life, A.H. Pascal (ed.), Lexington, Mass: Lexington Books, 1972, pp. 119-156.

Prescott, E. and M. Visscher, "Organization Capital," JPE, 88, 3, June 1980, 446-461.

Ross, S., P. Taubman, and M. Wachter, "Learning by Observing and the Distribution of Wages," NBER Conference Paper #44, June 1980.

Salop, J. and S. Salop, "Self-Selection and Turnover in Labor Markets," QJE, 90, Nov. 1976, 619-627.