Early Starters versus Late Beginners

Pierre-André Chiappori

University of Chicago

Bernard Salanié

Institut National de la Statistique et des Études Économiques, Centre de Recherche en Economie et Statistique, and Centre National de la Recherche Scientifique

Julie Valentin

Eurequa, Université Paris 1

We consider a model of wage formation characterized by two features: learning and downward rigidity. We show that wages should exhibit a late-beginner property: when one controls for the wage at date t, the wage at date $t + 1$ should be negatively correlated with the wage at date $t - 1$. We test this property on a sample of about 1,000 executives of a French state-owned firm whose careers we observe for 15 years. This organization exhibits the features that characterize internal labor markets; in particular, careers consist of sequences of discrete promotions, a fact that generates specific econometric problems. The results confirm the prediction.

I. Introduction

Most quantitative predictions of contract theory bear on the relationship between outcomes (''performances'') and monetary transfers.

In the case of labor contracts, different sets of assumptions about the information structure generate different conclusions about how wages depend on the worker's ''performance,'' as defined by some adequate index. In principle, such predictions are empirically testable. A major source of difficulty, however, lies in the fact that, in many situations, performance, although observed by the employer, is not available to the econometrician.[1] In most cases, indeed, data sets contain information only on wage or career profiles, plus a few worker-specific characteristics (age, training, etc.). In such cases, the static predictions of the models can turn out to be quite difficult to test.

In order to overcome the lack of performance proxies, a natural solution is to turn to dynamics and to derive, from the various theoretical explanations at stake, different implications for the random process followed through time by wages or promotions. Instead of focusing attention, as usual, on the way wages at date t depend on performances at the same period, we may, in other words, concentrate on the way wages at time t depend on wages at time $t - 1$ and before. For empirical investigations of this kind, data on performances or individual outcomes are no longer necessary. What is needed is simply a record of wage profiles plus, of course, a theoretical characterization of the dynamic properties of wages.

The main objective of this paper is to provide an illustration of this general research direction. We consider a class of models that is characterized by two features: the employer and the worker simultaneously learn about the worker's ability from the observation of his or her performance, and wages are downwardly rigid.

From now on, we shall refer to this class as the LDR (for learning and downward rigidity) models. The class includes in particular the model of insurance and Bayesian learning proposed by Harris and Holmstrom (1982). In this framework, downward rigidity arises because wage contracts are used to share the risk in an optimal way, under the constraint that the employee is always free to quit. An alternative explanation of wage rigidity relies on incomplete contracting (as in MacLeod and Malcomson [1993]). Also, downward rigidity is a standard finding of the empirical literature on wage setting (see, e.g., the recent study by Bewley [1997]).

Our test is based on a characteristic property of wage dynamics

---

[1] An important exception is Murphy's (1986) investigation of a longitudinal sample of 1,488 chief executive officers spanning more than 1,000 firms. Murphy's idea is to use the rate of return on common stock as a measure of the executive's performance. In a similar way, Farber and Gibbons (1991) study the correlation through time between wages and proxies of ability.

that we call the *late-beginner effect*. Consider two workers who start from the same wage level at date 0 and receive exactly the same wage at date 2. However, they have followed a different path between dates 0 and 2. Worker A is an early starter, in the sense that her wage at date 1 is high, and B is a late beginner, with a lower salary at date 1. We show that, in the LDR framework, B's future career prospects then are better than A's. In other words, if we control for the wage at date 2, then the wage at date 1 should be *negatively* correlated with future wages (at date 3 and beyond). As this property concerns only the time process followed by wages, it can be tested even if performances have not been recorded.

The rest of the paper is devoted to the construction and estimation of an econometric model aimed at testing this prediction. We estimate the model on a data set in which the careers of a panel of about 1,000 executives of a French state-owned firm are recorded for 15 years.[2]

A characteristic of this firm is that it provides a typical example of an internal labor market (see Doeringer and Piore 1971). This is a useful feature since we believe that LDR models provide especially accurate descriptions of internal labor markets, if only because their oversimplified representation of competition on the (external) labor market does not raise major problems in this case, as it would in a more ''open'' type of organization.[3] However, using data from an internal labor market has some drawbacks. Within an internal labor market, wages are, in general, determined by the hierarchical level and seniority, according to some explicit formula (which has generally been negotiated with unions). Thus promotions should be seen as the appropriate index to appraise the link between compensation and performances, a point already emphasized by Baker et al. (1994*b*) and Bewley (1997). In the empirical part, we shall therefore consider a model in which the dependent variable is qualitative; this raises specific econometric problems.

The paper proceeds as follows. In Section II, we summarize the main features of the model we consider and provide restrictions on compensation dynamics that lead to a testable property of the wage process. A brief description of the data is given in Section III. Some preliminary statistics are computed in Section IV. Section V is de-

[2] We believe that, as argued in Baker, Gibbs, and Holmstrom (1994*a*), for this purpose, longitudinal monographic data relative to a given firm should be preferred to cross-section samples gathering workers from many firms and involving all kinds of hierarchical levels.

[3] In particular, the wage rigidity assumption is much more likely to hold for agents remaining within the same firm than for agents who change jobs (see Bils 1985).

voted to the econometric model, whose estimation results are presented in Section VI. Finally, Section VII considers some alternative explanations for the late-beginner effect.

II.   The Theoretical Model

A.   The Basic Framework

The class of models for which our approach is relevant is characterized by two features. On the one hand, the labor relationship entails learning on the worker's ability. On the other hand, the wage process exhibits a downward rigidity property: the wage at date t must depend on the employer's beliefs but also be bounded below by some function of the wage at date $t - 1$.

These two requirements can be formalized as follows.

1.   Learning

Our representation of the learning process exactly follows that of Harris and Holmstrom (1982) in their seminal paper. Information is incomplete but symmetric between the employer and the employees. To capture this feature, it is assumed that each employee i is characterized by, besides observable variables (education etc.), a (scalar) ability index η^i. At the beginning of the relationship, the worker's ability is unknown *to both parties*. They know only the distribution from which η^i is drawn; we assume that this distribution is normal.

During the relationship, both parties learn (symmetrically) about η^i. Specifically, at each period t, the employee's performance y_t^i is publicly revealed. Performance is a noisy function of the employee's ability.

ASSUMPTION 1. Performance at date t is given by $y_t^i = \eta^i + \epsilon_t^i$, where the ϵ_t^i are independently and identically distributed centered normal random shocks.

Given the current observation, both parties revise their perception of the worker's ability in a Bayesian way; we call m_t^i the posterior beliefs at date t. A basic property of this setting, which will be crucial below, is given by the following easy lemma.

LEMMA 1. Under assumption 1, the posterior distribution of m_{t+1}^i, conditional on $y_1^i, \ldots, y_t^i$, depends only on m_t^i and on the parameters of the model. Moreover, it is increasing in m_t^i in the sense of first-order stochastic dominance.

Proof. Standard results on Bayesian Gaussian learning show that, conditionally on the past, m_{t+1}^i is normally distributed with mean

m_t^i and with a variance that depends only on t. Lemma 1 follows immediately.

2. Wage Formation and Downward Wage Rigidity

The second feature of our framework concerns the process of wage formation. Formally, we make the following assumption.

ASSUMPTION 2. *Downward rigidity*.—At any date t, the wage of employee i is given by

$$w_t^i = \max[\psi_t(w_{t-1}^i), \varphi_t(m_t^i)] + \alpha_t, \qquad (1)$$

where α_t is a random shock and the mappings ψ_t and φ_t are increasing.

A few comments are in order. Note, first, that wages are assumed to actually depend on current beliefs: since workers are always free to quit, the salary received by "good" workers must be high enough to match the potential offers from other employers. If m_t^i is publicly observable, at least partially,[4] these outside offers must be increasing in m_t^i.

The rigidity ingredient, on the other hand, is introduced by the term $\psi_t(w_{t-1}^i)$ in the max operator: even a very poor performance cannot drive the current wage below a given function of the past wage. Such a downward rigidity property is a standard finding of the empirical literature. Labor sociologists, in particular, have provided empirical evidence of this effect. Recently, Bewley (1997), after many interviews with heads and top managers of private firms, has emphasized the feeling of "insult" that agents associate with a decrease in their nominal wages; this argument is put forth by almost all executives interviewed by Bewley to explain, in particular, why wages could not be lowered even during recessions.

Finally, α_t is some noninsurable random shock. One may think of α_t as a firm-specific demand shock or as stemming from some bargaining process with unions. Alternatively, α_t can be interpreted as a general inflationary shock. In particular, the formulation in (1) is compatible with the finding by Baker et al. (1994*b*) and others that, although nominal wages generally are downwardly rigid, real wages need not be in general.

3. Examples

Several models that have been developed in the literature provide illustrations (and theoretical justifications) of our general framework.

---

[4] For a more complex information setting, see, e.g., Ricart i Costa (1988) and de Garidel (1997).

The Harris-Holmstrom model.—A first illustration is provided by the model of insurance and Bayesian learning proposed by Harris and Holmstrom (1982). In this framework, the employer is risk neutral and the worker is risk averse, and it is assumed that no insurance company covers productivity risks. Wage contracts are then used to share the risk in an optimal way, under the constraint that the employee is always free to quit. The learning process is as defined above; and since commitment is unilateral, the expected value of future profits, conditional on perceived ability at that time, must be nonpositive at each period.

The solution of the model displays the required downward rigidity feature. Specifically, Harris and Holmstrom show that the wage received by i at date t has the form

$$w_t^i = \max[w_{t-1}^i, \varphi_t(m_t^i)], \tag{2}$$

where φ_t is a strictly increasing function, the precise form of which depends on the parameters of the model. The intuition is that, since information is symmetric, the worker will be insured not only against the random shocks but also against the risk of being a poor performer (having a low η). Thus, at the optimum, no performance can result in a lowering of wages. Actually, should commitment be bilateral, the contract would provide full insurance, that is, constant wages independently of performances. Here, however, insurance can be only partial because of the no-commitment constraint on the worker's side. At each period, the worker must face a present value of future expected wages at least equal to the expected profit she will generate. The term $\varphi_t(m_t^i)$ represents precisely the market value of the employee, as determined by present beliefs about her ability. Clearly, (2) is a particular case of (1), and Gaussian Bayesian updating satisfies lemma 1. So the Harris-Holmstrom model does fit into our general framework.

Incomplete contracts and efficient investments: avoiding "holdups."—The insurance role of the employment relationship in Harris and Holmstrom is somewhat controversial. On the one hand, Bils (1985) finds that, over the business cycle, "the wages of those who stay at the same job are only slightly procyclical, but the wages of those who change jobs and those who move in and out of the work force are very procyclical" (p. 684)—a result that seems to support the insurance story. On the other hand, insurance suggests that the variance of wages of salaried workers should be much lower than that of self-employed workers, which is not really supported by the evidence (see Lazear and Moore 1984).

In any case, insurance is not the only justification for downward rigidity. In a recent paper, MacLeod and Malcomson (1993) reexam-

ine the holdup problems raised by incomplete contracting in a buyer-seller relationship in which unverifiable investments are required. Their main insight is that the contract cannot generate efficient investment unless the investor receives the full marginal returns on investment, taking into account any subsequent renegotiation. The authors exhibit simple contracts that achieve this in various settings. Most relevant to us is their section 7, in which they consider a multiperiod model with general investments and a cost of switching partners. They show that in this model, what they call a fixed-price contract then induces efficient investments. This contract takes the following form, after renegotiation:

$$p_t = \min[\max(p_{t-1}, p_t^S), p_t^B],$$

where p_t^B (p_t^S) makes the buyer's (seller's) outside option bind (see fig. 1 in MacLeod and Malcomson). In our case, the buyer is the employer, the seller is the worker, and p is the wage. Now assume that the employer's outside option never binds. Then, while MacLeod and Malcomson's approach is much more abstract than ours, the properties of their solution exhibit the type of rigidity that we assume in assumption 2. If, moreover, the worker's outside option is generated by learning and outside offers, then we are back within the LDR model.

B. The Influence of Early Performances

What are the specific properties of the wage paths under assumptions 1 and 2? Some properties are somewhat obvious. For instance, it can readily be shown that wages should exhibit the so-called fast track effect: a good start (i.e., a high wage in the first periods) generally signals a good future career. Thus the wage at date $t + 1$ is always positively correlated with the wage at date t and, as a matter of fact, with the wage at any date $t - s$, $s > 0$. Note, however, that this is true only for *unconditional* distributions: the correlation of the wage at date $t + 1$ with the wage at date, say, $t - 1$, *once we also condition on the wage at date t*, is much more complex and will be discussed in detail below.

A second remark is that, in such a setting, the order in which performances take place influences the final wage. Downward rigidity indeed generates a ''ratchet'' effect: whenever a given level of wage has been reached, future wages will never go below (a function of) this level, whatever future performances may be. Thus it is always better to perform well in the first period than later on (see Chiappori, Salanié, and Valentin [1996] for a formal statement).

C.   *Testable Properties of the Wage Process: The Late-
     Beginner Effect*

We now show that the LDR model implies interesting testable prop-
erties for the optimal wage path. To grasp the intuition, assume that
two workers start at date 0 with identical observable characteristics,
so that they receive the same first-period wage. Then, because of
disparities in performances (due to differences in abilities, luck, or
both), their wages at period 1 differ; say, A's wage is higher ($w_1^A >
w_1^B$). Finally, assume that second-period performances are such that,
at period 2, their wages again are equal. Proposition 1 states that B
then should be expected to have a better future career.

PROPOSITION 1. Assume that A and B are such that $w_0^A = w_0^B$,
$w_1^A > w_1^B$, and $w_2^A = w_2^B$. Then $m_2^A \leq m_2^B$. In particular, worker B's
expected wage at date 3 is not worse than A's:

$$E[w_3^B - w_3^A \,|\, w_0^A = w_0^B, \, w_1^A > w_1^B, \, w_2^A = w_2^B] \geq 0.$$

Proof. See the Appendix.

The intuition can be summarized as follows. Consider B's perfor-
mance profile. We know that B's first performance was not as good
as A's. Also, we know that the *order* of occurrence matters and that
starting with a bad result is a handicap. But B has still been able to
catch up at period 2. This can be the case only if her *average* perfor-
mance over the two periods has been at least as good as—and possi-
bly strictly better than—A's. In the former case, then perceived abili-
ties, and hence expected career profiles, are identical. In the latter
case, A's perceived ability is lower than B's; since future prospects
depend only on wage and perceived ability at period 2, we conclude
that B is strictly better off (on average) than A. The latter case is
illustrated in figure 1.

We call this property the late-beginner effect. To the best of our
knowledge, it has not been derived so far as a consequence of any
of the existing alternative models of wage formation. For instance,
in a repeated moral hazard framework, with free access to financial
markets, monetary cost of effort, and constant absolute risk aver-
sion,[5] theory shows that optimal incentives can be provided through
memoryless contracts; poor performance today should result in
lower wages tomorrow but then should be forgotten in the future
(see Fudenberg, Holmstrom, and Milgrom [1990] and Chiappori et
al. [1994] for a general presentation).

[5] Models in which agents cannot access a financial market (i.e., cannot save) are
not relevant for the present situation. Moreover, free access to financial markets
with general preferences generates very complex contracts if these contracts are to
be renegotiation-proof (in particular, effort and savings may have to be random-
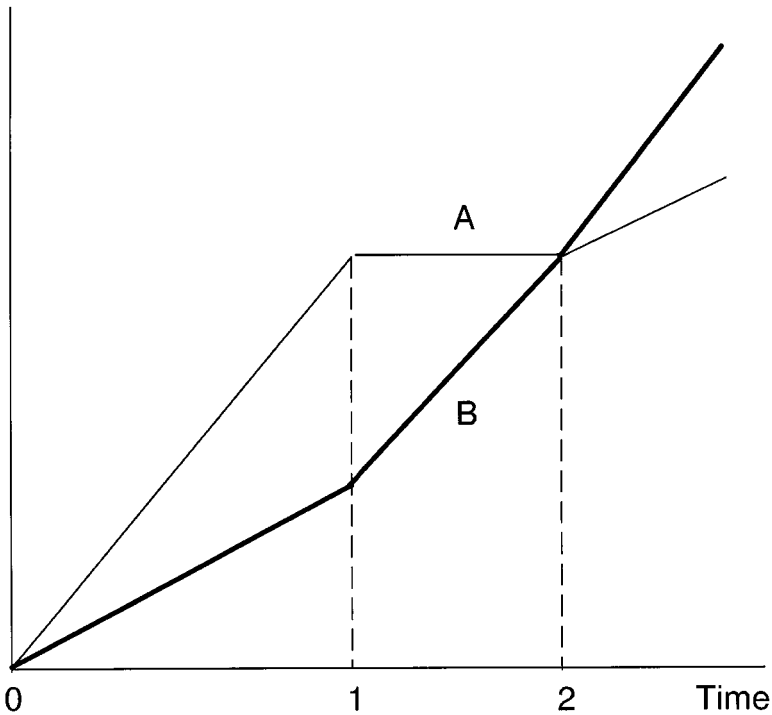ized). See Chiappori et al. (1994) for a detailed discussion.

Wage

FIG. 1.—Late-beginner effect

D. Testing for the Late-Beginner Effect

As proposition 1 characterizes the *intrinsic* dynamics of wages, independently of the underlying performances, it can be tested even in the absence of any data on workers' performances. Write down (expected) wages at date $t + 1$ as a function of observables (age, education, etc.) and of wages at both dates t and $t - 1$. Then the impact of the wage at date $t - 1$ should be found negative. Such a test can be performed on any data set in which wages are observable.

In the real world, however, observed wages do not exactly and exclusively reflect perceived ability (plus possibly last-period wages through rigidity). The wage formation process entails more complex features, such as discrete promotions within a predetermined hierarchical structure.

An important strand of the literature on labor contracts, for instance, emphasizes the role of tournaments in the wage-setting process (see Green and Stokey 1983). In such a context, wage increases

actually take the form of discontinuous jumps, often linked with a promotion to a higher level in the hierarchy. Another argument, provided by Malcomson (1984), relies on the existence of moral hazard on the employer's side: a supervisor can always, ex post, underdeclare the worker's performance in order to minimize his wage increase. Then promotions provide a simple but efficient commitment device: the employer can commit on the number of promotions for a given group, and this is more easily contractible.

This mechanism is especially relevant in our case. Indeed, the existence of a rigid hierarchical structure, where wage increases systematically take the form of promotion to some superior level, is a well-known characteristic of internal labor markets. Though the relevance of the LDR framework is probably general, one may argue that the details of the modelization fit the specific context of internal labor markets especially well. For one thing, other potential employers are considered in that framework only insofar as they provide outside opportunities that the agent may and does use in the bargaining process. While this modeling limitation does not raise specific problems for internal labor markets, it would in the context of an "open" organization.[6] Also, from the empirical point of view, the assumption of downwardly rigid wages is certainly much more likely to hold when agents spend their career within the same organization.

Of course, testing the late-beginner property on internal labor markets raises specific difficulties. The main one is that, in this context, wages do not have the same information content as (implicitly) postulated in the model. They often depend on the hierarchical level and seniority in a rigid way; in our data set, wages are actually determined by an explicit formula based on hierarchical level and seniority within the firm and within the job level. This does not mean that wages are uncorrelated with ability, but rather that the key information is given by the job level, or, in other words, that the dynamics of wages should be modeled as a discrete process within the hierarchical structure. In the remainder of this section, we argue that the late-beginner effect can still be tested with such data, and we describe two ways of performing this task.

---

[6] Ricart i Costa (1988) provides an extension of the Harris-Holmstrom framework to competition between employers. However, this competition typically takes place under some form of adverse selection, which raises the usual problems of the existence of an equilibrium. Also, the dynamics of competition in this framework is particularly complex, as illustrated by de Garidel (1997).

1. Desired versus Actual Wages

A first solution is to recognize that the w_t in the LDR model should be understood as ''desired'' wages, whereas actual wages depend on the hierarchical level, which is observable, according to some mechanism involving other observables (such as seniority). The dynamics of desired wages are characterized by the same equations as in the theoretical model above; in particular, they should, accordingly, exhibit the late-beginner effect. But desired wages are not directly observable from the data; what is observed is the hierarchical level, S_t, that belongs to some finite set $\Sigma = \{S^0, \ldots, S^k\}$.[7] The level S_t can be seen as a discretization of the desired wage in the sense that there exist thresholds $s_t^1, \ldots, s_t^k$ such that

$$S_t = S^j \Leftrightarrow s_t^j \leq w_t < s_t^{j+1}. \tag{3}$$

Now the dynamics of promotions, as recorded in the data, exactly reflects that of desired wages; we can thus try to estimate the latter from the former. Once this goal has been reached, testing the late-beginner effect amounts to testing that

$$\frac{\partial}{\partial w_{t-1}} E[w_{t+1} | w_t, w_{t-1}, \mathbf{X}] < 0,$$

where $\mathbf{X}$ denotes a vector of observables.

2. The Dynamics of Promotions

An alternative approach, which follows from Waldman (1984a, 1984b) and Ricart i Costa (1988), is to directly derive, from the setting above, predictions on the promotion mechanism. Assume again that agents, at any date t, can be assigned a level within some finite set $\Sigma = \{S^0, \ldots, S^k\}$. Suppose, following Gibbons and Waldman (1998), that each level corresponds to some specific task; in addition, tasks are increasingly demanding in the sense that a worker's productivity is highest in the low job when his expected ability is in a low range, highest in the middle job when his expected ability is in a middle range, and so on. Since, within our learning context, ability cannot be directly observed, but is proxied by average performance, the technical translation of this is that there exist thresholds $s_t^1, \ldots, s_t^k$ such that

$$S_t = S^j \Leftrightarrow s_t^j \leq m_t = \frac{y_1 + \cdots + y_t}{t} < s_t^{j+1}.$$

[7] In what follows, an upper index refers to a level in the hierarchy, whereas a lower index refers to time.

Finally, the downward rigidity property can be directly translated in terms of job levels by imposing that $S_t \geq S_{t-1}$ for all t.

It turns out that the late-beginner effect does not always hold in such a context. To understand why, consider again the situation depicted in figure 1, but with a slightly different interpretation. Here, A was promoted to S^1 at the end of period 1, whereas B remained at the initial level S^0. After period 2, however, A is still at S^1, and so is B, who has been able to catch up.

Proposition 1 holds when it is possible to condition on the *exact* value of the wage at period 2. By continuity, it still holds when the wage at period 2 is known with a small enough uncertainty, that is, when the interval $[s_2^1, s_2^2]$, corresponding to level S^1, is narrow enough. This suggests that the late-beginner effect is more likely to be satisfied when (i) the lower threshold s_2^1 is high enough and (ii) the upper threshold s_2^2 is low enough.

This claim can be made more precise (see Chiappori, Salanié, and Valentin [1996] for a formal statement), but the intuition is simple. Assume, first, that all agents are automatically promoted to level 1 at the end of period 2 (this corresponds to $s_2^1 = -\infty$). Then being a late beginner does not convey any information, except for the fact that the first performance has been poor. Consequently, expectations on B will be quite low. On the contrary, if being promoted to level 1 at the end of period 2 is extremely difficult and requires a very good average performance, then being a late beginner is a quite favorable signal. On the other hand, A may have remained at the same level as a result of good (although not excellent) performances; but, alternatively, the reason may entirely be downward rigidity, in spite of a very poor second-period performance. The argument of proposition 1 then applies: B is likely to perform better in the future.

A similar remark also applies to the upper threshold s_2^2. First note that, in our initial model, a better start is always, *unconditionally*, a signal of higher ability. Agents who start fast *and* continue at the same pace (the "champions") are likely to be the most productive overall. On the contrary, the early starters, according to our definition, are those who started well but experienced a slowdown at the next stage. For the empirical applications, it is crucial to be able to distinguish between these two categories. This is precisely the interpretation of the threshold s_2^2: early starters are characterized by the fact that their average performance, at the end of period 2, is not very good; technically, it remains below this threshold. Now, consider, as a third case, the situation in which s_2^2 is very high—say, $+\infty$—so that both champions and early starters are gathered in the same class. Then proposition 1 does not apply, and the late-beginner

effect will not hold. The intuition, again, is clear. Worker A's failure to be promoted at the end of period 2 has absolutely no informational content since promotion was basically out of reach by that time, whatever the performance. The only difference between A and B is that, while they both reached the same level, A did it earlier. Then the odds favor A.

It follows that the late-beginner effect is likely to hold when a promotion to level 2 is frequent enough, so that nonpromotion from level 1 to level 2 signals a relatively weak average performance. Empirically, this will raise problems for the relevant definition of the "periods" under consideration. If period 2 is taken to be very short, early starters are not singled out from champions, which will contradict the late-beginner effect. But, conversely, if it is too long, then basically everybody will have been promoted at level 1 by that time, which also weakens the prediction. We shall discuss this aspect in more detail in the empirical section.

Finally, it should be emphasized that the two approaches just described are not exclusive. On the one hand, from the data on promotions, one can recover the implicit dynamics of desired wages and check for the negative impact of wages at $t - 1$ on wages at date $t + 1$, controlling for wages at date t. On the other hand, the model, once estimated, allows one to compute the promotion probabilities conditionally on observables. These probabilities lead to a test of the kind described just above; in addition, they can be compared to actual frequencies, which provides an interesting specification test. This is the path we follow in the empirical sections below.

E. An Alternative Explanation: Human Capital Accumulation

Before we start the empirical part of the paper, it is worth considering an alternative explanation for the late-beginner effect that relies on human capital accumulation (HCA; see Mincer [1974]). Assume that agents differ by their willingness or ability to accumulate human capital. Some of them will choose a high-accumulation strategy. In a first stage, this results in lower wages since part of their effort is devoted to investment rather than production. But, at the same time, their capital stock increases quickly. After some time, their productivity will be boosted by the returns of their capital; wages will then rise at a much faster pace. Now consider the agents who adopted a low-accumulation profile. Their wage profile will start with a relatively steeper slope, but this slope will remain more or less constant, or at least increase in a much slower way. Eventually, agents of the first category will typically exhibit a late-beginner profile, and the

model predicts that they will end up with higher wages, just as in the LDR framework.

However, it is easy to devise an empirical test to distinguish between the LDR and the HCA explanations of the late-beginner effect. In the HCA case, agents whose career starts fast accumulate little human capital, and this lowers their future wages. On the other hand, we know that in the LDR model, a good start is interpreted as a signal of a high innate ability and hence predicts high future wages (this is the ''fast track'' effect). We shall come back to this point in the empirical section.

III. The Data

We now turn to the empirical part of the paper. The data set we use describes the career profiles of a sample of 1,058 executives in a large French state-owned firm from 1960 to 1982.[8]

This firm presents the typical attributes of an internal labor market. In particular, (1) all executives enter the firm at the same hierarchical level; (2) almost all jobs are filled by internal promotion; (3) quits and layoffs are exceptions; and (4) finally, and more important for our purpose, agents cannot be demoted. This is a provision of the labor contracts that turns out to be strictly observed since in our data no demotion is observed over the 15 years.

A. Wage Setting

Wages are determined as a function of job level and seniority. Specifically, there is an explicit formula, publicly announced, that defines the wage of a given agent as a function of (i) his seniority, (ii) his hierarchical level, and (iii) a general coefficient A_t, identical for all executives, and reflecting the general evolution of purchasing power; in general, A_t is bargained over with the unions. Technically, the internal rules of the firm indicate the following way of computing wages:

$$S_t = K_t^i \cdot A_t. \tag{4}$$

Here, K_t^i is the wage, as expressed in some internal numeraire (called ''points''); then (4) translates the internal numeraire into francs, using the coefficient A_t. The value of K_t^i in turn is given by

$$K_t^i = L^i \cdot (1 + \gamma_t^i) k_t^i,$$

---

[8] For a first investigation on the same data, see Bourguignon and Chiappori (1990).

where L^i is a coefficient specific to the hierarchical level, γ_t^i depends only on the employee's seniority within the firm, and k_t^i depends only on the employee's seniority within his current hierarchical level.

Apparently, this formula was strictly applied. As a matter of fact, the files we received did not indicate the wage levels at all. When we asked for them, the answer was that they could be computed from the formula; we were given a complementary file with the historical values of the coefficient A_t over the whole period.

The conclusion is that in this organization, as in most internal labor markets, wage changes either reflect seniority rules or translate promotion decisions into monetary terms. The key steps of a career, in particular, appear to be the transitions between the different hierarchical levels. Indeed, the analysis of hazard rates by hierarchical level displays no consistent administrative policy. Thus we consider promotion as the relevant reward that conveys the employer's view of the worker's ability.[9]

B. Observables

The data set contains biographical information, such as date of birth and diploma, and the promotion history, as recorded each year between 1960 and 1982. Our sample is restricted to agents entering the firm between 1960 and 1967; as a consequence, we observe the career of each of them for at least 15 years.

Below we focus on two observable characteristics that are correlated with ability at time of entry, namely, initial training and age at entry. As a matter of fact, the executives of the sample form a remarkably homogeneous population. They are predominantly men (98 percent). Race is not recorded (French law forbids that); neither is citizenship or geographic origin. None of these variables is likely to affect promotion for executives in any case. Initial training also is quite homogeneous, in the sense that all 1,058 executives of the sample graduated with a masters of engineering and entered the firm almost immediately. In France, however, not all masters of engineering are equally recognized. A crucial difference, for instance, is made between *Grandes Ecoles*[10] and alternative university programs that deliver an equivalent degree. In addition, there is also an impor-

---

[9] Another point is that, throughout the period, the real wage never declined, in contrast with the findings of Baker et al. (1994*b*). This probably reflects the strong bargaining position of unions in this particular context.

[10] The *Grandes Ecoles* are the most prestigious source of young executives in France. Wage surveys show that their diplomas strongly affect a beginner's value on the labor market.
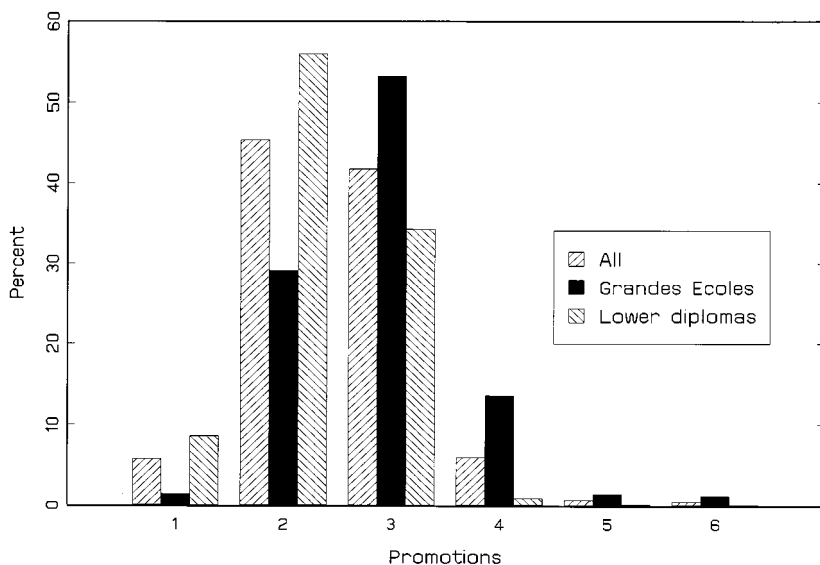
FIG. 2.—Number of promotions by diploma

tant dispersion between the *Grandes Ecoles*. Following most empirical
work on similar data, we divided the sample into two classes: the
best-known *Grandes Ecoles* and the other diplomas. One can observe
this dispersion when looking at the total number of promotions a
worker gets at the end of the first 15 years in the firm. Figure 2
displays the percentage of the sample who obtained from one to six
promotions at the end of the period of observation.

   The other explanatory variable is the age at entry in the firm. This
varies between 21 and 32, with a mode at 26.

C.   *Defining the Periods*

The late-beginner property involves three subperiods, so we shall
need to split the data accordingly. This, of course, is an important
part of the empirical work since it relates to the construction of the
specific data set used to test for the late-beginner effect.

   During their careers, the executives are promoted along a fixed
hierarchical scheme. Concerning the first 15 years, four levels of the
scale are in fact relevant: they are classified as 10, 11, 12, and 13.
Every worker starts at the bottom of level 10 in his first year of em-
ployment. All but a very small minority (less than 3 percent) are at
level 13 or below after 15 years. So the first promotion we consider
(which may or may not occur during period 1) occurs from level

10 to level 11. Among those promoted early to level 11, some (the "champions") are then promoted to level 12 during the next period, whereas others (our "early starters") are not. Also, among the rest, some (our "late beginners") are promoted to level 11 (or even to level 12) during the second period, whereas others are not. Finally, we can consider later promotions (to level 13 or above) during the last period.

The next step is to precisely define the periods. Our raw data set spans the first 15 years of every worker's career. Thus, a priori, there is a lot of leeway to break the 15 years into three periods. The end of the third period will be the end of the last year. But when should the first and second periods end?

A basic guideline is suggested by the conclusions of the theory section. On the one hand, each period must be long enough for promotions to happen "often," so that the downward rigidity phenomenon has time to appear (and, incidentally, that the samples are of sufficient size for the empirical conclusions to be statistically significant). On the other hand, during longer periods, most workers will be promoted, and promotion may lose most of its information content, as discussed above. This remark is especially important in the case of period 2. We want the number of promotions from level 11 to level 12 to be nonnegligible in order to distinguish our early starters from the best performers overall (the champions). There is, however, a price to pay for that: promotions from level 10 to level 11 may then concern a majority of the agents at stake, which somewhat blurs the distinction between our late beginners and the worst performers overall. The nature of this trade-off is essentially empirical; a closer look at the data is needed at this point.

A key empirical element to be considered, hence, is the promotion probability per year at each level. Figure 3 indicates, for each possible tenure, the percentage of the total population at each level. These are descriptive statistics; no correction has been made for truncation. Let us first consider level 10. Less than one-fourth of the sample was promoted before year 3; conversely, almost 90 percent had been promoted at year 5. This strongly suggests that the first period should last 3 or 4 years.

Let us now consider the second period. After 6 years, only 10 percent of the total population has been promoted to level 12; the corresponding agents are almost exclusively champions, that is, agents following the fast tracks. Period 2, hence, should end at least at year 7, and preferably 8 or 9. Going beyond 9, on the other hand, is excluded because the population remaining at level 10 after that date is too small (about 1 percent of the population).

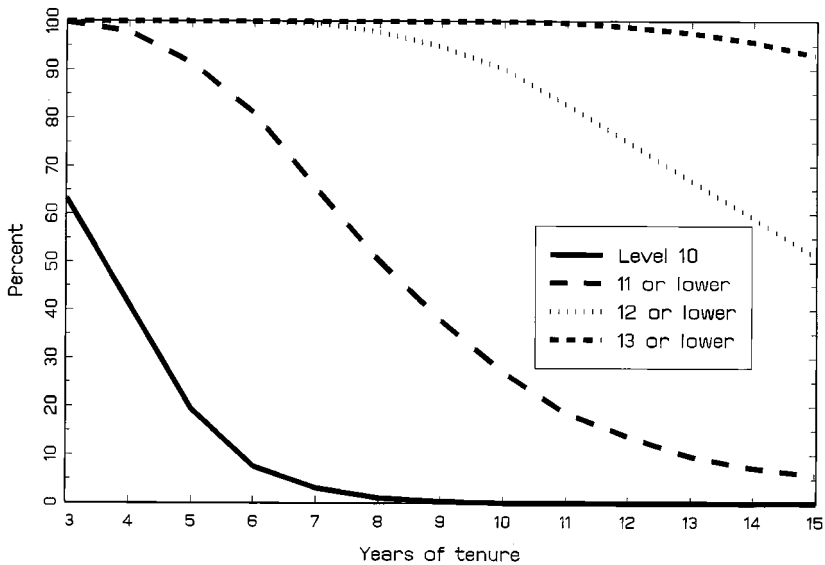The conclusion is that, given the requirements previously de-

FIG. 3.—Empirical probability of promotion by tenure

scribed, period 1 can reasonably be taken to end 3 or 4 years after entry into the firm, whereas for period 2, year 7, 8, or 9 seems acceptable. In the section on descriptive statistics below, we consider the various combinations and indicate their consequences. As for the econometric estimation, on the other hand, we select the intermediate definitions, that is, respectively, 4 years and 8 years.

In each case, we define

$$Y_1 = \begin{cases} 1 & \text{if the worker was promoted to level 11 in} \\ & \text{the first period} \\ 0 & \text{otherwise;} \end{cases}$$

$$Y_2 = \begin{cases} 2 & \text{if the worker was promoted to level 12 during} \\ & \text{the first two periods} \\ 1 & \text{if the worker was promoted to level 11 but not to} \\ & \text{level 12 during the first two periods} \\ 0 & \text{otherwise;} \end{cases}$$

$$Y_3 = \begin{cases} 3 & \text{if the worker was promoted to level 13 or} \\ & \text{above before the end of the 15 years} \\ 2 & \text{otherwise.} \end{cases}$$

Finally, our explanatory variables $\mathbf{X}^i$ refer to age and education. These are dichotomic variables. The variable X_1^i takes the value one

TABLE 1

SAMPLE PERCENTAGE
FREQUENCIES BY CATEGORY

	X_2^i	
X_1^i	0	1
0	29.7	30.7
1	18.2	21.4

when worker i has a diploma from one of the top *Grandes Ecoles* and zero otherwise; X_2^i takes the value one when the worker was recruited before reaching his twenty-sixth birthday and zero otherwise. The cutoff age of 26 was chosen because it splits the sample about evenly. Table 1 shows how the sample splits among the resulting four categories. In order to control for possible changes during the period, we also added a third dummy variable that takes the value one for employees who were hired after 1964 (the midpoint in the sample).

IV. Descriptive Statistics

Given our data, it is easy to test the late-beginner effect using only descriptive statistics. For any given definition of periods, denote p_{ijk} the true probability that a worker has the promotion path ($Y_1 = i$, $Y_2 = j$, $Y_3 = k$) and N_{ijk} the number of such workers in our data set. Then early starters are clearly those for whom $Y_1 = Y_2 = 1$, so that their conditional promotion probability in the third period is p_{113}/p_{11}. On the other hand, there are two possible ways to define the late beginners. The first one consists in selecting those workers for whom $Y_1 = 0$ and $Y_2 = 1$. Then the late-beginner effect translates, in terms of conditional promotion probabilities, into $p_{013}/p_{01} > p_{113}/p_{11}$. Let us call this inequality I1. The conditional probabilities on both sides of the inequality can be estimated by their empirical analogues N_{013}/N_{01} and N_{113}/N_{11}; the standard errors of these estimators follow by assuming a binomial process, and finally we can compute the Student of the difference of the two conditional probabilities if we make the further (natural) assumption that their estimators are independent.

To illustrate this, we use the whole sample of 1,058 executives and consider various possible definitions for the first two periods. In each case, we estimate the ratios N_{013}/N_{01} and N_{113}/N_{11} and the Student of the difference. The results are given in table 2.

When the second period ends 8 or 9 years after entry, the comparison of conditional promotion probabilities is consistent with our pre-

TABLE 2

PROMOTION PROBABILITIES FOR EARLY STARTERS AND LATE BEGINNERS
(with Inequality I1)

DEFINITION OF PERIODS	PROMOTION PROBABILITY		t-TEST FOR THE DIFFERENCE
	Late Beginners	Early Starters	
$t_1 = 3, t_2 = 7$	.31 (.02)	.35 (.04)	−.90
$t_1 = 4, t_2 = 7$	.30 (.03)	.34 (.03)	−.97
$t_1 = 3, t_2 = 8$	.23 (.02)	.17 (.04)	1.18
$t_1 = 4, t_2 = 8$	.22 (.02)	.21 (.03)	.32
$t_1 = 3, t_2 = 9$	.12 (.02)	.07 (.04)	1.10
$t_1 = 4, t_2 = 9$	.11 (.02)	.10 (.03)	.18

diction: late beginners are more likely to reach the top than early starters. It is not the case, however, when the second period is shorter (i.e., 7 years). Seven years may not be long enough for downward rigidity to operate, as argued above. In any case, the differences are not statistically significant.

These results, however, are somewhat biased by the fact that at the end of the second period, some of the late beginners are already at the next level (i.e., level 12; this corresponds to $Y_1 = 0$ and $Y_2 = 2$). These workers are not taken into account in the statistics above. Still, they provide an even better illustration of our late-beginner effect: typically, their ability is high, although they have been especially unlucky during the first period. Table 3 presents the promotion probabilities when this particular subsample is added to the late-beginner population so that the conditional promotion probability of late beginners now is defined as $(p_{013} + p_{023})/(p_{01} + p_{02})$. We call the resulting inequality I2.

TABLE 3

PROMOTION PROBABILITIES FOR EARLY STARTERS AND LATE BEGINNERS
(with Inequality I2)

DEFINITION OF PERIODS	PROMOTION PROBABILITY		t-TEST FOR THE DIFFERENCE
	Late Beginners	Early Starters	
$t_1 = 3, t_2 = 7$	.40 (.02)	.35 (.04)	1.1
$t_1 = 4, t_2 = 7$	.34 (.02)	.34 (.03)	.0
$t_1 = 3, t_2 = 8$	.39 (.02)	.17 (.04)	4.7
$t_1 = 4, t_2 = 8$	.33 (.02)	.21 (.03)	3.0
$t_1 = 3, t_2 = 9$	.38 (.02)	.07 (.04)	7.0
$t_1 = 4, t_2 = 9$	.32 (.02)	.10 (.03)	6.0

Now the differences of probabilities *always* have the predicted sign; in addition, for $t_2 = 8$ and $t_2 = 9$, they are highly significant. This suggests that the predictions of the theoretical model are at least worth more precise consideration. In the next section, we impose more structure on the data-generating process in order to reach sharper results.

V.   The Econometric Model

Our general strategy in this section is the one described at the end of Section II. We first specify a desired wage formation equation that can be estimated from our promotion data. A first test concerns the predicted negative correlation between wages at dates 1 and 3. Then, using the estimated model, we can estimate the conditional promotion probabilities and directly check the late-beginner effect.

The first task, hence, is to specify the equations for the desired wages w_t. We proceed as follows. First, we subtract the effect of observables: $W_t = w_t - \mathbf{X}\delta$, where δ is a set of time-invariant parameters to be estimated. Second, conditionally on observables, the first-period desired wage should depend on the realization of the first-period performance:

$$W_1 = \alpha_1 + u + \epsilon_1, \tag{5}$$

where ϵ_1 is the first period's random shock and u is a random fixed effect.

What we want to test now is a prediction on the expected value of the desired wage at date $t + 1$, conditionally on its values at dates t and $t - 1$. This suggests the following decomposition:

$$
\begin{aligned}
W_2 &= E[W_2 | W_1] + (W_2 - E[W_2 | W_1]), \\
W_3 &= E[W_3 | W_1, W_2] + (W_3 - E[W_3 | W_1, W_2]).
\end{aligned}
\tag{6}
$$

Given the complexity of the estimation process (it will rely on a trivariate probit model), we consider a linearization of both conditional expectations $E[W_2 | W_1]$ and $E[W_3 | W_1, W_2]$. Then (6) takes the form

$$
\begin{aligned}
W_2 &= \alpha_2 + \beta_2 W_1 + \epsilon_2, \\
W_3 &= \alpha_3 + \beta_3 W_2 + \gamma_3 W_1 + \epsilon_3,
\end{aligned}
\tag{7}
$$

where, by construction, ϵ_2 is independent from u and ϵ_1, and ϵ_3 is independent from u, ϵ_1, and ϵ_2.

The prediction to be tested is that $\beta_2 > 0$, $\beta_3 > 0$, and $\gamma_3 < 0$.

Finally, (5) and (6) give the following equations for the desired wages:

$$w_1 = c_1 + \mathbf{X}\delta + u + \epsilon_1,$$
$$w_2 = c_2 + \mathbf{X}\delta + \beta_2(u + \epsilon_1) + \epsilon_2, \tag{8}$$
$$w_3 = c_3 + \mathbf{X}\delta + \beta_3[\beta_2(u + \epsilon_1) + \epsilon_2] + \gamma_3(u + \epsilon_1) + \epsilon_3.$$

Given that we shall use a trivariate probit model, there is little we can do to estimate the variances of the disturbances. Thus we normalize the variances of $u + \epsilon_1$ and $u + \epsilon_2$ to one. This decomposes into $Vu = \rho$ and $V\epsilon_1 = V\epsilon_2 = 1 - \rho$. Note that ρ is a crucial parameter since it represents the variance of the random fixed effect and, thus, what is to be learned about the worker's ability. On the other hand, since the third period is longer than each of the first two periods, we allow for a different variance for ϵ_3: $V\epsilon_3 = v_3(1 - \rho)$ so that both ρ and v_3 are to be estimated.

Finally, we observe only the job level, interpreted as a discretization of the desired wage. Our dependent variables, hence, are defined as

$$Y_1 = \begin{cases} 1 & \text{if } w_1 > s_1^1 \\ 0 & \text{otherwise;} \end{cases}$$

$$Y_2 = \begin{cases} 2 & \text{if } w_2 > s_2^2 \\ 1 & \text{if } s_2^1 < w_2 < s_2^2 \\ 0 & \text{otherwise;} \end{cases} \tag{9}$$

$$Y_3 = \begin{cases} 3 & \text{if } w_3 > s_3^3 \\ 0 & \text{otherwise,} \end{cases}$$

with only one proviso due to downward rigidity: since there are no demotions in the sample, we code $Y_2 = 1$ whenever $Y_1 = 1$ and $w_2 < s_2^1$.

Our model is defined by (8) and (9) plus the requirement that u, ϵ_1, ϵ_2, and ϵ_3 are centered normal and independently distributed, with the variances given above. The constants c_i and α_i cannot be estimated independently of the thresholds, so we shall not try to interpret them. There are eight other parameters to be estimated: three in δ, β_2, β_3, γ_3, ρ, and v_3.

We are now ready to write our trivariate probit model. Note that because of the coefficients β_2, β_3, and γ_3 and the correlation structure they induce among (w_1, w_2, w_3), the model does not decompose across periods: it is formally similar to a probit model on panel data

with serial correlation. The likelihood function for the full sample can be written as

$$L_N = \frac{1}{N} \sum_{i=1}^{N} \log l^i,$$

where

$$l^i = \sum_{j=0,1} \sum_{k=0,1,2} \sum_{l=2,3} \mathbf{1}(Y_1^i = j) \mathbf{1}(Y_2^i = k) \mathbf{1}(Y_3^i = l) p_{jkl}^i$$

and $N = 1{,}058$ is the number of workers in our sample.

The computation of the probabilities p_{jkl}^i is tedious but simple. Take, for instance, the probability p_{002}^i that worker i is never promoted. This is just the probability that

$$u^i + \epsilon_1^i \leq s_1^1 - c_1 - X^i \delta,$$

$$\beta_2(u^i + \epsilon_1^i) + \epsilon_2^i \leq s_2^1 - c_2 - X^i \delta,$$

$$\beta_3[\beta_2(u^i + \epsilon_1^i) + \epsilon_2^i] + \gamma_3(u^i + \epsilon_1^i) + \epsilon_3^i \leq s_3^1 - c_3 - X^i \delta.$$

All left-hand terms in this system of inequalities are centered normal, and it is easy to compute their variance-covariance matrix, so that p_{002}^i is easily expressed using Φ_3, the cumulative distribution function of a trivariate normal vector with zero mean, unit variance, and arbitrary correlations. The function CDFTVN of the GAUSS software is a very accurate (if somewhat slow) implementation of the trivariate normal cumulative distribution function Φ_3.

It is readily checked that the likelihood function depends on exactly eight parameters (the constants are discarded), so that the model is just identified. While the expression for L_N looks forbidding, it can be computed and maximized fairly easily with GAUSS. This yields consistent and asymptotically efficient estimators $\hat{\boldsymbol{\theta}}_N$ of the parameter vector $\boldsymbol{\theta} = (\rho, \beta_2, \beta_3, \gamma_3, \delta, v_3)$. Their asymptotic variance-covariance matrix is consistently estimated by $\hat{I}^{-1}$, where $\hat{I}$ is the outer product of the scores

$$\hat{I} = \frac{1}{N} \sum_{i=1}^{N} \frac{\partial \log l^i}{\partial \boldsymbol{\theta}}(\hat{\boldsymbol{\theta}}_N) \frac{\partial \log l^i}{\partial \boldsymbol{\theta}'}(\hat{\boldsymbol{\theta}}_N).$$

These estimators can be tested for the theoretical restrictions on β_2, β_3, and γ_3; they also will give us immediately the conditional probabilities of promotion at the third period that we are most interested in.

TABLE 4

ESTIMATED PARAMETERS AND THEIR STUDENTS

Parameter	Estimated	Student
ρ (variance of fixed effect)	.46	3.8
$\beta_2 = \partial W_2/\partial W_1$	.67	8.6
$\beta_3 = \partial W_3/\partial W_2$	1.02	8.1
$\gamma = \partial W_3/\partial W_1$	−.30	−3.4
δ_1 (top diploma)	.78	9.6
δ_2 (young at entry)	−.34	−5.1
δ_3 (late entry)	.19	2.9
v_3 (third-period variance)	1.03	4.1

VI.   The Results

Maximizing the log likelihood yielded the estimates[11] contained in table 4. The algorithm converged without difficulty.[12]

First note that a high-level diploma increases the probability of a promotion; as expected, this effect is strongly significant. Age at recruitment also appears to have a positive effect on perceived ability; again, anecdotal evidence seems to confirm this. The value of δ_3 suggests either that the quality of newly hired executives improved slightly after 1964 or that demand factors created better opportunities from that date on. More interesting, the estimated variance of the individual random effect ρ implies that the individual random effect contributes about half the variance of the shock on each period's performance. This confirms that learning about innate abilities should be a central preoccupation of employers.

The key prediction of the LDR model is that, owing to the late-beginner effect, γ_3 should be negative; this is clearly corroborated by the data. Also note that the coefficients β_2 and β_3 have the expected positive sign.

Recall from the descriptive section that we had two ways to test the LDR model on conditional promotion probabilities. The first one is

$$\frac{p_{013}}{p_{01}} > \frac{p_{113}}{p_{11}}, \tag{10}$$

which we called inequality I1. Of course, these probabilities here depend on the individual through his (age, education, year of entry)

---

[11] We do not present estimates for the constants and thresholds since they are not identifiable.

[12] It took about an hour on a Pentium microcomputer. Most of the CPU time is spent inside the trivariate normal cumulative distribution function CDFTVN.

TABLE 5

ESTIMATED CONDITIONAL PROBABILITIES
(with Inequality I1)

X_1	X_2	X_3	Late Beginners (1)	Early Starters (2)	t_Δ (3)
0	0	0	.21	.19	.5
0	0	1	.24	.22	.7
0	1	0	.16	.16	.1
0	1	1	.19	.18	.3
1	0	0	.34	.29	1.3
1	0	1	.37	.32	1.4
1	1	0	.28	.25	.9
1	1	1	.31	.27	1.1

characteristics. Table 5 shows that the hypothesis is indeed con-
firmed for all ages, education levels, and subperiods of entry. We
also computed the estimated standard errors of (p_{013}/p_{01}) −
(p_{113}/p_{11}) using Slutsky's theorem. In each case, the late-beginner
effect is satisfied, although the estimated Student t_Δ in column 3 of
table 5 shows that the difference is never statistically significant at
the conventional 5 percent level.

It turns out that our second way to test the late-beginner effect
(using inequality I2) yields similar but much more clear-cut results.
Recall that in this test we also consider those who were not promoted
in the first period but were promoted *twice* in the second period as
late beginners, so that the conditional promotion probability of late
beginners now is $(p_{013} + p_{023})/(p_{01} + p_{02})$. The results are given in
table 6. Again, the late-beginner effect is confirmed for all values

TABLE 6

ESTIMATED CONDITIONAL PROBABILITIES
(with Inequality I2)

X_1	X_2	X_3	Late Beginners (1)	Early Starters (2)	t_Δ (3)
0	0	0	.31	.19	3.3
0	0	1	.35	.22	3.7
0	1	0	.23	.16	2.5
0	1	1	.27	.18	2.9
1	0	0	.50	.29	4.7
1	0	1	.55	.32	4.8
1	1	0	.41	.25	4.2
1	1	1	.46	.27	4.5

TABLE 7

ACTUAL VS. SIMULATED PROMOTION PATHS

Y_1	Y_2	Y_3	Actual	Simulated (Rounded)
0	0	2	10	10
0	0	3	0	0
0	1	2	265	267
0	1	3	76	74
0	2	2	21	20
0	2	3	64	65
1	1	2	142	140
1	1	3	38	39
1	2	2	103	104
1	2	3	339	338

of the observables; but this time the difference is always statistically significant.

Thus table 5 and especially table 6 systematically back the hypothesis that late beginners are favored in the third period, which we derived as an implication of the LDR model. This clearly suggests that learning about innate abilities is a strongly relevant factor in determining career profiles. It also emphasizes the surprising predictive power of such a simple model as that in LDR.

It is difficult to invent a specification test since no simple and attractive alternative specification comes to mind. However, it is easy to look at the fit of the model by comparing the promotion paths simulated from the estimated models and those found in the actual sample. This is done in table 7. The results show that, simple as it is, our model predicts promotion paths very accurately.

VII. Conclusion

The first goal of the paper was to show how the dynamical consequences of a wage contracting model can lead to testable restrictions on career profiles. We identified a class of models that generate a "late-beginner" property. Our empirical results confirm this prediction, as we find that late beginners are systematically favored in future promotion decisions.

A natural question at this point is whether *other* types of models would also predict a late-beginner type of effect. We shall review a few candidates here and indicate which empirical consideration could be used to distinguish them from the LDR class of models.

Let us first deal briefly with the first two alternative explanations. The first one rests on the famous Peter principle, which holds that

every worker is promoted up to his or her level of incompetence. The trouble with this approach is that it finds it difficult to account for the mere existence of late beginners whose career stagnates and then accelerates. The second explanation holds that recent performance could be a better indicator of future success because future work is more closely related to recent than to past work. This would ignore the fact that, in our framework, early starters and late beginners were not employed at the same task at the beginning of the first period; in fact, early starters were employed at a higher job level.

We now consider in more detail two other possible stories.

A.   Human Capital Accumulation

We first consider the HCA story. As argued in the theory section, there exists a very simple test that allows one to decide between the two explanations. If differences in HCA strategies are the main cause of the observed differences in career profiles, then it should be the case that, *unconditionally*, a slow start leads to better careers than a quick one.

Two types of evidence can be put forth to check this property. For one thing, we can consider the subset of agents who turned out to be the most successful and see what their initial wage profile looked like. Our data clearly favor the LDR interpretation. For instance, among the 517 agents (out of 1,058) who reached the top levels in the third period, only 140 (i.e., 27 percent) were not promoted during the first period (our definition of a "slow start"). This feature appears to be fairly robust: a slow start is, unconditionally, correlated with a lower terminal position. This "fast track" effect is of course very well known in the literature, and there is little need to insist on it.

Going forward this time, we can also test whether a slow start is, *unconditionally*, correlated with a higher terminal position. Again, the answer is clearly negative. Take, for instance, the proportion of executives who eventually made it to job level 13 as a function of their start. Among the best starters (the champions: $Y_1 = 1$ and $Y_2 = 2$), the proportion is above 75 percent, whereas it is only 21 percent for those with a relatively fast start ($Y_1 = 1$ and $Y_2 = 1$); finally, *none* of the executives with a really slow start (the worst performers overall: $Y_1 = 0$ and $Y_2 = 0$) did it. Restricting attention to only two periods does not change this conclusion; for instance, we find that

$$\frac{N_{12}}{N_1} = .71 > .19 = \frac{N_{02}}{N_0},$$

and the difference is highly significant (the Student of the difference is 9.9).

We conclude that the HCA explanation can be rejected. It should be stressed, however, that this conclusion does *not* imply that agents do not accumulate human capital, but simply that differences in the respective rates of human capital accumulation do not seem to provide a convincing explanation of the late-beginner effect.

B.   Autocorrelated Individual Productivity

A second possible explanation relies on specific properties of the dynamics of individual productivity. Assume that wages are directly linked to productivity, say, $w_t^i = \phi(\gamma_t^i)$, where γ_t^i denotes i's productivity at date t and ϕ is increasing. Also, assume that productivity follows some AR(2) process:

$$\gamma_t^i = \alpha\gamma_{t-1}^i + \beta\gamma_{t-2}^i + \nu_t. \tag{11}$$

Assume, finally, that $\beta < 0$. It may be, for instance, that *changes* in productivity are positively correlated across time, say

$$\gamma_t^i - \gamma_{t-1}^i = \lambda(\gamma_{t-1}^i - \gamma_{t-2}^i) + \nu_t,$$

with $\lambda > 0$ (in which case $\alpha = 1 + \lambda$ and $\beta = -\lambda$); but other illustrations could be found.

This, quite obviously, would lead to exactly the kind of effect we get in our model. This suggests a couple of remarks.

A formulation like (11) is both more specific and more ad hoc than our setting. The key point—the negative correlation between γ_t^i and γ_{t-2}^i, with γ_{t-1}^i kept constant—has to be postulated rather than derived from a set of theoretical assumptions. Also, while autocorrelation of productivity shocks may sound like an acceptable assumption, it may be more difficult to understand why productivity *changes* should be positively correlated. Note, incidentally, that the AR(2) structure is needed; an AR(1) process would not produce a late-beginner effect.

In principle, the existence of an AR(2) structure with a negative β is a testable assumption. Unfortunately, in our case, this test is difficult to achieve because of the discrete nature of our data. In any case, it would require more than three periods, which would add further difficulties for the estimation. Data on wages, such as those used by Baker et al. (1994*a*, 1994*b*), would be more adequate. Obviously, further research is needed in this area.

Appendix

Proof of Proposition 1

Since $w_2^B = w_2^A$, we know that

$$\max[\psi_2(w_1^B), \varphi_2(m_2^B)] + \alpha_2 = \max[\psi_2(w_1^A), \varphi_2(m_2^A)] + \alpha_2.$$

A first consequence is that

$$w_2^B = \varphi_2(m_2^B) + \alpha_2 > \psi_2(w_1^B) + \alpha_2. \tag{A1}$$

Indeed, assume that

$$w_2^B = \psi_2(m_1^B) + \alpha_2.$$

Then

$$\psi_2(w_1^B) + \alpha_2 \geq \psi_2(w_1^A) + \alpha_2,$$

which implies $w_1^B \geq w_1^A$, a contradiction. Now

$$w_2^A = \max[\psi_2(w_1^A), \varphi_2(m_2^A)] + \alpha_2 \geq \varphi_2(m_2^A) + \alpha_2. \tag{A2}$$

From (A1) and (A2) it follows that $\varphi_2(m_2^A) \leq \varphi_2(m_2^B)$. Since φ_2 is increasing, this implies

$$m_2^A \leq m_2^B. \tag{A3}$$

Finally, for $i = A, B$, we have

$$E[w_3^i | y_1^i, y_2^i] = E[\max[\psi_3(w_2^i), \varphi_3(m_3^i)] | y_1^i, y_2^i]$$
$$= E[\max[\psi_3(w_2^i), \varphi_3(m_3^i)] | m_2^i]$$

from lemma 1. Then (A3) implies that

$$E[w_3^B | y_1^B, y_2^B] \geq E[w_3^A | y_1^A, y_2^A].$$

Q.E.D.

References

Baker, George P.; Gibbs, Michael; and Holmstrom, Bengt. "The Internal Economics of the Firm: Evidence from Personnel Data." *Q.J.E.* 109 (November 1994): 881–919. (*a*)

———. "The Wage Policy of a Firm." *Q.J.E.* 109 (November 1994): 921–55. (*b*)

Bewley, Truman. "There Is No Harm in Looking: A Close Look at Wage Rigidity." Alfred Marshall lecture presented at the 12th Congress of the European Econ. Assoc. 1997.

Bils, Mark J. "Real Wages over the Business Cycle: Evidence from Panel Data." *J.P.E.* 93 (August 1985): 666–89.

Bourguignon, François, and Chiappori, Pierre-André. "Executives' Promotion in an Internal Labor Market: An Econometric Analysis." In *Essays in Honor of Edmond Malinvaud*, vol. 3, *Empirical Economics*, edited by Paul Champsaur. Cambridge, Mass.: MIT Press, 1990.

Chiappori, Pierre-André; Macho, Ines; Rey, Patrick; and Salanié, Bernard.

''Repeated Moral Hazard: The Role of Memory, Commitment, and the Access to Credit Markets.'' *European Econ. Rev.* 38 (October 1994): 1527–53.

Chiappori, Pierre-André; Salanié, Bernard; and Valentin, Julie. ''Early Starters versus Late Beginners.'' Working paper. Paris: Département et Laboratoire d'Économie Théorique et Appliquée, 1996.

de Garidel, T. ''Pareto Improving Asymmetric Information in a Dynamic Insurance Market.'' Discussion Paper no. 266. London: London School Econ., Financial Management Group, 1997.

Doeringer, Peter B., and Piore, Michael J. *Internal Labor Markets and Manpower Analysis*. Lexington, Mass.: Heath, 1971.

Farber, Henry S., and Gibbons, Robert. ''Learning and Wage Dynamics.'' Working Paper no. 3764. Cambridge, Mass.: NBER, July 1991.

Fudenberg, Drew; Holmstrom, Bengt; and Milgrom, Paul. ''Short-Term Contracts and Long-Term Agency Relationships.'' *J. Econ. Theory* 51 (June 1990): 1–31.

Gibbons, Robert, and Waldman, Michael. ''A Theory of Wage and Promotion Dynamics in Internal Labor Markets.'' Manuscript. Chicago: Univ. Chicago, Grad. School Bus., 1998.

Green, Jerry R., and Stokey, Nancy L. ''A Comparison of Tournaments and Contracts.'' *J.P.E.* 91 (June 1983): 349–64.

Harris, Milton, and Holmstrom, Bengt. ''A Theory of Wage Dynamics.'' *Rev. Econ. Studies* 49 (July 1982): 315–33.

Lazear, Edward P., and Moore, Robert L. ''Incentives, Productivity, and Labor Contracts.'' *Q.J.E.* 99 (May 1984): 275–96.

MacLeod, W. Bentley, and Malcomson, James M. ''Investments, Holdup, and the Form of Market Contracts.'' *A.E.R.* 83 (September 1993): 811–37.

Malcomson, James M. ''Work Incentives, Hierarchy, and Internal Labor Markets.'' *J.P.E.* 92 (June 1984): 486–507.

Mincer, Jacob A. *Schooling, Experience and Earnings*. New York: Columbia Univ. Press (for NBER), 1974.

Murphy, Kevin J. ''Incentives, Learning, and Compensation: A Theoretical and Empirical Investigation of Managerial Labor Contracts.'' *Rand J. Econ.* 17 (Spring 1986): 59–76.

Ricart i Costa, Joan E. ''Managerial Task Assignment and Promotions.'' *Econometrica* 56 (March 1988): 449–66.

Waldman, Michael. ''Job Assignments, Signalling, and Efficiency.'' *Rand J. Econ.* 15 (Summer 1984): 255–67. (*a*)

———. ''Worker Allocation, Hierarchies and the Wage Distribution.'' *Rev. Econ. Studies* 51 (January 1984): 95–109. (*b*)