NBER WORKING PAPERS SERIES

LEARNING AND WAGE DYNAMICS

Henry S. Farber

Robert Gibbons

Working Paper No. 3764

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
July 1991

We thank David Card, Richard Freeman, Lawrence Katz, Kevin Lang, Kevin M. Murphy, Jeffrey Wooldridge, and workshop participants at Berkeley, Boston University, Chicago, Columbia, Cornell, Dartmouth, Maryland, MIT, NBER (the 1990 Summer Institute), NYU, Northwestern, Stony Brook, and San Diego. We thank Gerald Oettinger for able research assistance. Financial support from the DOL (Farber: E-9-J-9-0050), the NSF (Farber: SES-8912664; Gibbons: SES-9010759), and an NBER Olin Fellowship (Gibbons) is also gratefully acknowledged. This paper is part of NBER's research program in Labor Studies. Any opinions expressed are those of the authors and not those of the National Bureau of Economic Research.

NBER Working Paper #3764
July 1991

LEARNING AND WAGE DYNAMICS

ABSTRACT

We develop a dynamic model of learning and wage determination: education may convey initial information about ability, but subsequent performance observations also are informative. Although the role of schooling in the labor market's inference process declines as performance observations accumulate, the estimated effect of schooling on the level of wages is independent of labor-market experience. In addition: time-invariant variables correlated with ability but unobserved by employers are increasingly correlated with wages as experience increases; wage residuals are a martingale; and wage cuts are not rare, even for workers who do not change jobs. We present evidence from the National Longitudinal Survey of Youth that is generally consistent with all four of the model's predictions. We conclude that a blend of the learning model with an on-the-job-training model is more plausible than either model alone.

Professor Henry S. Farber
Princeton University
Industrial Relations Section
Firestone Library
Princeton, NJ 08544-2098
and NBER

Professor Robert Gibbons
Cornell University
Johnson Graduate School
of Management
507 Malott Hall
Ithaca, NY 14850
and NBER

1. Introduction

A worker's education level and other personal characteristics convey only partial information about the worker's productive ability. As the worker accumulates experience in the labor market, further information is revealed. To analyze the effect of such learning on wage determination, we develop a dynamic model in which education may signal initial information about a worker's ability, but employers also learn from subsequent observations of the worker's output. In this learning model, the role of schooling in the labor market's inference process declines as more output observations become available, but the *estimated* effect of schooling on the level of wages is *independent* of labor-market experience. In addition: time-invariant variables correlated with ability but unobserved by employers are increasingly correlated with wages as experience increases; wage residuals are a martingale; and wage cuts are not rare, even for workers who do not change jobs.

We examine this learning model's predictions using data from the National Longitudinal Survey of Youth (NLSY). A crucial feature of these data for our analysis is that they allow us to observe young workers as they begin long-term attachments to the labor market. We use this information to derive a true measure of labor-market experience that differs importantly from the standard definition used in cross-sectional data.

The standard model for analyzing wage dynamics emphasizes human-capital accumulation, including both investment in education and investment in on-the-job training (OJT) (Becker, 1964; Mincer, 1974). As Stigler (1962) observed, however, disentangling the effects of learning and OJT on wage dynamics is "especially difficult" (p. 101). We see our learning model as a complement, not an alternative to the human-capital model. Our theoretical strategy is to derive the implications of the learning model in a framework that abstracts from human-capital investment considerations. Our empirical strategy is to examine data on new labor-force entrants (for whom learning about ability is likely to be most important) to determine whether the patterns we find are consistent with the learning model.

Our learning model is silent about important features of the data, especially the measured returns to experience.[1] Thus, even though we find at least moderate empirical support for all of the predictions of the learning model, it cannot offer a complete explanation. The OJT model, on the other hand, can support the full range of our empirical results only with unusual (and perhaps unreasonable) assumptions regarding the relationship among ability, schooling, and investment in OJT. Thus, while the OJT model is the leading explanation for other features of the data, we believe that our findings are not explained by this model alone. We conclude that a blend of the learning and OJT models offers the best explanation of the data.

2. Theory

In this section we develop a dynamic model of learning and wage determination. In Section 2A we introduce the main ideas by analyzing the simple case in which schooling is the only worker characteristic relevant to wage determination. In Section 2B we enrich the model to include three other kinds of time-invariant worker characteristics: those observed by employers and included in the data, those observed by employers but not included in the data, and those included in the data but not observed by employers. Allowing for these time-invariant worker characteristics yields a very general

[1] In contrast to our learning model, two other learning models---Harris and Holmstrom (1982) and MacDonald (1982)---yield measured returns to experience as their main prediction. Harris and Holmstrom focus on insurance, MacDonald on matching. We ignore both these issues so as to focus on the effect of learning on the estimated coefficients in an earnings regression.

specification of the pure learning model but completely ignores life-cycle changes in worker characteristics, such as productivity growth due to on-the-job training. In Section 2C we allow a limited role for time-varying worker characteristics, in order to give a precise statement of our empirical implementation of the learning model.

A. The Univariate Case

Let η_i and s_i denote the i^{th} worker's productive ability and schooling, respectively. We allow the joint distribution of ability and schooling, $F(\eta_i, s_i)$, to be arbitrary. Let y_{it} denote the output of the i^{th} worker in the worker's t^{th} period in the labor market. We assume that the outputs $\{y_{it}: t = 1, \ldots, T\}$ are independent draws from the conditional distribution $G(y_{it}|\eta_i, s_i)$, but we allow this distribution to be arbitrary. The output y_{it} can also be interpreted as the information the market extracts from output when it knows that the worker can take unobservable actions that influence output, as in Holmstrom (1982).

Information held by employers is symmetric but imperfect: all employers know the joint distribution $F(\eta_i, s_i)$ and the conditional distribution $G(y_{it}|\eta_i, s_i)$, all observe schooling s_i, and all observe the sequence of outputs $\{y_{i1}, \ldots, y_{it}\}$ through period t. The wage paid to a worker in period t equals expected output given all information available at t about the worker (here, schooling and the worker's observed output history):

$$(2.1) \qquad w_{it} = E(y_{it}|s_i, y_{i1}, \ldots, y_{i,t-1}) \ .$$

We can (slightly) relax the assumption that the wage is the conditional expectation of output. Parallel arguments hold if wages are a linear function of this conditional expectation and of output, so firms can earn a profit, and wages can be a base wage plus a piece rate. Also, the conditional expectation can be replaced by the analogous linear projection, as would be the case if employers lacked either the capacity or the information to compute conditional expectations but instead (implicitly) used regressions to predict output and set wages.

This learning model can be interpreted as a dynamic extension of Spence's (1973) static signaling model.[2] By allowing the joint distribution $F(\eta_i, s_i)$ to be arbitrary, we capture any equilibrium (from pooling, through partially revealing, to separating) in a conventional signaling model. We also capture the richer signaling model in which academic ability is imperfectly correlated with productive ability (so that schooling could perfectly reveal academic ability but provide only imperfect information about productive ability). Furthermore, although we interpret η_i as the i^{th} worker's productive ability, it could just as well reflect the quality of the i^{th} worker's schooling, provided that employers cannot observe school quality perfectly and so learn gradually by observing schooling and output, as reflected in (2.1).[3]

The learning model yields several new predictions concerning the estimated coefficients in earnings regressions. For the univariate case

---

[2]While the static signaling model provides a potentially important reinterpretation of the measured return to education, it generates frustratingly few (if any) new testable hypotheses. See Taubman and Wales (1973), Layard and Psacharopoulos (1974), Riley (1979), Albrecht (1981), and especially Lang and Kropp (1986) for discussion of the difficulties in testing the static signaling model.

[3]Given the significant effect of school quality on earnings documented by Card and Krueger (1990), this reinterpretation may be quite important empirically. Interestingly, the static signaling model cannot be reinterpreted in this way (unless the worker's marginal cost of schooling declines with school quality), since there is no opportunity for school quality to affect wages.

analyzed in this section, we derive the first of these new predictions, concerning the estimated coefficient on schooling.

Suppose there exists a panel dataset covering a single cohort of workers all entering the labor market in the same calendar year. Assume the data reveal the schooling of each worker in the cohort and the wage (but not the output) of each worker in each year of the panel ($t = 1, \ldots, T$). Consider using wage data from year t to estimate the earnings regression

$$w_{it} = \alpha_t + \beta_t s_i + \xi_{it} . \tag{2.2}$$

Note that the dependent variable in (2.2) is the level, not the log, of earnings. Because we assume that the wage is a conditional expectation in (2.1), our model yields several predictions about the level of earnings but none directly about the log of earnings.

In regression (2.2), the estimated effect of education on wages for workers in their t^{th} year in the labor market is

$$\hat{\beta}_t = \frac{\operatorname{cov}(s_i, w_{it})}{\operatorname{var}(s_i)} . \tag{2.3}$$

By the law of iterated expectations,[4]

$$E[E(y_{it}|s_i, y_{i1}, \ldots, y_{i,t-1})|s_i] = E(y_{it}|s_i) , \tag{2.4}$$

or

$$w_{it} = E(y_{it}|s_i, y_{i1}, \ldots, y_{i,t-1}) = E(y_{it}|s_i) + \psi_{it} , \tag{2.5}$$

---

[4]The law of iterated expectations states that $E\{E(y|x,z)|x\} = E(y|x)$. See Chung (1974, Chapter 9).

where ψ_{it} satisfies $E(\psi_{it}|s_i) = 0$, and hence $E(s_i\psi_{it}) = E(\psi_{it}) = 0$ and $cov(s_i, \psi_{it}) = 0$. Thus, (2.3) can be rewritten as

$$(2.6) \qquad \hat{\beta}_t = \frac{cov(s_i, E(y_{it}|s_i))}{var(s_i)} .$$

Because a worker's outputs are independent draws from a given conditional distribution, $E(y_{it}|s_i)$ is independent of t, so $\hat{\beta}_t$ is independent of t. That is, the estimated effect of education on the level of wages is independent of experience.

The intuition behind this result is as follows. Our assumption that wages equal expected output implies not only that the first-period wage, w_{i1}, is the expectation of first-period output given schooling, but also that no part of the innovation in wages between the first and second periods, $w_{i2} - w_{i1}$, can be forecasted from the information used to determine w_{i1} (here, schooling). (Equation (2.5) restates this fact for the innovation in wages between the first and t^{th} periods: ψ_{it} is orthogonal to s_i.) Thus, the dependent variable in the second-period regression, w_{i2}, equals the dependent variable in the first-period regression, w_{i1}, plus a term orthogonal to schooling, $w_{i2} - w_{i1}$. Therefore, the estimated coefficient on schooling is the same in the first and second periods.

B. The Multivariate Case

We now substantially generalize the conditions under which the estimated effect of education on wages is independent of experience, and we derive other implications of the learning model. We continue to abstract from life-cycle changes in worker characteristics, such as productivity growth due to on-the-job training, in order to analyze the pure effect of learning in wage determination.

Let X_i denote a vector of time-invariant worker characteristics (other than schooling) that are observable to employers and are included in the data. Let Z_i denote a vector of time-invariant worker characteristics that are observable to all employers but are *not* included in the data. Note that Z_i differs from η_i because employers observe the former but must learn about the latter; Z_i might include the worker's grade-point average, for example. Finally, let B_i denote a vector of time-invariant background variables that are included in the data but are *not* observable to employers, such as whether there was a library card in the household when the worker was age fourteen.

As in Section 2A, the joint distribution of worker characteristics, $F(\eta_i, s_i, X_i, Z_i, B_i)$, is arbitrary, and the outputs $\{y_{it}: t = 1, \ldots, T\}$ are independent draws from the conditional distribution $G(y_{it}|\eta_i, s_i, X_i, Z_i)$, which also is arbitrary. To distinguish the background variables from the worker's ability, we assume that B_i has no direct effect on output: the conditional distribution $G(y_{it}|\eta_i, s_i, X_i, Z_i, B_i) = G(y_{it}|\eta_i, s_i, X_i, Z_i)$ for every B_i. The assumption that the outputs are conditionally independent is convenient but can be relaxed. What is important here is that the outputs have identical conditional distributions. For some of our later results we also require that the outputs not be perfectly correlated.

Since employers observe X_i and Z_i, the wage-determination equation (2.1) now becomes

$$(2.7) \qquad w_{it} = E(y_{it}|s_i, X_i, Z_i, y_{i1}, \ldots, y_{i,t-1}) ,$$

and given data on the vector X_i as well as on schooling, the regression

(2.2) now becomes

$$w_{it} = \alpha_t + \beta_t s_i + X_i \gamma_t + \xi_{it} \quad (2.8)$$

The estimated coefficients $(\hat{\alpha}_t, \hat{\beta}_t, \hat{\gamma}_t)$ from this regression are the coefficients from the linear projection[5] of w_{it} on s_i and X_i, denoted $E^*(w_{it}|s_i, X_i)$:

$$E^*(w_{it}|s_i, X_i) = \hat{\alpha}_t + \hat{\beta}_t s_i + X_i \hat{\gamma}_t \quad (2.9)$$

The linear projection obeys the analog of the law of iterated expectations: $E^*\{E^*(y|x,z)|x\} = E^*(y|x)$. Furthermore, the linear projection of the conditional expectation is the projection itself: $E^*\{E(y|x)|x\} = E^*(y|x)$.[6] These two results imply the following analog of the laws of iterated expectations and projections: $E^*\{E(y|x,z)|x\} = E^*(y|x)$.[7] Because the wage equals expected output, applying the latter result to our model yields

$$E^*(w_{it}|s_i, X_i) = E^*(y_{it}|s_i, X_i) \quad (2.10)$$

---

[5]The linear projection $E^*(y|x)$ is the minimum-mean-square-error linear predictor. The conditional expectation $E(y|x)$, in contrast, is the minimum-mean-square-error predictor. If the conditional expectation is linear then it is equivalent to the linear projection.

[6]Define $\varepsilon = y - E(y|x)$. Then $E(\varepsilon|x) = 0$, so $E^*(\varepsilon|x) = 0$ because the linear projection equals the conditional expectation when the latter is linear. Thus, $0 = E^*\{y - E(y|x)|x\} = E^*(y|x) - E^*\{E(y|x)|x\}$.

[7]By the law of iterated projections, $E^*\{E(y|x,z)|x\} = E^*\{E^*[E(y|x,z)|x,z]|x\}$. Because the projection of the expectation is the projection itself, $E^*[E(y|x,z)|x,z] = E^*(y|x,z)$, so the result follows from the law of iterated projections.

But $E^*(y_{it}|s_i, X_i)$ is independent of t because the outputs are identically distributed, so the effect of schooling on wages is independent of experience. We summarize this argument with the following Proposition.

Proposition: If the outputs $\{y_{it}: t = 1, \ldots, T\}$ are conditionally independent draws from the distribution $G(y_{it}|\eta_i, s_i, X_i, Z_i)$ and the wage at t is given by $w_{it} = E(y_{it}|s_i, X_i, Z_i, y_{i1}, \ldots, y_{i,t-1})$ then the estimated effect of schooling on wages is independent of t in the sequence of regressions $w_{it} = \alpha_t + \beta_t s_i + X_i\gamma_t + \xi_{it}$ for $t = 1, \ldots, T$.

We now derive three additional results: (1) given a mild regularity condition, time-invariant worker characteristics that are unobserved by employers become increasingly correlated with wages as experience increases; (2) wage residuals are a martingale; and (3) wage cuts are not rare, even for workers who do not change jobs.

(1) *Unobserved Characteristics*: Recall that B_i is a vector of background variables that are included in the data but are unobservable to employers. The distribution of worker characteristics, $F(\eta_i, s_i, X_i, Z_i, B_i)$ is arbitrary, however, so the other variables observable to employers (s_i, X_i, and Z_i) could be correlated with B_i. To create a vector of variables that is orthogonal to employers' information when the worker enters the labor market, define B_i^* to be the residual from a regression of B_i on all the other variables included in the data, namely s_i and X_i, *and on* the worker's initial wage, w_{i1}:

$$(2.11) \qquad B_i^* = B_i - E^*(B_i|s_i, X_i, w_{i1}) .$$

Regressing B_i on the worker's initial wage purges B_i^* of the correlation between Z_i and B_i (provided that there is no measurement error in the observed initial wage).

Now add B_i^* as a regressor in (2.8):

$$(2.12) \qquad w_{it} = \alpha_t + \beta_t s_i + X_i\gamma_t + B_i^*\pi_t + \xi_{it} .$$

We are interested in how the estimated coefficients $\hat{\pi}_t$ vary with experience. For ease of exposition, take B_i to be a scalar. Since B_i^* is orthogonal to the other regressors in (2.12), the estimated coefficient $\hat{\pi}_t$ is given by

$$(2.13) \qquad \hat{\pi}_t = \frac{\text{cov}(B_i^*, w_{it})}{\text{var}(B_i^*)} .$$

To solve for $\text{cov}(B_i^*, w_{it})$, note that

$$\begin{aligned}(2.14) \qquad w_{it} &= E(y_{it}|s_i, X_i, Z_i, y_{i1}, \ldots, y_{i,t-1}) \\ &= E(y_{it}|s_i, X_i, Z_i, y_{i1}, \ldots, y_{i,t-2}) + \zeta_{it} \\ &= E(y_{i,t-1}|s_i, X_i, Z_i, y_{i1}, \ldots, y_{i,t-2}) + \zeta_{it} \\ &= w_{i,t-1} + \zeta_{it} \\ &= w_{i1} + \textstyle\sum_{\tau=2}^{t} \zeta_{i\tau} .\end{aligned}$$

Since B_i^* is orthogonal to w_{i1} by construction, we have $\hat{\pi}_1 = 0$ and

$$(2.15) \qquad \text{cov}(B_i^*, w_{it}) = \text{cov}(B_i^*, \textstyle\sum_{\tau=2}^{t} \zeta_{i\tau})$$

$$= \textstyle\sum_{\tau=2}^{t} \text{cov}(B_i^*, \zeta_{i\tau})$$

for $t > 1$. For many commonly encountered specifications of the distributions $F(\eta_i, s_i, X_i, Z_i, B_i)$ and $G(y_{it}|\eta_i, s_i, X_i, Z_i)$, $\text{cov}(B_i^*, \zeta_{i\tau})$ is positive for every τ. Given this regularity condition, $\hat{\pi}_t$ increases with t. Stated less formally, if B_i^* is correlated with ability then the estimated effect of B_i^* on wages should increase with experience, because wages progressively incorporate output signals and output is correlated with ability.

Compare the effect of worker characteristics the market cannot observe (B_i^*) to the effect of characteristics the market can observe (s_i and X_i). By definition, the former play no role in the market's wage-determination equation, but their estimated effect increases as the market learns ability by observing output. The latter, in contrast, play a declining role in the market's inference process, but have a constant estimated effect.

(2) *Wage Residuals*, and (3) *Wage Cuts*: In the pure learning model developed above, the assumptions that worker characteristics are time-invariant and that outputs have identical conditional distributions rule out productivity growth with experience, and the assumption that wages are determined by the conditional expectation of output rules out non-productivity explanations for measured wage growth with experience (such as the insurance model of Harris and Holmstrom). We address time-varying worker characteristics and the return to experience in Section 2C. For now, however, we continue to focus on the pure learning model.

Because $E(\zeta_{it}|w_{i,t-1}) = 0$ in (2.14), wages are a martingale:[8] $E(w_{it}|w_{i,t-1}) = w_{i,t-1}$. In the data, however, measured wage growth with experience implies that wages are not a martingale, so in our empirical work we focus on wage residuals. Since we have not yet introduced time-varying worker characteristics, we cannot yet define the wage residuals we will later use in our empirical analysis. For purposes of illustration, however, consider the residual from (2.8),

$$(2.16) \qquad \hat{\xi}_{it} = w_{it} - (\hat{\alpha}_t + \hat{\beta}_t s_i + X_i \hat{\gamma}_t)$$

$$= E(y_{it}|s_i, X_i, Z_i, y_{i1}, \ldots, y_{i,t-1}) - E^*(y_{it}|s_i, X_i) .$$

We have $E(\hat{\xi}_{it}|\hat{\xi}_{i,t-1}) = \hat{\xi}_{i,t-1}$, because $E^*(y_{it}|s_i, X_i)$ is independent of t.

Turning to wage cuts, (2.14) implies $E(w_{it} - w_{i,t-1}) = 0$, so

$$(2.17) \qquad 0 = \text{Prob}\{w_{it} - w_{i,t-1} < 0\} \cdot E\{w_{it} - w_{i,t-1}|w_{it} - w_{i,t-1} < 0\}$$

$$+ \text{Prob}\{w_{it} - w_{i,t-1} \geq 0\} \cdot E\{w_{it} - w_{i,t-1}|w_{it} - w_{i,t-1} \geq 0\}.$$

Thus, in the absence of wage growth, the product of the frequency and the conditional expectation of wage cuts is equal (in absolute value) to the analogous product for wage increases. For plausible distributions of wage innovations, therefore, real wage cuts will not be rare. In the presence of wage growth, both the frequency and conditional expectation of wage cuts will

---

[8]A martingale is a generalization of a random walk: in the latter, innovations are independently and identically distributed; in the former, independence suffices. In the learning model, the variance of innovations in the conditional expectation is not constant over time, so the conditional expectation is a martingale but not a random walk.

be closer to zero.

C. *Time-Varying Worker Characteristics*

We have so far restricted attention to time-invariant worker characteristics. We now enrich the learning model to allow for productivity growth with labor-market experience, for two reasons: to give a precise statement of the empirical model we will analyze in Section 3, and to assess whether the learning model's predictions continue to hold in the presence of a variety of human-capital considerations. The arguments involved in the latter assessment also allow us to determine whether the learning model's predictions can be derived from a pure human-capital model, in the absence of learning.

In the empirical model we analyze in Section 3, we assume that the i^{th} worker's output in period t is Y_{it}, where

$$Y_{it} = y_{it} + h(t) \ . \tag{2.18}$$

We continue to assume that $\{y_{it}: t = 1, \ldots, T\}$ are independent draws from the conditional distribution $G(y_{it}|\eta_i, s_i, X_i, Z_i)$, exactly as in Section 2B. We now also assume that productivity grows with labor-market experience according to h(t), due to on-the-job training or learning by doing. (Since X_i denotes a vector of time-invariant worker characteristics, we use t to measure labor-market experience.) For simplicity, we take h(t) to be deterministic, but the same conclusions would hold if h(t) also included a (serially uncorrelated) noise term that is independent of η_i, s_i, X_i, and Z_i.

Given (2.18), the wage-determination equation (2.7) now becomes

$$w_{it} = E(Y_{it}|s_i, X_i, Z_i, Y_{i1}, \ldots, Y_{i,t-1}) \ . \tag{2.19}$$

and in the T regressions by experience level in (2.8), the estimated intercepts $\{\hat{\alpha}_t\colon t = 1, \ldots, T\}$ now increase according to h(t). These T regressions can be collapsed into the single pooled regression

$$(2.20) \qquad w_{it} = \alpha_0 + \alpha_1 t + \beta_0 s_i + \beta_1 s_i t + X_i \gamma + \xi_{it} ,$$

where for expositional convenience we take h(t) to be linear.

Our empirical implementation of the learning model is based on (2.18), (2.19), and (2.20). In this model, the first three predictions from Section 2B continue to hold, and the fourth holds in a modified form: The effect of schooling on the level of the level of wages is independent of experience (i.e., $\hat{\beta}_1 = 0$). Time-invariant variables correlated with ability but unobserved by employers are increasingly correlated with wages as experience increases (i.e., given the regularity condition from Section 2B, including the interaction of B_i^* and experience as a regressor in (2.20) yields a positive coefficient). Wage residuals (i.e., $\hat{\xi}_{it}$ from (2.20), analogous to $\hat{\xi}_{it}$ from (2.8), as computed in (2.16)) are a martingale. Finally, negative real wage changes will occur less frequently than would be the case in the absence of wage growth with experience.

We now assess whether the predictions of the learning model continue to hold in the presence of human-capital considerations beyond those captured by (2.18). We consider two possibilities: (1) productivity growth is a function of experience and schooling, $h(t,s_i)$; (2) productivity growth is a function of experience and ability, $h(t,\eta_i)$. Suppose first that (2.18) is replaced by

$$(2.21) \qquad Y_{it} = y_{it} + h(t,s_i) ,$$

where the cross-partial derivative of $h(t,s_i)$ is not zero. (This derivative is positive if having more schooling makes investments in on-the-job training more productive.) In this case, $\hat{\beta}_1$ in (2.20) will reflect the cross-partial derivative of $h(t,s_i)$, and so will not equal zero. The second and third predictions remain unchanged, and the fourth remains as modified above.

Now suppose instead that (2.18) is replaced by

$$(2.22) \qquad Y_{it} = y_{it} + h(t,\eta_i) \ ,$$

where the cross-partial derivative of $h(t,\eta_i)$ is not zero. (This derivative is positive if having more ability makes investments in on-the-job training more productive.) In this case, $\hat{\beta}_1$ will reflect both the cross-partial derivative of $h(t,\eta_i)$ and the correlation between η_i and s_i, as determined from the distribution $F(\eta_i,\ s_i,\ X_i,\ Z_i,\ B_i)$. Thus, $\hat{\beta}_1$ could be zero if these two effects are of opposite sign, but this seems unlikely since both are correlations between ability and investments in human capital (the first on the job, the second in school). Turning to the second prediction, the coefficient on the interaction of B_i^* and experience will reflect both learning (as described in Section 2B) and the cross-partial derivative of $h(t,\eta_i)$, and so could be negative if the latter is sufficiently negative. Finally, wage residuals are no longer a martingale, but may have an equally distinctive covariance matrix (such as occurs when $h(t,\eta_i) = h(t)\cdot\eta_i$, for example).

In summary, for each of the four predictions derived in Section 2B, we have shown that the prediction ceases to hold if certain kinds of human-capital considerations are introduced. To the extent that the data are consistent with the four predictions of the learning model, we can therefore conclude that these particular human-capital considerations are not important in the data (but by no means that all human-capital considerations are unimportant).

To conclude this section, we consider whether a pure human-capital model (i.e., a model without learning) can produce the learning model's four predictions. The first prediction, that the return to education does not vary with labor-market experience, is consistent with an OJT model in which investment in OJT is uncorrelated with education (and with ability, if ability is correlated with education), so that higher education yields a parallel shift in the experience-earnings profile. In contrast, the second prediction, that the return to market-unobserved measures of skill is increasing with experience, is consistent with an OJT model in which investment is positively correlated with these measures. Thus, a pure OJT explanation of these two predictions requires the unlikely condition that worker heterogeneity related to investment in training must be independent of heterogeneity related to education.

The third prediction of the learning model, that wage residuals are a martingale, is consistent with the existence of time-varying worker characteristics that are not included in the data (such as health status), but does not follow naturally from an OJT model. Similarly, our last finding, that wage declines are relatively common, even for workers who do not change jobs, is consistent with the existence of productivity shocks, but is not a feature of the standard OJT model.[9]

---

[9]The standard life-cycle model of human capital accumulation (Ben Porath, 1967) predicts that investment net of depreciation will almost certainly be positive for workers early in their working life, and that earnings net of new investment will be growing. If we enrich this model to allow workers to delay their investment in human capital when they first enter the labor market until they are sure that they are committed to working in the longer

3. Empirical Analysis

In this section we first describe the data from the NLSY that we use for most of our empirical analyses. Next, we examine how the returns to both education and variables correlated with ability but not directly observed by the market (such as aptitude-test scores) vary with labor-market experience in the NLSY. Third, we examine the extent to which the covariance structure of wage residuals is consistent with wage residuals being a martingale. Finally, we investigate the frequency of negative real wage changes in the NLSY. The results are generally consistent with the learning model developed in Section 2.

A. Data

The National Longitudinal Survey of Youth (NLSY) has a number of advantages for our analysis of learning. First, learning about worker quality is likely to be most important early in a worker's career, and the NLSY is focused on precisely this part of the life cycle. Second, the NLSY allows us to use longitudinal information to determine relatively precisely when workers make their first long-term transition to the labor force. Most cross-section data sets (e.g., the Current Population Survey) must use an arbitrary definition of labor market experience (typically age-education-6).[10].

---

run, then workers will earn more at the start since they are not giving up any productivity to invest in skills. Later, those who recognize their commitment to the labor force will begin to invest, so their real wages will then decline.

[10]Inspection of Table 1 indicates that the average difference between age and education varies with entry cohort, ranging from 7.4 years in 1979 to 9.2 years in 1984. Our measure of experience is different from the usual definition of potential experience (age-education-6) to a surprising degree. The simple correlation between actual age at entry and years of education for our sample of 3630 workers is only .53. If potential experience were the

Other longitudinal data sets are not as well suited to our analysis as the NLSY: the Panel Study of Income Dynamics does not have as many young people just starting their working life, and the earlier National Longitudinal Surveys of Young Men and Young Women do not contain work histories as consistent and detailed as those in the NLSY.

Individuals in the NLSY were between the ages of fourteen and twenty-one on January 1, 1979. We eliminate from our analysis the 1280 workers in the military sample. The remaining sample is comprised of 11406 workers, including 6111 workers from a representative cross-section sample and 5295 individuals from a supplemental sample of under-represented minorities and economically disadvantaged workers.[11] At the time we carried out our analysis, there were data available for the 1979 through 1987 interview years.

In order to focus on the learning process from the time workers first make a primary commitment to the labor market, we limit our sample to individuals who make their first long-term transition from non-work to work during the sample period. We define a long term-transition to occur when an individual spends three consecutive years (i.e., intervals between interviews) primarily working after at least a year spent not primarily working. An individual is classified (by us) as primarily working if he/she worked in at least half of the weeks since the last interview and averaged at least thirty hours per week in the working weeks.[12] Only individuals aged 16

---

correct measure, this correlation would be one.

[11]We have carried out our empirical analyses using workers from the representative cross-section alone. The results are qualitatively identical to those reported below.

[12]At the 1979 interview date, the last interview was assumed to be January 1, 1978.

or older were asked the relevant questions on employment history. Thus, we could not classify the youngest cohorts (aged 14 and 15 in 1979) in the earliest years of the survey.

There are 2856 individuals whom we classify as primarily working at the first interview for which there is valid data to classify them. We dropped these individuals from the analysis because we could not determine whether the first observation for these workers was their first year primarily working. On this basis, the first year individuals could make their first long-term transition to the labor force was between the 1979 and 1980 interviews. This sample-construction procedure implies that the lower educational categories are under-represented among the older workers in the sample. For example, the high-school graduates in the oldest cohort (aged 21 in 1979) could have been working from one to three years since graduation (and so would be dropped from our sample), while college graduates in the same birth cohort would just be entering the labor market in 1979 and 1980.

Individuals were dropped from the sample if they were not classified as primarily working in three consecutive interviews. On this basis, the last year individuals could make their first long-term transition to the labor force was between the 1984 and 1985 interviews. (Because our data end in 1987, we cannot be sure that workers who enter after 1984 were primarily working for three years.) We dropped 4669 individuals who never made a long-term transition to the labor force by this definition, as well as ten individuals who were self-employed at all interviews after entry. Finally, we dropped three individuals who were not born in the 1957 through 1964 period (they were born in 1965) and 508 individuals with missing data on key variables.

The final sample consists of 3630 individuals who made their initial long-term transition to the labor force (by our definition) between 1979 and

1984.[13] Table 1 summarizes average age and average education by year of labor-market entry. The earlier cohorts are younger and less well educated. Females are slightly older and better educated upon entry.

Our definition of a worker's initial long-term transition to the labor force is arbitrary. Redefining our criteria with regard to minimum weekly hours or minimum weeks worked had very little effect on the final sample size. Changing the three-year consecutive history requirement had a predictably larger effect on the final sample size. Some information is available to evaluate how sharply we have defined the transition into the labor force. Only thirteen individuals out of the 3630 in the final sample reported three consecutive years of experience from 1975 through 1977. More workers were classified as primarily working for some years prior to their first long-term transition: 2901 were never classified as primarily working prior to their first long-term transition, but 477 were primarily working for one year, 221 for two years, and 31 for three or more years. Overall, our rule captures what seems to be a reasonably sharp transition from not working to working.[14]

While our procedure precisely determines the year of the first long-term transition to the labor force, it does not determine when within the year the transition occurred. We need to make some imputation of this in

---

[13]To recapitulate, there are 11406 individuals in the basic sample of the NLSY. Of these, 2586 were primarily working in the first year we could classify them, 4669 others did not have three consecutive years primarily working, 10 were self-employed in each year, 508 had missing data, and 3 were born outside the 1957-1964 period. This leaves 3630 individuals in the final sample.

[14]The analyses reported below were repeated using only the 2901 workers who were never classified as primarily working prior to their first long-term transition. The results were qualitatively identical to those derived using all 3630 workers.

order to select the first post-transition wage observation appropriately. Assume (for clarity) that interviews are exactly one year apart and that workers can enter at any date between interviews. In this case, an individual who is first classified as primarily working for the year between interviews t and t+1 must have made his/her transition in the interval from (t-.5) to (t+.5). For such an individual, we define experience to be zero at date t. If the individual made his/her transition between t and (t+.5) then there is of course no zero-experience post-transition wage observation available. Our first wage observation for the individual is then from date t+1, when the individual has one year of experience. If the individual made his/her transition between (t-.5) and t, however, then a zero-experience post-transition wage observation should be available (on either the current or most recent job). We use this wage observation only if the individual is working at date t in a full-time job (at least 35 hours per week). We drop zero-experience wage observations for workers who either are not currently on a job at t or who are working part-time (less than 35 hours per week) at t, because the jobs to which these wages correspond are likely to be prior to the long-term transition we are attempting to discern.[15]

Table 2 contains summary statistics for key variables broken down by years of labor-market experience for the sample of workers with complete data on the variables required for the analysis of earnings. Our sample contains

[15]We use this procedure in an attempt to eliminate from the sample wage observations from summer jobs or part-time jobs held while an individual is attending school. (Suggestive evidence that the part-time jobs held at t are likely to be prior to the transition is that 63 percent of jobs held at t are part-time while less than 15 percent of jobs held at any later interview date are part-time.) Of the 3630 individuals in our sample, only 776 have valid zero-experience wage observations (by our definition). There are 531 observations for workers who are not currently with a job at t and 1310 observations for individuals working part-time at t. A further 1013 individuals either have no reported wage at t or are missing data for other variables at t.

18521 valid observations for the 3630 workers. If there were complete data for each of the workers for every year since entry, our sample would have 23818 observations. Of the 5297 observations lost, 2854 are missing at zero years of experience (for various reasons, as noted above). The remainder were lost due to missing data or to workers being self-employed or without a job. Note that part-time observations are used in the analysis except at zero years of experience.

The statistics in Table 2 illustrate a number of features of our data. Workers with more experience are older on average, but age does not generally increase one-for-one with experience in our sample. This is because new entrants with more education enter when they are older. As one would expect, the average wage is strongly positively related to experience. The fraction married is also strongly positively related to experience.

While most of the data we require for our analysis are of the sort labor economists generally analyze (e.g., earnings, education, and experience), we also require measures that are not so common. Specifically, we need measures that are correlated with ability but not observed by the market. While it is difficult to think of an individual attribute that we are sure is not even indirectly observed by the market, we attempt to construct such measures using data on aptitude-test scores and family background available in the NLSY.

Let B_i be a variable that reflects some combination of innate abilities, background, experience, and education. Since much of what determines B_i is observable by the market, B_i would not serve well for our purposes, but the residual ($B_i^{\bullet}$) from a regression of B_i on attributes observable by the econometrician (X_i) is more promising:

(3.1) $$B_i^{\bullet} = B_i - X_i\hat{\gamma} .$$

where X_i is a vector of observable attributes for individual i and $\hat{\gamma}$ is the parameter vector from an OLS regression of B_i on X_i.

Although B_i^* is orthogonal to X_i by construction, it is not orthogonal to attributes observed by the market but not observed by the econometrician (Z_i). Under the hypotheses of the learning model in Section 2, we can ensure that B_i^* is orthogonal to all that is observed by the market by including the first observed wage of worker i (W_{i0}) as part of the vector of observable characteristics:

$$B_i^* = B_i - X_i\hat{\gamma} - \hat{\delta}W_{i0} , \qquad (3.2)$$

where $\hat{\gamma}$ and $\hat{\delta}$ are the estimated parameters from an OLS regression of B_i on X_i and W_{i0}.[16] In the model in Section 2, the wage at any date incorporates *all* the information the market has at that time about the worker's ability: anything orthogonal to that wage is unobserved by the market. However, if the initial wage is measured with error or influenced in other ways so that it does not accurately reflect expected output, then the B_i^* we compute in (3.2) is only partially purged of its correlation with attributes observed by the market but unobserved by the econometrician.

The NLSY has a number of background measures that can be used as measures of B^* for this purpose, including 1) results of the Armed Forces Vocational Aptitude Battery of tests administered to all respondents in the

---

[16]The approach used in this section differs from but was inspired by Hause (1972), who finds using the Thorndike data that the residual from a regression of 1969 earnings on 1955 earnings is correlated with a 1943 test score. The fact that the 1943 test score is not fully incorporated into earnings by 1955 can be interpreted as evidence that learning about ability goes on for quite a long time.

NLSY in 1980, 2) measures of parental educational and occupational attainment, and 3) measures of household intellectual environment, including whether there were newspapers, magazines, and/or library cards in the household. We focus (somewhat arbitrarily) on two measures. First, we construct the Armed Forces Qualifying Test (AFQT) score for each worker.[17] Second, we use information on whether anyone in the home had a library card when the individual was age fourteen. The AFQT score was selected because it is a widely recognized aptitude-test score that is available for all workers in our sample. The library card measure was selected as representative of the set of household-environment variables.[18] No measure of parental educational or occupational attainment was used because these variables were missing for a significant fraction of the workers in our sample.[19]

In order to generate the residual B_i^* from (3.2), we create a sample with only one observation per individual by taking the first valid observation for each of the 3630 individuals in our sample.[20] To compute the AFQT residual using this sample, we regressed the AFQT score on education, part-time status, the interaction of education and part-time status,

---

[17]The AFQT score is constructed as a linear combination of scores on four sub-tests of the battery of vocational aptitude tests. The AFQT score is the sum of the scores on the word knowledge, arithmetic reasoning, and paragraph comprehension tests plus one-half the score on the numeric operations test.

[18]The three household-environment variables (newspapers, magazines, and library cards at age 14) are, as expected, significantly positively correlated.

[19]For example, mother's education was missing for 173 and father's education was missing for 416 of the 3630 individuals in the sample.

[20]As noted above, only 776 of the 3630 workers have their first valid observation at zero years experience. Of the rest, 2607 are first included at one year, 201 at two years, and the remaining 46 at three or more years.

collective bargaining status, race, sex, marital status, the interaction of sex and marital status, age, calendar year, and the real wage.[21] This regression accounted for 52 percent of the variance in AFQT scores. Using the same sample, we also regressed the library-card indicator variable on the same set of variables. This regression has less explanatory power (R^2 = .09), which is not surprising given that the dependent variable is discrete. The residuals from this pair of regressions serve as our two measures of ability that are not observed by the market.

B. The Returns to Education, and to Other Variables Correlated with Ability.

The goal of the analysis in this section is to test the implications of the learning model that 1) the estimated effect of education on the level of wages does not vary with experience and 2) the estimated effect of variables correlated with ability but not observed by the market increases with experience. The earnings functions estimated in this section are specified in the *level* of the real wage rather than the logarithm because our theoretical model provides clear implications for how education and other variables are related to the wage level.

Our data contain multiple observations for each individual, and our learning model implies that the wage equation errors will be correlated across observations within each individual.[22] This correlation needs to be accounted for in deriving both efficient estimates of the earnings function parameters and appropriate tests of the implications of the learning model.

---

[21]All variables with the exception of the real wage were entered as complete sets of dummy variables for each distinct value of each variable. Education was measured in four categories: <12 years, 12 years, 13-15 years, ≥16 years.

[22]The model further implies that the wage residuals evolve as a martingale. This implication of the model is examined in detail in the next sub-section.

Our approach is 1) to estimate the earnings function by OLS, 2) to use the resulting residuals to derive an unrestricted estimate of the within-worker covariance matrix of residuals, and 3) to use this estimated covariance matrix to compute GLS estimates of the earnings function.

Our data are unbalanced, in the sense that there are different numbers of wage observations for different workers. In order to keep the computation of the GLS estimates tractable, we use subsets of our workers with balanced data (i.e., the same number of cross-section observations per worker).[23] We select three (overlapping) balanced subsamples of our full sample of 3630 workers: 1) one through four years experience (n=2217), 2) one through five years experience (n=1610), and 3) one through six years experience (n=1100).[24] These panels were selected as compromises between having a panel of adequate length (enough variation in experience) and having enough workers (longer panels necessarily have fewer workers). Table 3 contains sample statistics for the three balanced panels.

Table 4 contains estimates of the within-worker covariance matrix of wage residuals for each of the three balanced samples. For each balanced sample, these estimates are based on residuals computed from separate OLS regressions for each year-experience cell of the real wage on dummy variables for age (all values), education category (four values), part-time status, the interaction of the education dummies with part-time status, collective bargaining status, nonwhite, female, marital status, and the interaction of

---

[23]The validity of using balanced samples relies on the assumption that the process that generates the balanced samples is independent of the wage determination process. We briefly investigate this assumption at the end of Section 3C.

[24]All workers in the six-year balanced sample are in the two shorter balanced samples, and all workers in the five-year balanced sample are in the four-year balanced sample.

female and marital status. The estimated covariance of the residuals at experience levels j and k is computed as

(3.3) $$\text{Cov}(\varepsilon_j,\varepsilon_k) = \frac{1}{n}\sum_{i=1}^{n} e_{ji}e_{ki}$$

where e_{ji} and e_{ki} represent the OLS residuals for worker i and n is the number of workers in the balanced sample. The standard errors of the elements of the covariance matrix are computed as

(3.4) $$\text{se}(\text{Cov}(\varepsilon_j,\varepsilon_k)) = \frac{1}{n}\left(\sum_{i=1}^{n}\left(e_{ji}e_{ki} - \text{Cov}(\varepsilon_j,\varepsilon_k)\right)^2\right)^{1/2}$$

Let c denote the unrestricted TxT covariance matrix of residuals whose elements are defined by equation 3.3. Assuming independence of residuals across workers, the (nT)x(nT) covariance matrix of residuals for the entire sample is block diagonal and equal to $C = c \otimes I_n$ where I_n is an nxn identity matrix. The unrestricted estimates of c presented in Table 4 are used to compute C for each balanced sample. These estimates of C are then used to compute GLS estimates of the earnings function for each balanced sample. We defer discussion of the specific form of the covariance matrix to the next sub-section.

Table 5 contains GLS estimates of regressions of the real wage for the three balanced panels. There are two specifications for each panel. Both specifications include a constant, age at entry, and dummy variables for part-time employment, three (out of four) education categories, the interaction of part-time status and the education dummies, year fixed effects, the interaction of year fixed effects with the education dummies, collective bargaining status, race, sex, marital status, and the interaction of sex and marital status. The base educational group is twelve years and the base year is 1987. The specifications in columns (2), (4), and (6) also include the AFQT and Library Card residuals and the interaction of these residuals with experience. This is in order to investigate the relationship between the wage and variables unobserved by the market but correlated with ability. The first observation for each worker is omitted from the analysis of the second specification because the first observed wage was used to compute the AFQT and Library Card residuals. The reported R^2 is computed using the GLS coefficients with the untransformed data.[25]

The results show the usual strong positive relationship of earnings with experience and education in all three panels. Among the education categories, only the ≥16 years category is consistently significantly different from the base group (=12 years). Examining the first specification, without the AFQT and Library Card variables, the joint hypothesis that the interactions of experience with the three education dummies are all zero cannot be rejected using a Wald test for the four- and five-year panels (p-values .871 and .673 respectively), but, we can reject this hypothesis for the six-year panel (p-value of Wald test = .0391). This rejection occurs because of a strongly significant positive interaction of experience with the education 13-15 years dummy. The point estimates of both the education <12 and the education ≥16 dummies, while not significant, are negative suggesting that there is not a monotonic relationship between education and the rate of growth of wages even in the six-year panel.

---

[25] Since a GLS regression does not maximize this R-squared (unlike an OLS regression), it is possible for this R-squared to decline when a variable is added to the regression. Also, this R-squared does not relate to a known distribution as simply as the OLS R-squared relates to the F-distribution.

The second specification, which includes the AFQT and Library Card residual variables, yields similar results. We still fail to reject the joint hypothesis that the interactions of experience with the three education dummies are all zero for the four- and five-year panels (p-values of Wald tests .514 and .804 respectively). In contrast to the first specification, we also fail to reject this hypothesis for the six-year panel (p-value = .152). Note, however, that the interaction of experience with the education 13-15 years dummy is still significantly different from zero at conventional levels. Taken together, these six regressions fail to show evidence of a systematic relationship between experience and the return to education, as predicted by our learning model.[26]

The estimates of the coefficients on the AFQT and Library Card residual variables provide further evidence consistent with the learning model. In no case is the wage related significantly to the level of the AFQT residual or the level of the Library Card residual, and in all but one case the interactions of these residuals with experience are significantly positive at conventional levels.[27] Wald tests of the joint hypothesis that the two interactions of the residuals with experience are zero can be rejected at conventional levels in all three panels (p-values = .0260, .0112, and .00016 in the four-, five-, and six-year panels respectively). This is precisely the pattern predicted by the learning model, where ability measures that are unobserved by the market have an increasing relationship with the wage as additional output signals that are correlated with these measures are

---

[26]Re-estimation of the model using a linear specification for education rather than the four education categories also resulted in no significant movement in the return to education with experience.

[27]The one exception is the interaction of the library card residual with experience in the four-year panel, and this has a p-value of .06.

incorporated into the wage.

C. Are Wage Residuals a Martingale?

The learning model implies that an individual's residuals from the estimated earnings function are a martingale. In this sub-section, we investigate how closely the covariance structures of residuals presented in Table 4 for the three balanced panels correspond to a martingale overlaid with classical measurement error. First, we present the restrictions that a martingale puts on the covariance structure. Second, we impose those restrictions using the optimal minimum-distance (OMD) estimator proposed by Chamberlain (1982, 1984) and used by Abowd and Card (1989). This yields estimates of the covariance structure under the martingale assumption, as well as a test statistic for the martingale assumption.

Suppose that the process generating the wage for worker i in year t is

$$W_{it} = X_{it}\beta + \varepsilon_{it} \;. \tag{3.5}$$

where X_{it} denotes a generic vector of regressors and the error ε_{it} is a martingale:

$$\begin{aligned} \varepsilon_{it} &= \varepsilon_{it-1} + \mu_{it-1} \\ &= \varepsilon_{i1} + \sum_{\tau=1}^{t-1} \mu_{i\tau} \;. \end{aligned} \tag{3.6}$$

where μ_{it} has mean zero and variance $\sigma^2_{\mu t}$ and is uncorrelated across workers, over time within workers, and with ε_{i1}. Suppose also that the wage is observed with error (ϕ_{it}) that has zero mean and variance σ^2_{ϕ} and is uncorrelated across workers and over time within workers. The observed wage is then

$$W^*_{it} = X_{it}\beta + \varepsilon_{it} + \phi_{it} , \tag{3.7}$$

and the observed wage error is[28]

$$\begin{aligned} \theta_{it} &= \varepsilon_{it} + \phi_{it} \\ &= \varepsilon_{i1} + \sum_{\tau=1}^{t-1} \mu_{i\tau} + \phi_{it} . \end{aligned} \tag{3.8}$$

The variance of the observed errors at experience level t is thus

$$\begin{aligned} \mathrm{Var}(\theta_{it}) &= \mathrm{Var}(\varepsilon_{it}) + \mathrm{Var}(\phi_{it}) \\ &= \sigma_1^2 + \sum_{\tau=1}^{t-1} \sigma_{\mu\tau}^2 + \sigma_\phi^2 , \end{aligned} \tag{3.9}$$

where σ_1^2 represents the variance of initial unmeasured expected worker ability. Clearly, $\mathrm{Var}(\theta_{it})$ grows with experience (t). The covariance between within-worker errors at experience levels t and s is

---

[28]This analysis ignores the contribution to the observed error from the difference between the true value of β and the OLS estimate of β.

$$(3.10) \qquad \mathrm{Cov}(\theta_{it}, \theta_{is}) = \mathrm{Var}(c_{i,\min(t,s)})$$

$$= \sigma_1^2 + \sum_{\tau=1}^{\min(t,s)-1} \sigma_{\mu\tau}^2 .$$

Intuitively, the covariance between wage residuals grows with the number of prior output signals that the two residuals have in common, namely $\min(t,s)-1$.

Consider the lower-triangular covariance matrices in Table 4 for a T-period panel. A T-period panel has a TxT covariance matrix with $T(T+1)/2$ unique elements. The hypothesis that the error structure is a martingale overlaid with measurement error imposes two types of restrictions on these $T(T+1)/2$ elements. The first restriction is that the off-diagonal elements within each column should be equal, because the covariance in (3.10) depends only on the minimum of the indices (the number of signals in common). This imposes $(T-1)(T-2)/2$ restrictions on the covariance structure. The second restriction is that the diagonal elements (variances) should be larger than the off-diagonal elements in the same column by the same amount in each column. This is due to measurement error that has a common variance across periods, and it imposes T-2 restrictions on the data. Thus, there are a total of $(T^2-T-2)/2$ restrictions from these two sources, and they can be tested by estimating the restricted covariance structure using an OMD estimator and testing the fit of the restricted model.

More formally, the $(t,s)^{th}$ element of the restricted covariance structure (σ_{ts}) can be expressed as

$$(3.11) \qquad \sigma_{ts} = \alpha_0 + \alpha_1 D_{ts} + \sum_{j=2}^{T} \alpha_j C_{tsj} ,$$

where D_{ts} is a dummy variable that equals one if t=s (i.e., if σ_{ts} is a variance), and C_{tsj} is a dummy variable that equals one if $j \leq \min(t,s)$. This is a linear model for σ_{ts}, and the parameters correspond to the error processes in (3.6) and (3.8): $\alpha_0 = \sigma_1^2$, $\alpha_1 = \sigma_\phi^2$, and $\alpha_j = \sigma_{\mu,j-1}^2$. This model has T+1 parameters for a T-period panel, so we are fitting the T(T+1)/2 unique elements of the covariance matrix with T+1 parameters. The number of restrictions is $T(T+1)/2 - (T+1) = (T^2-T-2)/2$ as described above.

In addition to the two restrictions just discussed, the prediction that the error structure is a martingale implies that all parameter estimates should be positive, because they all represent variances: α_0 is the variance of initial unmeasured expected ability, α_1 is the variance of the measurement error, and the α_j's are the variances of the innovations to market beliefs about workers' abilities.

The OMD estimator of the model of the covariance structure is a GLS estimator given the linear form of (3.11).[29] Let m_i represent the vector of the T(T+1)/2 unique elements in the cross-product matrix of residuals for worker i and let m represent the vector of means of the elements of m_i across the sample of n workers. Thus, m is our estimate of the covariance matrix of residuals. The variance matrix of the vector of covariance elements is

$$(3.12) \qquad \Omega = \frac{1}{n} \sum_{i=1}^{n} (m_i - m)(m_i - m)' ,$$

where (3.12) is simply a recasting in matrix notation of the square of (3.4). The OMD estimator minimizes the quadratic form

[29]The following derivation is a straightforward adaptation of the model used by Abowd and Card (1989).

$$(3.13) \qquad \Pi = n\cdot(m-g[\alpha])'\Omega^{-1}(m-g[\alpha])$$

where $g[\alpha]$ is the model of the covariance structure in (3.11). It is straightforward to show that Π is minimized by the GLS estimator of α, which is

$$(3.14) \qquad \hat{\alpha} = (P'\Omega^{-1}P)^{-1}(P'\Omega^{-1}m) \ ,$$

where P is the (T(T+1)/2) by (T+1) matrix of explanatory variables implied by (3.11) and m is the (T(T+1)/2)-vector of covariance elements. The covariance matrix of $\hat{\alpha}$ is

$$(3.15) \qquad \mathrm{Cov}(\hat{\alpha}) = \frac{1}{n}\cdot(P'\Omega^{-1}P).$$

Under the null hypothesis that the structure specified in (3.11) is correct, the minimized value of the quadratic form in (3.12) is distributed as χ^2 with degrees of freedom equal to the number of restrictions on the model, $(T^2-T-2)/2$. Along with computation of this test statistic, we will test the hypotheses that each element of $\hat{\alpha}$ is zero against the alternative that it is positive. To the extent that we cannot reject the $(T^2-T-2)/2$ restrictions embodied in (3.11) and the elements of $\hat{\alpha}$ are all significantly positive, the error structure is consistent with a martingale overlaid with classical measurement error.

Our estimation of the covariance structure is based on the empirical covariance matrices of residuals for the three balanced panels from the NLSY presented in Table 4. Casual examination of these three matrices shows general patterns that are consistent with the predictions of the martingale overlaid with measurement error. First, the off-diagonal elements are roughly equal in size within each column. Second, the diagonal elements in each column are larger than other elements in the column. Third, the elements increase in size moving from left to right within each row. Further casual evidence along the same lines is that OLS regressions of the covariance elements on the variables implied by (3.11) have R-squareds greater than .98 in all three panels.[30] This suggests that our model of the covariance data fits rather well.

Table 6 contains the OMD estimates of the covariance structure model for the three balanced panels specified in (3.11). In two of the three panels we cannot reject that wage residuals are a martingale overlaid with measurement error (p-value=.16 in the four-year panel and p-value=.21 in the six-year panel). Only in the five-year panel can we reject the model (p-value=.009).

Overall, the testable pattern of results is strongly consistent with the martingale prediction. All estimated elements of α (i.e., all estimated variances) are significantly positive for all three balanced panels. In particular, the variance of wage innovations is significantly positive in every period. Thus, the variance of wage residuals is growing with experience, and the covariances are growing with the number of common prior innovations to the wage.

For many commonly encountered specifications of the distributions $F(\eta_i, s_i, X_i, Z_i, B_i)$ and $G(y_{it}|\eta_i, s_i, X_i, Z_i)$, one can derive a more

---

[30]For T=4, for example, we regressed the 10 covariance elements σ_{ts} on five regressors: a constant, a variance dummy (equal to one if $t = s$), and three dummies for min(t,s) equals 1,2 and 3. The R-squared was .996. The R-squareds were .984 and .989, respectively, in the analogous regressions for the five-, and six-year balanced panels.

detailed prediction than simply that wage residuals should be a martingale: the variance of wage innovations should decline with experience. Table 6 offers no support for this prediction. The most striking feature of the estimates of α_2 through α_T, however, is that α_T greatly exceeds α_2 through α_{T-1}. This finding suggests that our use of balanced panels may not be innocuous. It could be, for example, that workers who leave the labor market after four years of experience do so because they received especially low wages in their fourth year. In keeping with this possibility, note that the estimate of α_4 for the four-year panel (1.37) exceeds that for the five-year panel (.880), and that the analogous inequality holds for α_5 in the five- and six-year panels (1.04 versus .733, respectively).

Except for the fact that α_T greatly exceeds α_2 through α_{T-1}, no clear pattern emerges from the variances of wage innovations in Table 6. The estimates are sufficiently imprecise that a gradual decrease would be difficult to reject.

We consider the evidence in Table 6 to be generally consistent with the predictions of the learning model. The evidence strongly supports the martingale prediction. Furthermore, Table 4 provides strong evidence *against* an alternative OJT-based model that predicts that the variance of wage residuals, $\text{var}(\theta_{it})$, increases with experience, as follows.

Suppose that there is no learning but that OJT amplifies ability differences, as would occur if $h(t,\eta_i) = H(t)\cdot\eta_i$ in (2.22) and $H(t)$ increases in t, for example. The analog of (3.6) is then

$$\varepsilon_{it} = H(t)\cdot\eta_i + \mu_{it}, \tag{3.16}$$

so the analog of (3.10) is

$$(3.17) \qquad \mathrm{Cov}(\theta_{it}, \theta_{is}) = H(t)\cdot H(s)\cdot \sigma_1^2 ,$$

which increases not only in min(t,s) but also in max(t,s). Thus, if residual variances increase with experience because OJT amplifies ability, then the covariances in Table 4 should increase not only across rows but also down columns. As noted earlier, casual inspection of Table 4 suggests that the covariances are constant within columns, as implied by the martingale prediction.[31]

D. How Common are Negative Wage Changes?

A central implication of the learning model is that, absent wage growth due to other factors, the expected change in the real wage paid to a worker is zero. This suggests that wage declines will not be rare events. Given that the real wage is likely to grow with experience due to investment in human capital, we expect the fraction of wage changes that are negative to be significantly less than one-half (especially for young workers, who are likely to benefit from the steep part of the concave experience profile).

Our approach in this section is to determine the frequency with which the real wage declines in the NLSY.[32] Because measurement error in wage levels is potentially a very serious problem in investigating wage changes, much of what we observe in the raw data as real wage declines could well be spurious. Our estimates of the covariance structure of wages, contained in Table 6, suggest that there is a significant amount of measurement error in

[31]Given that the covariance structure in (3.17) is nonlinear in the parameters, it is more difficult to compute the OMD estimates. For this reason, we do not present a formal test of this alternative model.

[32]McLaughlin (1990) presents an analysis of data mainly from the Panel Study of Income Dynamics and concludes that wage declines are quite common.

our data. On the other hand, recent work by Bound and Krueger (1990) comparing CPS reports of annual earnings with Social Security administrative records suggests that self-reported earnings data may be more reliable than is generally thought. We address this issue carefully in our analysis, and we make conservative assumptions about the importance of measurement error.[33]

We start with the four-year balanced panel of wage observations from our NLSY sample and focus on the three wage changes per worker within the panel. We use only wage changes for cases where both wages (W_t and W_{t-1}) refer to full-time jobs (≥ 35 hours per week) currently held at the respective interview dates. Of the 6651 (= 3x2217) potential wage changes, 5165 meet our criteria and have valid wage data. Of these, 31.8% of the measured real wage changes (and 22.2% of the nominal changes) are negative. Negative real wage changes are more common for workers who change employers (movers) than for stayers. There are 35.4% negative real wage changes (29.5% nominal) for movers compared with 30.6% negative real wage changes (19.8% nominal) for stayers. These numbers are summarized in left side of the "total" rows of Table 7.

The left side of Table 7 contains a further breakdown of the frequency of measured negative wage change by experience. The results show no clear pattern. The left side of Table 8 (8a for real changes and 8b for nominal) contains a similar breakdown by calendar year. Again, there is no clear pattern.

A key question is how many of the 30 percent reported negative wage changes are due to measurement error in the wages. In order to address this, we make a downward adjustment in the fraction negative based on a

---

[33]By conservative assumptions, we mean assumptions that err on the side of attributing more of the variance in wage changes to measurement error rather than to actual wage innovations.

decomposition of the variance of wage changes into a component due to true innovations in wages and a component due to classical measurement error.

To start, we build on our earlier results and assume that wage innovations follow a martingale overlaid with classical measurement error.[34] Using the error structure defined in (3.8), the first-difference of measured wage innovations is

$$(3.18) \qquad \Delta\theta_{it} = \theta_{it} - \theta_{it-1}$$

$$= (\varepsilon_{it} - \varepsilon_{it-1}) + (\phi_{it} - \phi_{it-1})$$

$$= \mu_{it-1} + (\phi_{it} - \phi_{it-1})$$

with variance

$$(3.19) \qquad \mathrm{Var}(\Delta\theta_{it}) = \sigma^2_{\mu t-1} + 2\sigma^2_{\phi} \ .$$

where $\sigma^2_{\mu t-1}$ is the variance of the wage innovation at t and σ^2_{ϕ} is the variance of the measurement error.[35] The first column of Table 6 contains

---

[34]It is straightforward to show that within the class of first-order autoregressive models for wage innovations, the martingale (ρ=1) attributes more of the observed variation in wage changes to measurement error than does any other value of ρ. Thus, the martingale is a conservative assumption for our purpose.

[35]This assumes classical measurement error that is independent across periods. Bound and Krueger (1990) compare reported wages from the CPS to (linked) administrative records from the Social Security Administration and find evidence that measurement error is positively serially correlated, rather than classical. To the extent that this is true in our data, the measurement error will tend to be self-correcting in first differences. Thus, we once again make a conservative assumption for our purpose.

direct estimates of $\sigma^2_{\mu t-1}$ for each level of experience as well as an estimate of σ^2_{ϕ} for the four-year panel. Our estimates suggest that measurement error accounts for a large fraction of the variance of measured wage changes each year (73.5% at year two, 78.6% at year three, and 62.9% at year four).

In order to determine what this implies for the fraction of wage changes that are negative, we must adjust the observed wage changes. One approach would be to choose a parametric form for the distribution of changes in log wages (e.g., Normal) and simply adjust the variance of that distribution to the level implied by the above calculation. We use an alternative method to reduce the spread of the distribution of wage changes without assuming a particular functional form for the distribution and without changing its mean or altering its basic shape. First, we compute an adjusted wage-change series ($\Delta\tilde{W}$) that is a weighted average of the observed wage change and the mean of the observed wage change distribution ($\Delta\bar{W}$):

$$\Delta\tilde{W}_{it} = \nu_t \Delta W_{it} + (1-\nu_t)\Delta\bar{W}_t , \tag{3.20}$$

where i indexes individuals and t indexes experience. The weight is

$$\nu_t = \left[\frac{\sigma^2_{\mu,t-1}}{\sigma^2_{\mu,t-1} + 2\sigma^2_{\phi}}\right]^{1/2} . \tag{3.21}$$

which is simply the square root of the ratio of the variance of the innovation in wages to the variance of measured changes. This adjusted wage change has mean $\Delta\bar{W}_t$ and variance $\sigma^2_{\mu,t-1}$. Next, we compute the fraction of adjusted wage changes that are negative.

The fraction of negative real wage changes is reduced from 31.8% in the unadjusted data to 16.4% in the adjusted data. The fraction negative nominal wage changes is reduced from 22.2% in the unadjusted data to 10.5% in the adjusted data. These results are presented in "total" rows of Table 7. Table 7 also contains the fractions of negative adjusted wage changes broken down by experience, and Tables 8a and 8b contain the fractions broken down by calendar year.

Since the distributions of wages changes are different for movers and stayers, we repeated our analysis for these two sub-groups; the results are contained in Tables 7 and 8. Our adjustments yield the results that 14.1% of real (8.1% of nominal) wage changes for stayers are negative while 23.3% of real (17.6% of nominal) wage changes are negative for movers.

To summarize, this analysis of the NLSY data shows that a substantial minority of workers have negative real and nominal wage declines, even after making a generous adjustment for measurement error.

4. Concluding Remarks

We generated four predictions from our learning model: 1) there should be no systematic relationship between the return to market-observed skill (education) and experience, 2) there should be a positive relationship between the return to market-unobserved skill (the AFQT and library-card residuals) and experience, 3) wage residuals should be a martingale, and 4) wage cuts should not be rare. Our empirical evidence supports all four predictions.

While there is general empirical support for the learning model, there are configurations of the investment in OJT model that are consistent with each of our findings taken separately. In a pure-OJT explanation of our full array of findings, however, worker heterogeneity related to investment in training must be independent of heterogeneity related to education. Since training and education are both forms of human capital, we find this independence condition awkward, especially in comparison to the simple way the learning model predicts the relationships we find. The learning model, however, cannot account for measured returns to experience. We conclude that a blend of the learning model with an on-the-job-training model is more plausible than either model alone.

REFERENCES

Abowd, J. and D. Card (1989), "On the Covariance Structure of Earnings and Hours Changes," *Econometrica*, 57: 411-445.

Albrecht, J. (1981), "A Procedure for Testing the Signaling Hypothesis," *Journal of Public Economics*, 15: 123-32.

Becker, G. (1964), *Human Capital*. New York: NBER.

Ben Porath, Y. (1967), "The Production of Human Capital and the Life-Cycle of Earnings," *Journal of Political Economy*, 75: 352-365.

Bound, J., and A. Krueger (1991), "The Extent of Measurement Error in Longitudinal Earnings Data: Do Two Wrongs Make a Right?," *Journal of Labor Economics*, 9: 1-24.

Card, D., and A. Krueger (1990), "Does School Quality Matter? Returns to Education and the Characteristics of Public Schools in the United States," Princeton University Industrial Relations Section Working Paper #265, May.

Chamberlain, G. (1982), "Multivariate Regression Models for Panel Data," *Journal of Econometrics*, 18: 5-46.

__________ (1984), "Panel Data," Chapter 22 in *The Handbook of Econometrics*, Vol. 2, ed. by Z. Griliches and M. Intrilligator. New York: North Holland.

Chung, K.L. (1974), *A Course in Probability Theory*. New York: Academic Press.

Harris, M. and B. Holmstrom (1982), "A Theory of Wage Dynamics," *Review of Economic Studies*, 49: 315-333.

Hause, J. (1972), "Earnings Profiles: Ability and Schooling," *Journal of Political Economy*, 80: S108-S138.

Holmstrom, B. (1982), "Managerial Incentive Problems---A Dynamic Perspective," in *Essays in Economics and Management in Honor of Lars Wahlbeck*. Stockholm: Swedish School of Economics.

Lang, K. and D. Kropp (1986), "Human Capital versus Sorting: The Effects of Compulsory Attendance Laws," *Quarterly Journal of Economics*, 101: 609-24.

Layard, R. and G. Psacharopoulos (1974), "The Screening Hypothesis and the Returns to Education," *Journal of Political Economy*, 82: 985-98.

MacDonald, G. (1982), "A Market Equilibrium Theory of Job Assignment and Sequential Accumulation of Information," *American Economic Review*, 72: 1038-55.

McLaughlin, K. (1990), "Rigid Wages?," University of Rochester Department of Economics Working Paper #229, May.

Mincer, J. (1974), *Schooling, Experience, and Earnings*. New York: Columbia University Press for the NBER.

Riley, J. (1979), "Testing the Educational Screening Hypothesis," *Journal of Political Economy*, 87: S227-52.

Spence, A.M. (1973), "Job Market Signalling," *Quarterly Journal of Economics*, 87: 355-79.

Stigler, G. (1962), "Information in the Labor Market," *Journal of Political Economy*, 70: 94-105.

Taubman, P. and T. Wales (1973), "Higher Education, Mental Ability andScreening," *Journal of Political Economy*, 81: 28-55.

Table 1:

Average Age and Education by Entry Cohort and Sex

Overall

Entry Cohort	N	Average Age	Average Education
1979	675	19.6	12.2
1980	577	19.8	12.5
1981	636	20.1	12.6
1982	550	20.9	12.9
1983	607	21.4	13.1
1984	585	22.4	13.2
All	3630	20.7	12.7

Males

Entry Cohort	N	Average Age	Average Education
1979	339	19.4	11.9
1980	262	19.6	12.1
1981	339	20.0	12.3
1982	282	20.7	12.5
1983	308	21.2	12.8
1984	295	22.1	12.9
All	1825	20.5	12.4

Females

Entry Cohort	N	Average Age	Average Education
1979	336	19.8	12.5
1980	315	20.1	12.9
1981	297	20.2	12.9
1982	268	21.1	13.3
1983	299	21.6	13.3
1984	290	22.8	13.5
All	1805	20.9	13.0

Note: Year of entry for each worker is defined as interview year preceding first three-year spell classified as primarily working. Experience is defined as zero at year of entry. See text for details.

Table 2

Summary Statistics
NLSY
(Means and Standard Deviations)

All Years of Employment (1979-1987)

by Experience

Exper- ience	N	Wage	Age	Educ- ation	Part- Time	Coll. Barg	Non- White	Female	Mar- ried	Marr.& Female
0	776	4.76 (1.76)	21.1 (2.07)	12.4 (2.08)	0.0	.182	.420	.499	.159	.106
1	3327	5.12 (2.10)	21.7 (2.14)	12.8 (2.10)	.141	.159	.389	.500	.167	.104
2	3382	5.67 (2.45)	22.7 (2.15)	12.9 (2.13)	.0920	.176	.393	.495	.247	.139
3	3363	6.09 (2.67)	23.7 (2.15)	12.9 (2.15)	.0737	.190	.397	.501	.314	.180
4	2661	6.42 (2.93)	24.4 (2.01)	12.9 (2.11)	.0973	.188	.380	.500	.372	.203
5	2050	6.61 (2.95)	25.1 (1.94)	12.9 (2.15)	.101	.185	.380	.500	.419	.224
6	1507	6.96 (3.22)	25.8 (1.85)	12.8 (2.14)	.110	.202	.378	.504	.464	.247
7	968	7.22 (3.30)	26.7 (1.75)	12.8 (2.10)	.112	.215	.362	.519	.486	.260
8	487	7.56 (3.39)	27.6 (1.62)	12.7 (2.10)	.113	.263	.361	.476	.476	.216
Total:	18521	6.06 (2.78)	23.7 (2.62)	12.9 (2.13)	.0985	.185	.387	.500	.314	.175

Notes: The numbers in parentheses are standard deviations. The Part-Time, Collective Bargaining, Nonwhite, Female, Married, and Married&Female variables are dummy variables. Wage data are in real 1982-1984 dollars (deflated by CPI). Observations at the time of entry (experience = 0) which are part-time are not included in this analysis. See text for details.

Table 3
Summary Statistics
NLSY - Balanced Samples
(Means and Standard Deviations)

Means of Worker Characteristics at Time of Entry

	1 thru 4 years experience	1 thru 5 years experience	1 thru 6 years experience
Number of Wrkrs.	2217	1610	1100
Age at entry	20.4 (1.98)	20.1 (1.94)	19.9 (1.84)
Education	12.8 (2.04)	12.7 (2.03)	12.6 (1.98)
Education < 12	.155	.177	.189
Education = 12	.445	.443	.451
Education 13-15	.227	.214	.205
Education ≥ 16	.174	.166	.155
Nonwhite	.363	.360	.354
Female	.498	.500	.505
AFQT Residual	.300 (14.65)	.483 (14.88)	.805 (14.90)
Library Card Residual	-.00670 (.428)	-.00532 (.427)	-.00524 (.424)

Means of Pooled Data

	1 thru 4 years experience	1 thru 5 years experience	1 thru 6 years experience
Number of Obs.	8868	8050	6600
Wage	5.86 (2.55)	5.98 (2.61)	6.17 (2.74)
Part-Time	.0959	.0960	.0964
Collective Barg.	.180	.183	.194
Married	.259	.286	.321
Married & Female	.146	.155	.169

Notes: The Part-time, Collective Bargaining, Nonwhite, Female, Married, and Married&Female variables are dummy variables. Wage data are in real 1982-1984 dollars (deflated by CPI). The AFQT and Library Card residuals are computed from regressions of the AFQT and Library Card variables on dummy variables for education category, a complete set of age dummies, nonwhite, female, part-time at the time of entry, the interaction of part-time with the education categories, a complete set of year-at-time-of-entry dummies, and the real wage at time of entry.

Table 4
Empirical Covariance Matrix of Within-Worker Wage Residuals
(standard error)

1 through 4 Years Experience Balanced Sample (n=2217)

	1	2	3	4
1	3.14 (.190)			
2	2.02 (.142)	4.10 (.213)		
3	1.84 (.137)	2.69 (.166)	4.45 (.187)	
4	1.94 (.158)	2.90 (.189)	3.44 (.194)	6.17 (.322)

1 through 5 Years Experience Balanced Sample (n=1610)

	1	2	3	4	5
1	3.09 (.235)				
2	1.94 (.154)	3.84 (.215)			
3	1.67 (.147)	2.47 (.162)	4.21 (.202)		
4	1.81 (.177)	2.67 (.188)	3.17 (.199)	5.97 (.372)	
5	1.68 (.160)	2.29 (.174)	2.85 (.192)	3.88 (.260)	6.03 (.306)

1 through 6 Years Experience Balanced Sample (n=1100)

	1	2	3	4	5	6
1	2.69 (.187)					
2	1.78 (.162)	3.63 (.250)				
3	1.57 (.164)	2.36 (.195)	4.14 (.245)			
4	1.67 (.181)	2.54 (.222)	3.05 (.228)	5.88 (.447)		
5	1.59 (.175)	2.27 (.205)	2.78 (.221)	3.76 (.274)	5.67 (.337)	
6	1.59 (.176)	2.48 (.231)	3.02 (.241)	3.95 (.338)	4.31 (.321)	7.65 (.472)

(see note on next page)

Note to Table 4:

Based on residuals from separate OLS regressions for each year-experience cell of the real wage on dummy variables for age (all values), education category (four values), part-time status, the interaction of the education dummies with part-time status, collective bargaining status, nonwhite, female, marital status, and the interaction of female and marital status. The elements of the covariance matrix are computed as defined in equation 3.3 and the standard errors of the elements of the covariance matrix are computed as defined in equation 3.4.

Table 5: GLS Estimation of Earnings Functions[a]
Balanced Samples

	1 thru 4 years experience		1 thru 5 years experience		1 thru 6 years experience	
	(1)	(2)	(3)	(4)	(5)	(6)
years of experience	.479 (.043)	.396 (.0542)	.506 (.0561)	.436 (.0649)	.406 (.0813)	.365 (.0943)
education < 12 years	-.998 (.635)	-1.17 (.742)	-.544 (.772)	-.863 (.908)	-.262 (1.10)	-.283 (.441)
education 13-15 years	.508 (.419)	.178 (.494)	.748 (.572)	.0332 (.655)	-1.20 (.914)	-1.32 (1.06)
education ≥ 16 years	3.15 (.477)	3.53 (.554)	3.04 (.679)	2.79 (.758)	3.96 (1.15)	3.19 (1.30)
(educ. < 12) *experience	-.0724 (.0915)	-.0315 (.120)	-.138 (.114)	-.0698 (.139)	-.117 (.156)	-.105 (.200)
(13 ≤ educ ≤ 15) *experience	.00289 (.0670)	.0681 (.0867)	-.0375 (.0844)	.0673 (.0999)	.285 (.127)	.294 (.148)
(educ. ≥ 16) *experience	-.0201 (.0801)	-.0935 (.0996)	-.0536 (.105)	-.00675 (.119)	-.0946 (.162)	.00642 (.185)
AFQT residual	---	.00289 (.00486)	---	.00182 (.00472)	---	-.00348 (.00496)
AFQT residual *experience	---	.00305 (.00155)	---	.00276 (.00127)	---	.00388 (.00121)
Library Card Residual	---	.0297 (.166)	---	.157 (.164)	---	.165 (.173)
Library Card Resid *experience	---	.0832 (.0529)	---	.0779 (.0442)	---	.0926 (.0423)
p-value of Wald test education*experience	.871	.514	.673	.804	.0391	.152
p-value of Wald test AFQT and Lib. Card interactions with exp		.0260		.0112		.000158
R^2 (GLS estimates)	.244	.249	.244	.243	.264	.261
number of workers	2217	2217	1610	1610	1100	1100

[a]Summary statistics are contained in table 3. The first observation for each individual is omitted from the analyses in columns (2), (4), and (6). The dependent variable is average hourly earnings on the most recent job deflated by the 1982-84=100 CPI. All regressions also include a constant, age at entry, and dummy variables for part-time, the interaction of part-time and the education dummies, year dummies, the interaction of year dummies with the education dummies, collective bargaining status, race, sex, marital status, and sex*marital status. The base year is 1987 and the base education level is 12 years. The R-squared is computing applying the GLS estimates to the untransformed data. Standard errors are in parentheses.

Table 6:
Optimal Minimum Distance Estimation of Covariance Structure
Random Walk Overlaid with Classical Measurement Error
NLSY Balanced Panels
(standard errors in parentheses)

Parameter	Associated Variance	1 thru 4 years experience	1 thru 5 years experience	1 thru 6 years experience
Variance of initial unmeasured expected ability:				
α_0	σ_1^2	1.89	1.70	1.59
		(.130)	(.101)	(.0825)
Variance of measurement error:				
α_1	σ_ϕ^2	1.16	1.19	1.10
		(.0837)	(.0825)	(.0842)
Variances of wage innovations each period:				
α_2	$\sigma_{\mu 1}^2$	.837	.682	.644
		(.103)	(.102)	(.118)
α_3	$\sigma_{\mu 2}^2$	.633	.589	.655
		(.0993)	(.109)	(.123)
α_4	$\sigma_{\mu 3}^2$	1.37	.880	.772
		(.218)	(.135)	(.139)
α_5	$\sigma_{\mu 4}^2$	---	1.04	.733
			(.203)	(.156)
α_6	$\sigma_{\mu 5}^2$	---	---	1.86
				(.344)
χ^2 statistic, structural test:		7.93	22.1	17.9
degrees of freedom:		5	9	14
p-value of test statistic:		.160	.0086	.211
number of workers		2217	1610	1100

Note: The optimal minimum distance estimator is the GLS regression of the unique elements of the appropriate covariance matrix in table 4 on the variables implied by 3.11. The covariance matrix of the covariance elements in table 4 is defined in 3.12. The chi-squared test statistic is the minimized value of the quadratic form in 3.13, and the degrees of freedom are the number of restrictions 3.11 places on the unrestricted covariance matrices in table 4.

Table 7
Analysis of Wage Changes
Fraction Negative Wage Changes with Adjustments for Measurement Error
Four-Year Balanced Sample, NLSY
by Experience

Fraction Negative Real Wage Change
(cell size)

	Raw			Adjusted		
Experience	All	Stayers	Movers	All	Stayers	Movers
2	.315 (1647)	.306 (1194)	.338 (453)	.140 (1647)	.116 (1194)	.203 (453)
3	.310 (1795)	.300 (1354)	.342 (441)	.148 (1795)	.125 (1354)	.218 (441)
4	.330 (1723)	.313 (1322)	.384 (401)	.204 (1723)	.180 (1322)	.284 (401)
Total	.318 (5165)	.306 (3870)	.354 (1295)	.164 (5165)	.141 (3870)	.233 (1295)

Fraction Negative Nominal Wage Change
(cell size)

	Raw			Adjusted		
Experience	All	Stayers	Movers	All	Stayers	Movers
2	.200 (1647)	.181 (1194)	.252 (453)	.0728 (1647)	.0570 (1194)	.115 (453)
3	.226 (1795)	.198 (1354)	.311 (441)	.0930 (1795)	.0665 (1354)	.175 (441)
4	.239 (1723)	.213 (1322)	.327 (401)	.149 (1723)	.119 (1322)	.247 (401)
Total	.222 (5165)	.198 (3870)	.295 (1295)	.105 (5165)	.0814 (3870)	.176 (1295)

Note: The adjusted wage change series ($\Delta\tilde{W}_{it}$) is computed as a weighted average of the observed wage change (ΔW_{it}) and the sample mean wage change ($\Delta\bar{W}_t$): $\Delta\tilde{W}_{it} = \theta_t \Delta W_{it} + (1-\theta_t)\Delta\bar{W}_t$ where $\theta_t = \left[\frac{\sigma^2_{\mu,t-1}}{\sigma^2_{\mu,t-1} + 2\sigma^2_{\phi}}\right]^{.5}$. The values of $\sigma^2_{\mu,t-1}$ and $2\sigma^2_{\phi}$ are taken from the estimates in the first column of table 6. Wage changes are computed using all non-part-time observations in the four-year balanced panel on workers with a job at the relevant interview dates.

Table 8a
Analysis of Wage Changes
Fraction Negative Wage Changes with Adjustments for Measurement Error
Four-Year Balanced Sample, NLSY
by Year

Fraction Negative Real Wage Change
(cell size)

	Raw			Adjusted		
Year	All	Stayers	Movers	All	Stayers	Movers
81	.370 (370)	.368 (261)	.376 (109)	.170 (370)	.149 (261)	.220 (109)
82	.310 (688)	.315 (502)	.296 (186)	.131 (688)	.112 (502)	.183 (186)
83	.341 (1041)	.321 (853)	.431 (188)	.175 (1041)	.142 (853)	.325 (188)
84	.342 (976)	.338 (723)	.356 (253)	.178 (976)	.151 (723)	.257 (253)
85	.275 (1039)	.253 (763)	.337 (276)	.150 (1039)	.128 (763)	.210 (276)
86	.294 (704)	.268 (508)	.362 (196)	.162 (704)	.146 (508)	.204 (196)
87	.323 (347)	.327 (260)	.310 (87)	.199 (347)	.189 (260)	.230 (87)
Total	.318 (5165)	.306 (3870)	.354 (1295)	.164 (5165)	.141 (3870)	.233 (1295)

Note: The adjusted wage change series $(\Delta\tilde{W}_{it})$ is computed as a weighted average of the observed wage change (ΔW_{it}) and the sample mean wage change $(\Delta\bar{W}_t)$: $\Delta\tilde{W}_{it} = \theta_t \Delta W_{it} + (1-\theta_t)\Delta\bar{W}_t$ where $\theta_t = \left[\frac{\sigma^2_{\mu,t-1}}{\sigma^2_{\mu,t-1} + 2\sigma^2_{\phi}}\right]^{.5}$. The values of $\sigma^2_{\mu,t-1}$ and $2\sigma^2_{\phi}$ are taken from the estimates in the first column of table 6. Wage changes are computed using all non-part-time observations in the four-year balanced panel on workers with a job at the relevant interview dates.

Table 8b
Analysis of Wage Changes
Fraction Negative Wage Changes with Adjustments for Measurement Error
Four-Year Balanced Sample, NLSY
by Year

Fraction Negative Nominal Wage Change
(cell size)

	Raw			Adjusted		
Year	All	Stayers	Movers	All	Stayers	Movers
81	.178 (370)	.169 (261)	.202 (109)	.0514 (370)	.0460 (261)	.0642 (109)
82	.179 (688)	.165 (502)	.215 (186)	.0683 (688)	.0418 (502)	.140 (186)
83	.276 (1041)	.250 (853)	.394 (188)	.109 (1041)	.0821 (853)	.229 (188)
84	.230 (976)	.199 (723)	.316 (253)	.114 (976)	.0858 (723)	.194 (253)
85	.201 (1039)	.165 (763)	.301 (276)	.109 (1039)	.0826 (763)	.181 (276)
86	.236 (704)	.209 (508)	.306 (196)	.129 (704)	.110 (508)	.179 (196)
87	.208 (347)	.189 (260)	.264 (87)	.141 (347)	.119 (260)	.207 (87)
Total	.222 (5165)	.198 (3870)	.295 (1295)	.105 (5165)	.0814 (3870)	.176 (1295)

Note: The adjusted wage change series ($\Delta\tilde{W}_{it}$) is computed as a weighted average of the observed wage change (ΔW_{it}) and the sample mean wage change ($\Delta\bar{W}_t$): $\Delta\tilde{W}_{it} = \theta_t \Delta W_{it} + (1-\theta_t)\Delta\bar{W}_t$ where $\theta_t = \left[\frac{\sigma^2_{\mu,t-1}}{\sigma^2_{\mu,t-1} + 2\sigma^2_{\phi}}\right]^{.5}$. The values of $\sigma^2_{\mu,t-1}$ and $2\sigma^2_{\phi}$ are taken from the estimates in the first column of table 6. Wage changes are computed using all non-part-time observations in the four-year balanced panel on workers with a job at the relevant interview dates.