LEARNING AND WAGE DYNAMICS*

Henry S. Farber
Princeton University

Robert Gibbons
Cornell University

First draft: August 1990

This version: May 1994

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 predicted to be independent of labor-market experience. The model also predicts that time-invariant variables correlated with ability but unobserved by employers should be increasingly correlated with wages as experience increases and that wage residuals should be a martingale.

We present evidence from the National Longitudinal Survey of Youth that is generally consistent with the model's predictions, but a chi-squared goodness-of-fit test does reject the martingale prediction for wage residuals even after accounting for classical measurement error. We investigate alternative specifications and find that a modification of the learning model that allows for worker ability to evolve as an AR1 process fits the data quite well.

*We are grateful to James Powell for advice on computing the covariance matrices for unbalanced samples. We thank David Card, Richard Freeman, Lawrence Katz, Kevin Lang, Kevin M. Murphy, Steve Pischke, Jeffrey Wooldridge, and participants in numerous workshops for comments on an earlier version of this paper. We thank Gerald Oettinger, Caroline Minter-Hoxby, and Paul Oyer for able research assistance. Financial support from the DOL (Farber: E-9-J-9-0050), the NSF (Farber: SES-8912664; Gibbons: SES-9010759), an Alfred P. Sloan Research Fellowship (Gibbons), and the Industrial Relations Section at Princeton University is also gratefully acknowledged.

1. Introduction

A worker's education level and other observable characteristics convey only partial information about the worker's productive ability. As the worker accumulates experience in the labor market, further information seems likely to be 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 our 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 predicted to be independent of labor-market experience. The model also predicts that time-invariant variables correlated with ability but unobserved by employers should be increasingly correlated with wages as experience increases, and that wage residuals should be a martingale.

These three predictions are derived from very mild assumptions on the distribution of worker characteristics and the form of the production function. We assume that wages equal expected output at each date and that the stochastic component of a worker's output has a time-invariant distribution. Given these two assumptions, we are able to apply a variant of the law of iterated expectations, which holds without requiring further assumptions on probability distributions or functional forms.

We examine the learning model's predictions using twelve years of 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, when learning is arguably most important. 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 evidence we uncover is broadly consistent with all three of the learning model's predictions. In particular, in a GLS regression involving 28,984 observations on 4,970 workers, we find no evidence that the relationship between earnings and education varies significantly with experience, but we find strong evidence that time-invariant worker characteristics correlated with ability but unobserved by employers are increasingly correlated with wages as experience increases. And while the χ^2 test from optimal-minimum-distance (OMD) estimation of the covariance structure of wage residuals rejects the martingale prediction, the main qualitative features of the empirical covariance structure of wage residuals are as predicted.

In an attempt to fit the empirical covariance structure of wage residuals more closely, we briefly consider a pair of alternative models of the covariance structure. First, we fit a model with serially correlated measurement error in wages rather than the classical measurement error assumed in the initial model. While this model fits the data significantly better than did the initial model, it is still strongly rejected by the data. Second, we fit a new learning model in which each worker's ability evolves as an AR1, rather than being time-invariant as in our initial learning model. Unlike our initial model, this AR1 model assumes normal distributions and linear functional forms, so as to deliver closed-form predictions. To our surprise, this AR1 model, with only four parameters in its error structure, is not strongly rejected by the χ^2 test from the OMD estimation of the covariance structure of wage residuals.

The organization of the paper is straightforward. Section 2 presents the theory and section 3 the evidence. Finally, in section 4 we own up to the fact that our analysis has completely ignored two large literatures, on

job mobility and human-capital acquisition, and we sketch how our theory and evidence might be integrated into these literatures in future work.

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 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 $\{y_{it} : t = 1, \dots, T\}$ are independent draws from the conditional distribution $G(y_{it} | \eta_i, s_i)$, but we allow this distribution to be arbitrary.¹

¹In Appendix A we analyze an example in which these schooling and production functions are linear with normal disturbances.

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 $\{y_{i1}, \dots, y_{it}\}$ 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):

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

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, as would be the case if wages were 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 yields several new predictions concerning the estimated coefficients in earnings regressions. For the univariate case analyzed in this sub-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 ($t = 1, \dots, T$). Consider using wage data from year t to estimate the earnings regression

$$(2.2) \quad w_{it} = \alpha_t + \beta_t s_i + \varepsilon_{it} .$$

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

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

By the law of iterated expectations,²

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

or

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

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

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

Because a worker's outputs are independent draws from the conditional distribution $G(y_{it} | \eta_i, s_i)$, the conditional expectation $E(y_{it} | s_i)$ is

²The law of iterated expectations states that $E\{E(y|x,z)|x\} = E(y|x)$. See Chung (1974, Chapter 9).

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.³

The intuition behind this result is as follows. Our assumption that wages equal expected output implies not only that the first-period wage, w_{11} , 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, $w_{12} - w_{11}$, can be forecasted from the information used to determine w_{11} (here, schooling). (Equation (2.5) restates this fact for the innovation in wages between the first and t^{th} periods: ψ_{1t} is orthogonal to s_1 .) Thus, the dependent variable in the second-period regression, w_{12} , equals the dependent variable in the first-period regression, w_{11} , plus a term orthogonal to schooling, $w_{12} - w_{11}$. Therefore, the estimated coefficient on schooling is the same in the first and second periods.

We conclude this sub-section by noting three possible reinterpretations of the model just described. First, after a suitable modification of the wage equation (2.1), what we call output (y_{1t}) could 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). Second, the learning model can be interpreted as a dynamic extension of Spence's (1973) static signaling model.⁴ By allowing the joint distribution

³To avoid potential confusion, we hasten to note that this time-invariant estimated coefficient (typically) is not a consistent estimate of any structural parameter of interest; see the example in Appendix A.

⁴While Spence's 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.

$F(\eta_i, s_i)$ 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). Third, 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).⁵

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 effect of learning about worker ability 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 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

⁵Given the significant effect of school quality on earnings documented by Card and Krueger (1992), this reinterpretation may be quite important empirically.

are included in the data but are *not* directly observable by employers, such as the score on an aptitude test taken at a specific time.

As in Section 2A, the joint distribution of worker characteristics, $F(\eta_1, s_1, X_1, Z_1, B_1)$, is arbitrary, and the outputs $\{y_{it} : t = 1, \dots, T\}$ are independent draws from the conditional distribution $G(y_{it} | \eta_1, s_1, X_1, Z_1)$, which also is arbitrary. To distinguish the background variables from the worker's ability, we assume that B_1 has no direct effect on output: the conditional distribution $G(y_{it} | \eta_1, s_1, X_1, Z_1, B_1) = G(y_{it} | \eta_1, s_1, X_1, Z_1)$ for every B_1 . 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_1 and Z_1 , the wage-determination equation (2.1) now becomes

$$(2.7) \quad w_{it} = E(y_{it} | s_1, X_1, Z_1, y_{i1}, \dots, y_{i,t-1}) ,$$

and given data on the vector X_1 as well as on schooling, the regression (2.2) now becomes

$$(2.8) \quad w_{it} = \alpha_t + \beta_t s_1 + X_1 \gamma_t + \varepsilon_{it} .$$

The estimated coefficients $(\hat{\alpha}_t, \hat{\beta}_t, \hat{\gamma}_t)$ from this regression are the coefficients from the linear projection⁶ of w_{it} on s_1 and X_1 , denoted

⁶The 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.

$E^*(w_{it}|s_i, X_i)$:

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

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: $E^*\{E(y|x)|x\} = E^*(y|x)$.⁷

These two results imply the following analogue of the laws of iterated expectations and projections: $E^*\{E(y|x, z)|x\} = E^*(y|x)$.⁸ Because the wage equals expected output, applying this last result to our model yields

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

But $E^*(y_{it}|s_i, X_i)$ is independent of t , because 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 $\{y_{it} : t = 1, \dots, T\}$ 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 $w_{it} = E(y_{it}|s_i, X_i, Z_i, y_{i1}, \dots, y_{i,t-1})$ then the estimated effect of schooling on wages is independent of t in the sequence of regressions $w_{it} = \alpha_t + \beta_t s_i + X_i \gamma_t + \varepsilon_{it}$ for $t = 1, \dots, T$.

⁷Define $\varepsilon = y - E(y|x)$. Then $E(\varepsilon|x) = 0$, so $E^*(\varepsilon|x) = 0$ because the linear projection equals the conditional expectation when the latter is linear. Thus, $0 = E^*\{y - E(y|x)|x\} = E^*(y|x) - E^*\{E(y|x)|x\}$.

⁸By 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.

We now derive two 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; and (2) wage residuals are a martingale.

(1) *Unobserved Characteristics*: Recall that B_1 is a vector of background variables that are included in the data but are unobservable by employers. The distribution of worker characteristics, $F(\eta_1, s_1, X_1, Z_1, B_1)$ is arbitrary, however, so the other variables observable to employers (namely, s_1 , X_1 , and Z_1) could be correlated with B_1 . To create a vector of variables that are orthogonal to employers' information when the worker enters the labor market, define B_1^* to be the residual from a regression of B_1 on all the other variables included in the data (namely s_1 and X_1) and on the worker's initial wage, w_{11} :

$$(2.11) \quad B_1^* = B_1 - E^*(B_1 | s_1, X_1, w_{11}) .$$

Regressing B_1 on the worker's initial wage purges B_1^* of the correlation between Z_1 and B_1 (provided that there is no measurement error in the observed initial wage).

Now add B_1^* as a regressor in (2.8):

$$(2.12) \quad w_{it} = \alpha_t + \beta_t s_{it} + X_{it} \gamma_t + B_1^* \pi_t + \epsilon_{it} .$$

We are interested in how the estimated coefficients $\hat{\pi}_t$ vary with experience. For ease of exposition, take B_1 to be a scalar. Since B_1^* is orthogonal to the other regressors in (2.12), the estimated coefficient $\hat{\pi}_t$ is

given by

$$(2.13) \quad \hat{\pi}_t = \frac{\text{cov}(B_1^*, w_{1t})}{\text{var}(B_1^*)} .$$

To solve for $\text{cov}(B_1^*, w_{1t})$, note that

$$\begin{aligned} (2.14) \quad w_{1t} &= E(y_{1t} | s_1, X_1, Z_1, y_{11}, \dots, y_{1,t-1}) \\ &= E(y_{1t} | s_1, X_1, Z_1, y_{11}, \dots, y_{1,t-2}) + \zeta_{1t} \\ &= E(y_{1,t-1} | s_1, X_1, Z_1, y_{11}, \dots, y_{1,t-2}) + \zeta_{1t} \\ &= w_{1,t-1} + \zeta_{1t} \\ &= w_{11} + \sum_{\tau=2}^t \zeta_{1\tau} . \end{aligned}$$

Since B_1^* is orthogonal to w_{11} by construction, we have $\hat{\pi}_1 = 0$ and

$$\begin{aligned} (2.15) \quad \text{cov}(B_1^*, w_{1t}) &= \text{cov}(B_1^*, \sum_{\tau=2}^t \zeta_{1\tau}) \\ &= \sum_{\tau=2}^t \text{cov}(B_1^*, \zeta_{1\tau}) \end{aligned}$$

for $t > 1$. For many commonly encountered specifications of the distributions $F(\eta_1, s_1, X_1, Z_1, B_1)$ and $G(y_{1t} | \eta_1, s_1, X_1, Z_1)$, $\text{cov}(B_1^*, \zeta_{1\tau})$ is positive for every τ . Given this regularity condition, $\hat{\pi}_t$ increases with t . Stated less formally, if B_1^* is correlated with ability then the estimated effect of B_1^* 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*: Because $E(\zeta_{it} | w_{i,t-1}) = 0$ in (2.14), wages are a martingale:⁹ $E(w_{it} | w_{i,t-1}) = w_{i,t-1}$. 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; we do so in Section 3. For purposes of illustration, however, consider the residual from (2.8):

$$\begin{aligned} (2.16) \quad \hat{\varepsilon}_{it} &= w_{it} - (\hat{\alpha}_t + \hat{\beta}_t s_i + X_i \hat{\gamma}_t) \\ &= E(y_{it} | s_i, X_i, Z_i, y_{i1}, \dots, y_{i,t-1}) - E^*(y_{it} | s_i, X_i), \end{aligned}$$

$$\text{so } E(\hat{\varepsilon}_{it} | \hat{\varepsilon}_{i,t-1}) = \hat{\varepsilon}_{i,t-1}.$$

C. Time-Varying Worker Characteristics

In the pure learning model developed above, the assumptions that worker characteristics are time-invariant and that outputs have identical

⁹A 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.

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. We now enrich the learning model to allow for productivity growth with labor-market experience. We do this for two reasons: 1) to give a precise statement of the empirical model we will analyze in Section 3, and 2) to assess whether the learning model's predictions continue to hold in the presence of various aspects of human-capital theory. 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

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

We continue to assume that the outputs $\{y_{it} : t = 1, \dots, T\}$ 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 independent of η_i, s_i, X_i , and Z_i .

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

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

and in the T regressions by experience level in (2.8), the estimated intercepts $\{\hat{\alpha}_t: t = 1, \dots, T\}$ are now larger than previously by the amount $h(t)$. These T regressions can be collapsed into the single pooled regression

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

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, all three predictions from Section 2B continue to hold: The effect of schooling on 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., $\hat{\varepsilon}_{it}$ from (2.20), analogous to $\hat{\varepsilon}_{it}$ from (2.8), as computed in (2.16)) are a martingale.

We now assess whether these 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

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

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 $h(t, s_1)$, and so will not equal zero. The second and third predictions remain unchanged.

Now suppose instead that (2.18) is replaced by

$$(2.22) \quad Y_{it} = y_{it} + h(t, \eta_1) ,$$

where the cross-partial derivative of $h(t, \eta_1)$ 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 $h(t, \eta_1)$ and the correlation between η_1 and s_1 , as determined from the distribution $F(\eta_1, s_1, X_1, Z_1, B_1)$. 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_1^* and experience will reflect both learning (as described in Section 2B) and the cross-partial derivative of $h(t, \eta_1)$, 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 $h(t, \eta_1) = h(t) \cdot \eta_1$, for example).

In summary, for each of the three 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 three 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 sub-section, we consider whether a pure human-capital model (i.e., a model without learning) can produce the learning model's three 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 unobserved skill. 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) that are themselves a martingale, but this prediction does not follow naturally from an OJT model.

3. Empirical Analysis

In this section we first describe the data from the NLSY that we use for our empirical analysis. 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 consider two alternative covariance structures. 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 made their first long-term transition to the labor force, from which we can compute their labor market experience. 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 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. At the time we carried out our analysis, there were data available for the 1979 through 1991

¹⁰ Inspection of Table 1 indicates that the average difference between age and education varies with entry cohort, ranging from 7.5 years in 1979 to 12.7 years in 1988. 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 4998 workers is only .37. If potential experience were the correct measure, this correlation would be one.

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.¹¹ Only individuals aged 16 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 2857 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.

¹¹At the 1979 interview date, the last interview was assumed to be January 1, 1978.

Individuals were dropped from the sample if there were not three consecutive interviews at which they were classified as primarily working during the preceding year. On this basis, the last year individuals could make their first long-term transition to the labor force was between the 1988 and 1989 interviews. (Because our data end in 1991, we cannot be sure that workers who entered after 1988 were primarily working for three years.) We dropped 2903 individuals who never made a long-term transition to the labor force by this definition, as well as thirteen individuals who were self-employed at all interviews after entry. Finally, we dropped four individuals who were not born in the 1957 through 1964 period (they were born in 1965) and 901 individuals with missing data on key variables.

The final sample consists of 4998 individuals who made their initial long-term transition to the labor force (by our definition) between 1979 and 1988.¹² Table 1 summarizes average age and average education by year of labor-market entry. The earlier entry 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-consecutive-years working-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 14 individuals out of the 4998 in

¹²To recapitulate, there are 12686 individuals in the NLSY. The 1280 workers in the military subsample were deleted, leaving 11406 individuals in the basic sample of the NLSY. Of these, 2587 were primarily working in the first year we could classify them, 2903 others did not have three consecutive years primarily working, 13 were self-employed in each year, 901 had missing data, and 4 were born outside the 1957-1964 period. This leaves 4998 individuals in the final sample.

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: 3563 were never classified as primarily working prior to their first long-term transition, but 836 were primarily working for one year, 465 for two years, and 134 for three or more years. Overall, our rule captures what seems to be a reasonably sharp transition from not working to working.

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 date in order to select the first post-transition wage observation appropriately. Assume (for expositional 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 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 the 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.¹³

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 34742 valid observations for the 4998 workers. If there were complete data for each of the workers for every year since entry, our sample would have 46367 observations. Of the 11625 observations lost, 3829 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 at a later interview. 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 the worker's ability but are not

¹³We 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 58 percent of jobs held at t are part-time while about 10 percent of jobs held at any later interview date are part-time.) Of the 4998 individuals in our sample, only 1169 have valid zero-experience wage observations (by our definition). There are 776 observations for workers who are not currently with a job at t and 1709 observations for individuals working part-time at t . A further 1344 individuals either have no reported wage at t or are missing data for other variables at t .

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, as follows.

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^*) from a regression of B_i on attributes observable by the econometrician (X_i) is more promising:

$$(3.1) \quad B_i^* = B_i - X_i \hat{\gamma},$$

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:

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

where $\hat{\gamma}$ and $\hat{\delta}$ are the estimated parameters from an OLS regression of B_i on X_i and w_{i0} .¹⁴ In the model in Section 2, the wage at any date incorporates *all*

¹⁴The 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

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_1^* 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 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.¹⁵ 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.¹⁶ No measures of parental educational or occupational attainment were used because these variables are

earnings by 1955 can be interpreted as evidence that learning about ability goes on for quite a long time.

¹⁵The 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.

¹⁶The three household-environment variables (newspapers, magazines, and library cards at age 14) are, as expected, significantly positively correlated.

missing for a significant fraction of the workers in our sample.¹⁷

In order to generate the residual B_1^* from (3.2), we create a sample with only one observation per individual by taking the first valid observation for each of the 4998 individuals in our sample.¹⁸ 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, race, sex, age, calendar year, and the real wage.¹⁹ This regression accounted for 53 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 ($R^2 = .07$), 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 sub-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

¹⁷For example, mother's education was missing for 264 and father's education was missing for 602 of the 4998 individuals in the sample.

¹⁸As noted above, only 1169 of the 4998 workers have their first valid observation at zero years experience. Of the rest, 3508 are first included at one year, 260 at two years, and the remaining 61 at three or more years.

¹⁹All 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.

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.²⁰ 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. 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.

Table 3 contains estimates of the within-worker covariance matrix of wage residuals. 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 female and marital status. Our data are unbalanced, in the sense that there are different numbers of wage observations for different workers.

We compute each covariance using all available observations for that element. Some notation is useful here. Let e_{it} represent the residual at experience level t for the i^{th} worker. Define a dummy variable $d_{i,ts} = 1$ if e_{it} and e_{is} are both observed. Let $N_{ts} = \sum_{i=1}^N d_{i,ts}$ denote the number of observations for which both e_{it} and e_{is} are observed. Now define $p_{ts} = N_{ts}/N$,

²⁰The 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.

the fraction of the sample for which both e_{it} and e_{is} are observed. Finally, define $r_{i,ts} = d_{i,ts} / p_{ts}$.

A natural estimator of the covariance between ε_t and ε_s using the unbalanced data is

$$(3.3) \quad \hat{\sigma}_{ts} = \sum_{i=1}^N d_{i,ts} e_{it} e_{is} = (1/N) \sum_{i=1}^N r_{i,ts} e_{it} e_{is}.$$

Next, we need to compute an estimator for the covariance of $\hat{\sigma}_{ts}$ and $\hat{\sigma}_{qr}$. A consistent estimator of this covariance is

$$(3.4) \quad V_{ts,qr} = (1/N) \sum_{i=1}^N r_{i,ts} r_{i,qr} (e_{it} e_{is} - \hat{\sigma}_{ts})(e_{iq} e_{ir} - \hat{\sigma}_{qr}).$$

The standard errors of the $\hat{\sigma}_{ts}$'s are simply the square roots of the appropriate $V_{ts,ts}$'s.

Let c denote the unrestricted TxT covariance matrix of residuals whose elements are defined by equation 3.3. Estimates of this covariance matrix based on the sample of 34742 wage observations for the 4998 workers are contained in Table 3. We also computed estimates of the covariance matrix of first differences in residuals. The first-differenced residuals were computed simply by differencing the residuals from the wage levels regressions. Expressions for the estimated covariance of first-differenced residuals and the covariance matrix of these estimated covariances of first-differenced residuals are defined analogously to equations 3.3 and 3.4. Our estimates of the covariance matrix of first-differenced residuals are contained in Table 4. We defer discussion of the specific form of the covariance matrices in Tables 3 and 4 to the next sub-section.

We now turn to using our estimates of the covariance matrix of wage residuals for GLS estimation of earnings functions. The covariance matrix

for individual i is c_i , and it is composed of elements of c corresponding to the pairs of residuals observed for individual i . Assuming independence of residuals across workers, the covariance matrix of residuals for the entire sample, C , is block diagonal with the block being c_i for each individual i . The unrestricted estimates of c presented in Table 3 are used to compute C for the overall sample. These estimates of C are then used to compute GLS estimates of the earnings function.

Table 5 contains GLS estimates of regressions of the real wage for the unbalanced sample. The specification includes a constant, age at entry, education, experience, the square of experience, a dummy variable for part-time employment, the interaction of part-time status and education, year fixed effects, the interaction of year fixed effects with education, collective bargaining status, race, sex, marital status, and the interaction of sex and marital status. The base year is 1991. The model also includes the AFQT and Library Card residuals and the interaction of these residuals with experience 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 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 unweighted data.

The results show the usual strong positive relationship of earnings with experience and education. Given that the earnings function is estimated in levels rather than logs, the coefficient of education cannot be interpreted directly as the return to education, but evaluating the proportional effect of education on earnings at the mean level of earnings yields an estimated return to education of 9.1 percent.

Consistent with the learning model, there is no evidence that the relationship between earnings and education varies with experience. The

estimated coefficient on the interaction of education and experience is not significantly different from zero (p-value = 0.33 using a two-tailed test).²¹

The estimates of the coefficients on the AFQT and Library Card residual variables provide further evidence consistent with the learning model. The estimated coefficients of the interactions of these residuals with experience are both significantly positive. A Wald test of the joint hypothesis that the two interactions of the residuals with experience are zero can be rejected at conventional levels of significance (p-value = 1.2×10^{-8}). This is 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 incorporated into the wage.

A small puzzle is that the estimated main effect of the AFQT residual is significantly positive. Given that the residuals are computed conditional on the first observed wage (and, hence, all information observed by the market at the point), we do not expect there to be strong relationship between the residuals and the wage at lower levels of experience. Indeed, the estimated main effect of the library card residual is not significantly different from zero. The significantly positive main effect of the AFQT residual suggests that variation in the AFQT score orthogonal to the measured observables including the first observed wage is correlated with earnings. One interpretation of this result consistent with the learning model is that substantial new information about worker quality correlated with the AFQT

²¹The interaction of education and experience is positive in a model that does not include year fixed effects and allow the return to education to vary by year. Given the well-documented increase in the return to education during the 1980's (e.g., Katz and Murphy, 1992) and the fact that experience in our sample is growing over time, it is not surprising that the return to education would seem to grow with experience if the return to education is not allowed to vary by calendar year.

score is revealed shortly after the first wage observation, and wages are adjusted at that time to reflect that new information. Thus, the relationship between the wage and the AFQT score may not be linear in experience.

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 Tables 3 and 4 correspond to a martingale overlaid with classical measurement error. First, we present the restrictions that a martingale puts on each of these covariance structures. 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

$$(3.5) \quad w_{it} = X_{it}\beta + \varepsilon_{it} ,$$

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

$$(3.6) \quad \begin{aligned} \varepsilon_{it} &= \varepsilon_{it-1} + \mu_{it-1} \\ &= \varepsilon_{i1} + \sum_{\tau=1}^{t-1} \mu_{i\tau} , \end{aligned}$$

where μ_{it} has mean zero and variance $\sigma_{\mu t}^2$ and is uncorrelated across workers, over time within workers, and with ε_{11} . Suppose also that the wage is observed with error (ϕ_{it}) that has zero mean and variance σ_{ϕ}^2 and is uncorrelated across workers and over time within workers. The observed wage is then

$$(3.7) \quad w_{it}^* = X_{it}\beta + \varepsilon_{it} + \phi_{it} ,$$

and the observed wage error is²²

$$(3.8) \quad \begin{aligned} \theta_{it} &= \varepsilon_{it} + \phi_{it} \\ &= \varepsilon_{11} + \sum_{\tau=1}^{t-1} \mu_{i\tau} + \phi_{it} . \end{aligned}$$

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

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

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

²²This analysis ignores the contribution to the observed error from the difference between the true value of β and the OLS estimate of β .

$$\begin{aligned}
 (3.10) \quad \text{Cov}(\theta_{it}, \theta_{is}) &= \text{Var}(\varepsilon_{i, \min(t,s)}) \\
 &= \sigma_1^2 + \sum_{\tau=1}^{\min(t,s)-1} \sigma_{\mu\tau}^2 .
 \end{aligned}$$

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 covariance matrices in Table 3 for a T-period panel. A T-period panel has a $T \times T$ 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 $(t,s)^{\text{th}}$ element of the restricted covariance structure (σ_{ts}) can be expressed as

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

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 initial variance of unmeasured 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 martingale prediction also has implications for the covariance structure of the first-difference of earnings, and the implied structure is quite simple. Let $\Delta\theta_{it} = \theta_{it} - \theta_{it-1}$ represent the first difference of observed earnings residuals. From equation 3.8, this is

$$(3.8') \quad \Delta\theta_{it} = \mu_{t-1} + (\phi_{it} - \phi_{it-1}).$$

Assuming classical measurement error, the variance of the first difference of the observed errors at experience level t is thus

$$(3.9') \quad \text{Var}(\Delta\theta_{it}) = \sigma_{\mu t-1}^2 + 2\text{Var}(\phi_{it}).$$

Given that the innovations to the wage process are independent, the covariances between first differences of within-worker errors at experience levels t and s have a particularly simple structure. The first covariance is

simply the negative of the measurement error variance:

$$(3.10') \quad (\text{Cov}(\Delta\theta_{it}, \Delta\theta_{it-1}) = -\text{Var}(\phi_{it}).$$

Higher-order covariances are all zero:

$$(3.10'') \quad \text{Cov}(\Delta\theta_{it}, \Delta\theta_{is}) = 0 \quad (|t-s| > 1).$$

As with levels, the $(t,s)^{\text{th}}$ element of the restricted covariance structure of first differences of earnings (σ_{ts}) can be expressed as a linear function. This is

$$(3.11') \quad \sigma_{ts} = \alpha_1 (2D_{ts} - A_{ts}) + \sum_{j=2}^T \alpha_j L_{tsj},$$

where D_{ts} is a dummy variable that equals one if $t=s$ (i.e., if σ_{ts} is a variance), A_{ts} is a dummy variable that equals one if $|t-s|=1$ (i.e., if σ_{ts} is a first-order covariance) and L_{tsj} is a dummy variable that equals one if $t=s=j$ (i.e., if σ_{ts} is the j^{th} diagonal element of the covariance matrix).

As before, this is a linear model for σ_{ts} , and the parameters correspond to the error processes in (3.8'): $\alpha_1 = \sigma_{\phi}^2$, and $\alpha_j = \sigma_{\mu, j-1}^2$. This model has T parameters for a T -period panel of first differences, so we are fitting the $T(T+1)/2$ unique elements of the covariance matrix with T parameters. The number of restrictions is $T(T+1)/2 - T = T(T-1)/2$.

The OMD estimator of the model of the covariance structure (either in levels or first differences) is a GLS estimator given the linear forms of

3.11 and 3.11'.²³ In order to avoid duplication, we will present this analysis in terms of the levels model and then apply it to the covariance structure in both levels and in first differences. 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 covariance matrix of the vector of covariance elements is

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

The OMD estimator minimizes the quadratic form

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

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

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

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

²³The following derivation is a straightforward adaptation of the model used by Abowd and Card (1989).

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

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 in both levels and first differences presented in Tables 3 and 4, respectively. Casual examination of the levels covariance structure in Table 3 shows general patterns that are consistent with the predictions of the martingale overlaid with classical measurement error. First, the diagonal elements in each column are larger than other elements in the column (consistent with measurement error). Second, the elements increase in size moving from left to right within each row (as new common innovations are added). On the other hand, the theory predicts that the covariances within each column should be roughly the same size, and they are not: the covariances shrink quite consistently moving from lower-order to higher-order covariances. A formal goodness-of-fit test requires the OMD estimates.

The first column of Table 6 contains the OMD estimates of the covariance model specified in (3.11). Using the χ^2 statistic defined above, we clearly reject that wage residuals are a martingale overlaid with

measurement error ($p\text{-value}=1\times 10^{-11}$). On the other hand, the pattern of estimates is consistent with the martingale prediction. All estimated elements of α (i.e., all estimated variances) are significantly positive. 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. We computed an R-squared as the fraction of (unweighted) variation in the elements of the covariance structure that can be accounted for by the model. The value of the R-squared using the covariance structure in levels and the estimates in the first column of Table 6 is 0.80.²⁴ The rejection of the structure seems to be due largely to the declining covariances (which the structure requires to be constant) within each column.

We turn now to first differences. Casual examination of the first differences covariance structure in Table 4 once again shows general patterns that are consistent with the predictions of the martingale overlaid with classical measurement error. First, the diagonal elements in each column are large relative to the other elements in the covariance matrix (consistent with measurement error). Second, the first covariances are strongly negative and of roughly the same size throughout (again consistent with measurement error). Third, the higher-order covariances (which are predicted to be zero) are all quite small, though predominantly negative.

The second column of Table 6 contains the OMD estimates of the model of the covariance structure of first differences of wage residuals specified in (3.11'). Using the χ^2 statistic defined above, we reject that wage residuals

²⁴Note that the OMD estimates are not designed to maximize this R-squared. An equally-weighted minimum distance estimator (equivalent to OLS) would maximize this R-squared.

are a martingale overlaid with measurement error (p-value=0.0009). On the other hand, the pattern of estimates is again consistent with the martingale prediction. All estimated elements of α (i.e., all estimated variances) are significantly positive. In particular, the variance of wage innovations is significantly positive in every period. The estimated variance of the measurement error is also significantly positive. As in the levels case, we ask how much of the variation in Table 4 can be accounted for by these estimates. The value of the R-squared using the covariance structure in first differences and the estimates in the second column of Table 6 is 0.97. The rejection of the martingale covariance structure seems to be due largely to the small negative values observed for the higher-order covariances, which the structure requires to be zero.²⁵

We consider the evidence in Table 6 to be generally consistent with the predictions of the learning model, despite the formal rejection of the martingale prediction. Furthermore, Table 3 provides strong evidence *against* an alternative model of on-the-job training (OJT) 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 $h(t, \eta_1) = H(t) \cdot \eta_1$ in (2.22) and $H(t)$ increases in t , for example. The analog of (3.6) is then

$$(3.16) \quad \varepsilon_{it} = H(t) \cdot \eta_1 + \mu_{it} ,$$

so the analog of (3.10) is

²⁵The covariance structure implied by the martingale can be rejected against a less constrained model with one more parameter where the higher order covariances can take on a common non-zero value. We return to estimation of alternative structures below.

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

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 3 should increase not only across rows but also down columns. As noted earlier, casual inspection of Table 3 suggests that the covariances decline within columns as opposed to being constant within columns, as implied by the martingale prediction. On the other hand, there is no evidence that these covariances are increasing.

We now consider of richer conceptions of the measurement error process and the learning process that might be consistent with the observed covariance structures.

D. Alternative Covariance Structures

There are at least two approaches to estimating a less constrained covariance structure. The first involves a less constrained specification for the measurement error process. The second involves a richer conception of the evolution of worker types in the learning model. In this sub-section, we investigate two alternative error structures and compare their goodnesses of fit to the structure analyzed in the previous sub-section.

We first investigate a simple AR1 process for the measurement error. This was suggested by the work of Bound and Krueger (1991), who find that measurement error in annual income is serially correlated. The model with classical measurement error that we analyzed in the previous sub-section is nested in this model. Specifically, the special case of the AR1 measurement-error model where the AR1 coefficient equals zero is equivalent to classical

measurement error.

Let the measurement error for worker i in period t be

$$(3.18) \quad \phi_{it} = \rho \phi_{it-1} + v_{it},$$

where v_{it} is i.i.d. and ρ is the serial-correlation coefficient. If $\rho=0$ then the measurement error is classical. In the steady-state, the variance of observed errors in levels is

$$(3.19) \quad \begin{aligned} \text{Var}(\theta_{it}) &= \text{Var}(\varepsilon_{it}) + \text{Var}(\phi_{it}) \\ &= \sigma_1^2 + \sum_{\tau=1}^{t-1} \sigma_{\mu\tau}^2 + \frac{1}{(1-\rho^2)} \sigma_v^2, \end{aligned}$$

and the covariance of observed errors is

$$(3.20) \quad \begin{aligned} \text{Cov}(\theta_{it}, \theta_{is}) &= \text{Var}(\varepsilon_{i, \min(t,s)}) + \text{Cov}(\phi_{it}, \phi_{is}) \\ &= \sigma_1^2 + \sum_{\tau=1}^{\min(t,s)-1} \sigma_{\mu\tau}^2 + \frac{\rho^{|t-s|}}{(1-\rho^2)} \sigma_v^2. \end{aligned}$$

This covariance structure has the property that the covariances are declining within columns, as the data in Table 3 suggest.

The AR1 process for measurement error also has implications for the covariance structure of first differences. In the steady-state, the variance of first differences of observed wages is

$$(3.21) \quad \text{Var}(\Delta\theta_{it}) = \sigma_{\mu t-1}^2 + \frac{2}{(1+\rho)} \sigma_v^2,$$

and the covariance of first differences of observed wages is

$$(3.22) \quad \text{Cov}(\Delta\theta_{it}, \Delta\theta_{is}) = \frac{(\rho-1)}{(1+\rho)} \rho^{|t-s|-1} \sigma_\nu^2.$$

These covariances are negative and declining with the order of the covariance; the sign is consistent with Table 4, but the decline is not apparent there.

In both levels and first differences, the covariance structure implied by this model is nonlinear in ρ and linear in the remaining parameters conditional on ρ . Let the parameters of the model (ρ , σ_ν^2 , σ_1^2 , and $\sigma_{\mu\tau}^2$, $\tau=1, \dots, T-1$) be denoted by the vector α , and let the appropriate nonlinear function of α (defined by equations 3.19 and 3.20 for the levels model and by equations 3.21 and 3.22 in first differences) be represented by the function $g(\alpha)$. Then the OMD estimates of α are obtained by minimizing the quadratic form in equation 3.13 with respect to α . The covariance matrix of $\hat{\alpha}$ is

$$(3.23) \quad \text{Cov}(\hat{\alpha}) = \frac{1}{n} \cdot (F' \Omega^{-1} F)^{-1},$$

where $F = \partial g(\alpha) / \partial \alpha$.²⁶ We computed these estimates using a Newton-Raphson algorithm.

The estimated value of ρ for the covariance structure of the level of wages is 0.136 (s.e. = 0.019), and the estimates of the other parameters are qualitatively identical to those in first column of Table 6.²⁷ The goodness-

²⁶Note that this reduces to the usual linear GLS estimator when g is a linear function of α , as in the simple learning model analyzed earlier.

²⁷Note that our estimate of ρ is approximately the same as the estimate of the AR1 coefficient in the measurement error process for annual

of-fit statistic is 132.5 with 53 degrees of freedom. This compares with the goodness-of-fit statistic of 157 with 54 degrees of freedom for the classical-measurement-error model presented in the first column of Table 6. The difference between these statistics is distributed as χ^2 with one degree of freedom, and it is clear that the hypothesis that $\rho=0$ can be rejected at any reasonable level of significance. However, the goodness-of-fit statistic of 132.5 in the AR1 measurement error model (distributed as χ^2 with 53 degrees of freedom) clearly rejects this model as well.²⁸

The estimated value of ρ for the covariance structure of first differences of wages is 0.095 (s.e. = 0.019), and the estimates of the other parameters are qualitatively identical to those in second column of Table 6. The goodness-of-fit statistic is 66.8 with 43 degrees of freedom. This compares with the goodness-of-fit statistic of 79.0 with 44 degrees of freedom for the classical-measurement-error model presented in the second column of Table 6. Once again, it is clear that the hypothesis that $\rho=0$ can be rejected at any reasonable level of significance. As in levels, however, the goodness-of-fit statistic of 66.8 in the AR1 measurement error model (distributed as χ^2 with 53 degrees of freedom) rejects that model as well (p-value = 0.012), though not as strongly as before.²⁹

We conclude that simply assuming AR1 measurement error on top of the learning model from Section 2 is not sufficient to explain the observed covariance structure of wages, either in levels or first differences.

income estimated by Bound and Krueger (1991).

²⁸The R-squared implied by the the AR1 measurement error model using the covariance structure in levels and the OMD estimates is 0.88.

²⁹The R-squared implied by the the AR1 measurement error model using the covariance structure in first differences and the OMD estimates is 0.97.

A richer approach to the learning model allows worker types to change over time. One relatively simple characterization allows worker types to evolve as an AR1, but even this simple model gets complicated rather quickly if one proceeds at the level of generality of Section 2. In order to derive a tractable empirical model for the covariance structure, we impose a linear production function and normal probability distributions. The details of this learning model are given in Appendix B; here we simply state the stochastic structure of the learning model and present the steady-state covariance structure.

Let η_{i0} represent the initial quality of worker i . Suppose that, rather than being fixed, η evolves as

$$(3.24) \quad \eta_{it} = \rho \eta_{it-1} + \nu_{it},$$

where ρ is the AR1 parameter (bounded between -1 and $+1$), and ν_{it} is the innovation to worker i 's ability in period t . (One interpretation is that the worker's characteristics are fixed but their value in a changing world is not.) A worker's ability is never observed directly, but the market sees a noisy signal of ability by observing output. We assume that all relevant random variables are distributed normally, including the initial distribution of worker ability (η_0), the innovation (ν_t), and output conditional on ability. Given that the wage in any period is equal to the worker's expected output in that period, we derive the covariance structure of wages given observable worker characteristics in any period. This derivation is contained in the appendix.

The linear-normal model where ability evolves as an AR1 and wages are observed with classical measurement error has five parameters: 1) the variance (σ_z^2) of worker characteristics not included in the data but observed

by the market and fixed over time (Z_1), 2) the variance of the measurement error (σ_ϕ^2), 3) the AR1 parameter (ρ), 4) the variance (σ_0^2) of initial worker ability (η_{01}), and 5) the variance of the innovation to worker ability (σ_ν^2). As described in Appendix B, we analyze the steady-state of this model in which the variances σ_0^2 and σ_ν^2 determine the steady-state variance of ability, which (with some abuse of notation) we denote by σ_1^2 . Thus, the model we estimate has only four parameters (σ_z^2 , σ_ϕ^2 , ρ , and σ_1^2).

The variance of the level of observed errors is

$$(3.25) \quad \text{Var}(\theta_{it}) = \text{Var}(\varepsilon_{it}) + \text{Var}(\phi_{it})$$

$$= \sigma_z^2 + (1 + \rho^2 + \dots + \rho^{2(t-1)})\sigma_1^2 + \sigma_\phi^2,$$

and the covariance is

$$(3.26) \quad \text{Cov}(\theta_{it}, \theta_{is}) = \rho^{(|t-s|)}(1 + \rho^2 + \dots + \rho^{2(\min(t,s)-1)})\sigma_1^2.$$

The variance of the first difference in observed errors is

$$(3.27) \quad \text{Var}(\Delta\theta_{it}) = \left(1 + (1-\rho)^2(1 + \rho^2 + \dots + \rho^{2(t-1)})\right)\sigma_1^2 + 2\sigma_\phi^2,$$

and the first-order covariance of the first difference in observed errors is

$$(3.28) \quad \text{Cov}(\Delta\theta_{it}, \Delta\theta_{it+1}) =$$

$$\left(- (1-\rho) + \rho(1-\rho)^2(1 + \rho^2 + \dots + \rho^{2(t-1)})\right)\sigma_1^2 - \sigma_\phi^2.$$

Finally, the higher-order covariances of the first differences in observed errors are

$$(3.29) \quad \text{Cov}(\Delta\theta_{1t}, \Delta\theta_{1s}) = \rho^{(|t-s|-1)} \left(-(1-\rho) + \rho(1-\rho)^2(1 + \rho^2 + \dots + \rho^{2(\min(t,s)-1)}) \right) \sigma_1^2.$$

As with the AR1 measurement error model analyzed earlier in both levels and first differences, the covariance structure implied by this model is nonlinear in ρ and linear in the remaining parameters conditional on ρ . Once again, let the parameters of the model (σ_ϕ^2 , ρ , and σ_1^2) be denoted by the vector α , and let the appropriate nonlinear function of α (defined by equations 3.25 and 3.26 for the levels model and by equations 3.27, 3.28, and 3.29 in first differences) be represented by the function $g(\alpha)$. Then the OMD estimates of α are obtained by minimizing the quadratic form in equation 3.13 with respect to α , and the covariance matrix of $\hat{\alpha}$ is as defined in equation 3.23. Once again we used the Newton-Raphson algorithm to minimize the quadratic form.

These models fit the data remarkably well, particularly since they have only a few parameters (four in levels, three in first differences). For the levels model, the estimates are contained in the first column of Table 7. The estimated AR1 parameter is 0.91 (s.e. = 0.0098), and the goodness-of-fit statistic is 91.5 with 62 degrees of freedom. The R-squared computed using this structure is 0.96. Recall that the goodness-of-fit statistic was 157 (with 54 degrees of freedom) in the base model discussed in the previous sub-section, and the goodness-of-fit statistic was 132.5 (with 53 degrees of freedom) in the model with AR1 measurement error. Thus, the model with an AR1 process determining worker ability does much better with many fewer parameters than either of these (4 parameters vs. 12 and 13 parameters, respectively, for the earlier models). In this analysis of wage levels we can still reject the AR1 model of worker ability (p-value = 0.009), but this

model clearly does a better job than either of the other models.

For the first-difference model, the estimates are presented in the second column of Table 7. This model cannot be rejected at conventional levels of significance, despite having only three free parameters. The estimated AR1 parameter is 0.933 (s.e. = 0.0123) and the goodness-of-fit statistic is 64.59 with 52 degrees of freedom. The R-squared computed using this structure is 0.97. The goodness-of-fit statistic was 79.0 (with 44 degrees of freedom) in the base model discussed in the previous sub-section, and the goodness-of-fit statistic was 66.8 (with 43 degrees of freedom) in the model with AR1 measurement error. Once again, the model with an AR1 process determining worker ability does better with many fewer parameters (3 parameters vs. 11 and 12, respectively, for the earlier models). In this analysis of first differences, we cannot reject the AR1 model of worker ability (p-value = 0.137).

4. Final Remarks

In brief, we find our initial learning model to be a parsimonious yet effective tool for understanding important aspects of wage dynamics. This finding is somewhat surprising in light of the facts that the model allows for human capital acquisition only in a very simple way and that it completely ignores job transitions. But it would be incorrect to infer from our results that future research should ignore job mobility and human capital. To the contrary, we see this paper as providing theory and evidence on the role of learning about worker quality in wage dynamics, and we expect the theory and empirical evidence presented here to be integrated into future work on mobility and human capital.

Most research on learning in labor markets has focused on the process through which workers find and evaluate matches with jobs³⁰. Such research studies the determinants of job mobility and the effect of work history on earnings. In this paper, in contrast, we completely ignore jobs, so as to focus on the implications of learning about worker quality for wage dynamics. One interpretation of the improved fit we find from our AR1 ability model, however, may be that jobs are (of course!) important. That is, the AR1 model may be a reduced form for a more traditional model of matching and mobility (although alternative interpretations are possible). This issue is far too complex to be analyzed in this paper, but we intend to explore it in future work.

While we use learning to animate wage dynamics, the standard model for analyzing wage dynamics is human-capital theory (Becker, 1964; Mincer, 1974). We completely ignore the relationship between human-capital acquisition and ability, again to focus on the implications of learning about worker quality. As Stigler (1962) observed, however, disentangling the effects of learning and on-the-job training (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. Since our learning model is silent about important features of the data, especially the measured return to experience, it cannot offer a complete explanation of the data. As we describe in Section 2C, however, the OJT model cannot support the full range of our empirical results

³⁰See Jovanovic (1979) and MacDonald (1982) for theory, Farber (forthcoming) and Flinn (1986) for evidence, and McCall (1980) for a nice integration of the two. Harris and Holmstrom (1982) is an exception, and may be the paper most closely related to ours: they impose linearity and normality assumptions but consider the added complexity of insurance.

unless one imposes awkward assumptions regarding the relationship among ability, schooling, and investment in OJT. We conclude that a blend of the learning and OJT models offers the best explanation of wage dynamics.

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 (1992), "Does School Quality Matter? Returns to Education and the Characteristics of Public Schools in the United States," *Journal of Political Economy*, 100: 1-400.
- 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.
- Farber, H.S., "The Analysis of Inter-Firm Worker Mobility," *Journal of Labor Economics*. (forthcoming)
- Flinn, C. (1986) "Wages and Job Mobility of Young Workers," *Journal of Political Economy*, 94: S88-S110.
- 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.
- Jovanovic, B. (1979), "Job Matching and the Theory of Turnover," *Journal of Political Economy*, 87: 972-990.
- Katz, L. and K.M. Murphy (1992), "Changes in Relative Wages, 1963-1987: Supply and Demand Factors," *Quarterly Journal of Economics*, 107: 35-78.
- 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.

McCall, B. (1990) "Occupational Matching: A Test of Sorts," *Journal of Political Economy*, 98: 45-69.

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 and Screening," *Journal of Political Economy*, 81: 28-55.

Appendix A

In this appendix we derive closed-form results for a linear-normal example of our main learning model. We assume that ability (η_i) is normally distributed with mean m and precision h . Schooling (s_i) is a linear function of ability: $s_i = \eta_i + \delta_i$, where δ_i is normally distributed with mean zero and precision h_δ , and is independent of ability. The market's belief about ability after observing schooling is therefore normal with mean $m_0(s_i) = (hm + h_\delta s_i)/(h + h_\delta)$ and precision $h_0 = h + h_\delta$.

Output is a linear function of schooling and ability: $y_{it} = c + bs_i + \eta_i + \xi_{it}$, where ξ_{it} is normal with mean zero and precision h_ξ , independent across dates, and independent of η_i and δ_i . Wages are therefore

$$\begin{aligned} w_{it} &= E(y_{it} \mid s_i, y_{i1}, \dots, y_{i,t-1}) \\ &= c + bs_i + E(\eta_i \mid s_i, y_{i1}, \dots, y_{i,t-1}) \\ &= c + bs_i + \frac{hm + h_\delta s_i + h_\xi(z_{i1} + \dots + z_{i,t-1})}{h + h_\delta + (t-1)h_\xi}, \end{aligned}$$

where $z_{it} = y_{it} - c - bs_i$ is the noisy signal about ability that the market extracts from output. In particular, $w_{i1} = c + bs_i + m_0(s_i)$.

Since we show in the text that regressions of earnings on schooling at each date yield the same estimated coefficients (that is, $w_{it} = \hat{c} + \hat{b}s_i + \hat{\epsilon}_{it}$ for all t), we can compute $\hat{b}$ and $\hat{c}$ at $t = 1$ (when the regression yields an $R^2 = 1$, in this example):

$$\begin{aligned} \hat{b} &= \frac{dw_{i1}}{ds_i} = b + \frac{dm_0(s_i)}{ds_i} = b + \frac{h_\delta}{h + h_\delta}, \\ \hat{c} &= E(w_{i1}) - \hat{b}E(s_i) = c + \frac{h}{h + h_\delta} m. \end{aligned}$$

Neither estimate is consistent (that is, $\hat{b} > b$, and $\hat{c} >(<) c$ when $m >(<) 0$), but the residual has a natural interpretation:

$$\begin{aligned}\hat{\varepsilon}_{it} &= w_{it} - \hat{c} - \hat{b}s_i \\ &= E(\eta_i \mid s_i, y_{i1}, \dots, y_{i,t-1}) - m_0(s_i),\end{aligned}$$

so $\hat{\varepsilon}_{it}$ is a forecast error—the difference between the expectation of η_i given $(s_i, y_{i1}, \dots, y_{i,t-1})$ and the earlier expectation given only s_i .

Appendix B

We now extend the linear-normal example from Appendix A to allow ability to be an AR1 rather than time-invariant. Let initial ability (η_{i0}) be normally distributed with mean m and precision H_0 . As before, schooling is $s_i = \eta_{i0} + \delta_i$, where δ_i is normal with mean zero and precision h_δ , and is independent of η_{i0} . The market's belief about initial ability after observing schooling is therefore normal with mean $m_0(s_i) = (hm + h_\delta s_i)/(h + h_\delta)$ and precision $h_0 = h + h_\delta$.

Ability at later dates is $\eta_{it} = \rho\eta_{i,t-1} + v_{it}$, where v_{it} is normal with mean zero and precision h_v , independent across dates, and independent of η_{i0} and δ_i . The market's belief about η_{i1} after observing schooling is therefore normal with mean $\rho m_0(s_i)$ and precision $H_1 = h_0 h_v / (\rho^2 h_v + h_0)$.

Output is $y_{it} = c + bs_i + \eta_{it} + \xi_{it}$, where ξ_{it} is normal with mean zero and precision h_ξ , independent across dates, and independent of η_{i0} , δ_i , and v_{it} for all t . The market's belief about η_{i1} after observing s_i and y_{i1} is therefore normal with mean

$$m_{i1} = \frac{H_1 \rho m_0(s_i) + h_\xi z_{i1}}{H_1 + h_\xi}$$

and precision $h_1 = H_1 + h_\xi$.

These updating rules have natural analogs at date t . Let η_{it} given s_i and $y_{i1}, \dots, y_{i,t-1}$ be normal with mean $\rho m_{i,t-1}$ and precision H_t . Then η_{it} given s_i and $y_{i1}, \dots, y_{it}$ is normal with mean

$$m_{it} = \frac{H_t \rho m_{i,t-1} + h_\xi z_{it}}{H_t + h_\xi}$$

and precision $h_t = H_t + h_\xi$. Therefore, $\eta_{i,t+1}$ given s_i and $y_{i1}, \dots, y_{it}$ is normal with mean ρm_{it} and precision $H_{t+1} = h_t h_v / (\rho^2 h_v + h_t)$.

In this model the earnings regressions $w_{it} = \hat{c}_t + \hat{b}_t s_i + \hat{\varepsilon}_{it}$ no longer yield time-invariant coefficients. Instead, one can derive that

$$\hat{b}_t = b + \rho^t \frac{h_\delta}{h + h_\delta} \quad \text{and}$$

$$\hat{c}_t = c + \rho^t \frac{h}{h + h_\delta} m.$$

As in Appendix A, however, the residual can be interpreted as a forecast error—the difference between the expectation of η_{it} given s_i and $y_{i1}, \dots, y_{i,t-1}$ and the expectation of η_{it} given only s_i :

$$\hat{\varepsilon}_{it} = \rho m_{i,t-1} - \rho^t m_0(s_i).$$

The law of iterated expectations implies that $\hat{\varepsilon}_{i,t+1} = \rho(\hat{\varepsilon}_{it} + \lambda_{it})$, where λ_{it} is orthogonal to $\hat{\varepsilon}_{it}$ and has mean zero. Thus, residuals are no longer a martingale (the point of the exercise), and $\text{var}(\hat{\varepsilon}_{i,t+1}) = \rho^2 \text{var}(\hat{\varepsilon}_{it}) + \rho^2 \text{var}(\lambda_{it})$.

While the law of iterated expectations tells us nothing about $\text{var}(\lambda_{it})$, our distributional assumptions do: $\text{var}(\lambda_{it}) = h_\xi / H_t (H_t + h_\xi)$. Also, in this example (until modified below), the earnings regression at $t = 1$ yields an $R^2 = 1$, so $\hat{\varepsilon}_{i1} \equiv 0$ and $\text{var}(\hat{\varepsilon}_{i1}) = 0$. Thus, one could in principle compute $\text{var}(\hat{\varepsilon}_{it})$ iteratively, but this produces a highly nonlinear function of ρ that is not tractable for estimation. Instead, we move to the steady state.

If $H_t = H$ for all t then $\text{var}(\lambda_{it}) = h_\xi / H (H + h_\xi) \equiv \text{var}(\lambda_i)$ for all t , so $\text{var}(\hat{\varepsilon}_{i2}) = \rho^2 \text{var}(\lambda_i)$ and

$$\text{var}(\hat{\varepsilon}_{it}) = [1 + \rho^2 + \dots + \rho^{2(t-2)}] \text{var}(\hat{\varepsilon}_{i2})$$

for $t \geq 2$. One can also derive that $\text{cov}(\hat{\varepsilon}_{it}, \hat{\varepsilon}_{i,t+s}) = \rho^s \text{var}(\hat{\varepsilon}_{it})$. We will take $\text{var}(\hat{\varepsilon}_{i2})$ as parametric and estimate the value of ρ that achieves the best fit between this predicted covariance structure and the empirical covariance structure in Table 3.

Before taking the model to the data, however, we make two changes. First, we add a constant term to the covariance structure. As in Section 2B, let there be a time-invariant worker characteristic z_i that is observed by employers but not included in the data. If $y_{it} = c + bs_i + \eta_{it} + z_i + \xi_{it}$, where z_i is orthogonal to the other variables in the model, then $\hat{\epsilon}_{it}$ includes z_i and each variance and covariance term includes $\text{var}(z_i)$.

The second change is essentially notational. The first wage in the model corresponds to what we have called the zero-experience wage in the data—both are conditioned only on schooling and other pre-employment observables, not on any output realizations. Recall from Section 3A that only 1169 of the 5998 individuals in our sample have valid zero-experience wage observations. We drop these observations from the analysis. Since zero experience in the dataset is $t = 1$ in the model, one year of experience in the dataset is $t = 2$ in the model, and the variance of year-one residuals in the dataset is the parameter $\text{var}(\hat{\epsilon}_{i2})$ in the model. Incorporating classical measurement error then yields (3.25) and (3.26).

We also need to derive the covariance structure of the first-difference of wage residuals, $\hat{\epsilon}_{i,t+1} - \hat{\epsilon}_{it}$. Naturally, z_i will disappear from this first-difference. Since $\hat{\epsilon}_{i1} \equiv 0$ in the model, $\text{var}(\hat{\epsilon}_{i2} - \hat{\epsilon}_{i1}) = \text{var}(\hat{\epsilon}_{i2})$, which will again be taken as parametric. Using the facts that $\hat{\epsilon}_{i,t+1} = \rho(\hat{\epsilon}_{it} + \lambda_{it})$ and that (in steady-state) $\text{var}(\hat{\epsilon}_{i2}) = \rho^2 \text{var}(\lambda_i)$, we have

$$\text{var}(\hat{\epsilon}_{i,t+1} - \hat{\epsilon}_{it}) = \left\{ 1 + (1-\rho)^2[1 + \rho^2 + \dots + \rho^{2(t-2)}] \right\} \text{var}(\hat{\epsilon}_{i2}) \quad \text{for } t \geq 2,$$

$$\text{cov}(\hat{\epsilon}_{i,t+1} - \hat{\epsilon}_{it}, \hat{\epsilon}_{i,t+2} - \hat{\epsilon}_{i,t+1}) =$$

$$\left\{ -(1-\rho) + \rho(1-\rho)^2[1 + \rho^2 + \dots + \rho^{2(t-2)}] \right\} \text{var}(\hat{\epsilon}_{i2}), \quad \text{and}$$

$$\text{cov}(\hat{\epsilon}_{i,t+1} - \hat{\epsilon}_{it}, \hat{\epsilon}_{i,t+s} - \hat{\epsilon}_{i,t+s-1}) = \rho^{s-2} \text{cov}(\hat{\epsilon}_{i,t+1} - \hat{\epsilon}_{it}, \hat{\epsilon}_{i,t+2} - \hat{\epsilon}_{i,t+1})$$

for $s \geq 2$. Incorporating classical measurement error and making the notational change described above then yields (3.27), (3.28), and (3.29).

Table 1:
Average Age and Education by Entry Cohort and Sex

Entry Cohort	Overall		
	N	Average Age	Average Education
1979	673	19.6	12.1
1980	576	19.8	12.5
1981	638	20.1	12.5
1982	550	20.9	12.8
1983	607	21.4	13.0
1984	588	22.4	13.0
1985	478	23.3	13.4
1986	407	24.1	13.7
1987	309	25.0	13.6
1988	172	26.2	13.5
All	4998	21.7	12.9
Entry Cohort	Males		
	N	Average Age	Average Education
1979	339	19.4	11.8
1980	261	19.6	12.1
1981	340	20.0	12.2
1982	282	20.7	12.4
1983	309	21.2	12.7
1984	296	22.1	12.7
1985	233	23.3	13.5
1986	206	24.3	13.6
1987	139	25.2	13.8
1988	79	26.3	13.7
All	2484	21.5	12.7
Entry Cohort	Females		
	N	Average Age	Average Education
1979	334	19.7	12.5
1980	315	20.1	12.8
1981	298	20.2	12.8
1982	268	21.1	13.2
1983	298	21.6	13.3
1984	292	22.8	13.4
1985	245	23.3	13.2
1986	201	24.3	13.7
1987	170	25.2	13.5
1988	93	26.3	13.3
All	2514	21.9	13.1

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 (1979-1991)
by Experience

Exper- ience	N	Wage	Age	Educ- ation	Part- Time	Coll. Barg	Non- White	Female	Mar- ried	Marr.& Female
0	1169	4.96 (2.02)	22.2 (2.56)	12.5 (2.33)	0.0	.170	.441	.483	.198	.119
1	4589	5.33 (2.33)	22.7 (2.68)	12.9 (2.35)	.132	.165	.415	.506	.202	.124
2	4622	5.90 (2.62)	23.6 (2.67)	13.0 (2.37)	.087	.180	.418	.504	.270	.152
3	4623	6.30 (2.86)	24.6 (2.67)	13.0 (2.39)	.071	.186	.421	.504	.332	.188
4	4182	6.70 (3.13)	25.5 (2.51)	13.1 (2.37)	.090	.186	.413	.505	.387	.213
5	3749	7.01 (3.24)	26.2 (2.38)	13.1 (2.40)	.096	.190	.416	.497	.441	.233
6	3269	7.27 (3.48)	27.0 (2.25)	13.1 (2.36)	.094	.185	.407	.501	.479	.252
7	2740	7.53 (3.62)	27.6 (2.10)	13.1 (2.31)	.093	.191	.397	.490	.510	.262
8	2170	7.77 (3.61)	28.3 (1.97)	13.0 (2.29)	.102	.191	.393	.486	.541	.270
9	1640	7.89 (3.74)	29.0 (1.90)	13.0 (2.29)	.098	.189	.387	.493	.573	.285
10	1230	7.77 (3.54)	29.8 (1.79)	12.9 (2.21)	.104	.196	.412	.496	.566	.281
11	759	7.88 (3.68)	30.6 (1.66)	12.9 (2.17)	.083	.202	.406	.511	.570	.283
Total:	34742	6.64 (3.21)	25.6 (3.27)	13.0 (2.35)	.092	.184	.412	.500	.386	.207

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
Empirical Covariance Matrix of Within-Worker Wage Residuals (Levels)
(standard error)
[cell size]

	0	1	2	3	4	5	6	7	8	9	10	11
0	2.58 (.389) [1169]											
1	1.82 (.379) [1081]	3.85 (.395) [4589]										
2	1.72 (.391) [1080]	2.45 (.335) [4292]	4.81 (.419) [4622]									
3	1.66 (.397) [1092]	2.45 (.346) [4282]	3.40 (.372) [4333]	5.67 (.443) [4623]								
4	1.57 (.400) [956]	2.32 (.354) [3875]	3.27 (.376) [3936]	4.11 (.399) [3945]	6.83 (.487) [4182]							
5	1.45 (.412) [834]	2.08 (.353) [3476]	2.95 (.382) [3514]	3.82 (.429) [3540]	4.74 (.437) [3479]	7.12 (.519) [3749]						
6	1.67 (.444) [705]	2.17 (.373) [3023]	3.03 (.411) [3077]	3.75 (.426) [3078]	4.73 (.481) [3016]	5.16 (.477) [3028]	8.42 (.562) [3269]					
7	1.58 (.527) [567]	2.01 (.396) [2529]	2.89 (.435) [2589]	3.53 (.450) [2588]	4.36 (.488) [2529]	4.91 (.498) [2520]	6.49 (.563) [2557]	9.14 (.600) [2740]				
8	1.40 (.515) [419]	1.84 (.405) [2006]	2.67 (.449) [2047]	3.11 (.451) [2051]	3.84 (.481) [2016]	4.35 (.504) [1991]	5.88 (.566) [2017]	6.59 (.580) [2029]	9.30 (.628) [2170]			
9	1.04 (.522) [316]	1.73 (.434) [1513]	2.54 (.455) [1543]	2.75 (.464) [1548]	3.70 (.510) [1521]	3.94 (.532) [1530]	5.38 (.586) [1519]	6.16 (.589) [1511]	7.32 (.634) [1495]	9.97 (.699) [1640]		
10	0.825 (.518) [240]	1.35 (.403) [1125]	2.33 (.495) [1154]	2.52 (.480) [1156]	3.22 (.549) [1125]	3.50 (.543) [1133]	4.64 (.599) [1151]	5.29 (.630) [1132]	5.89 (.632) [1105]	6.92 (.640) [1112]	8.97 (.714) [1230]	
11	0.566 (.467) [148]	1.24 (.466) [683]	2.22 (.579) [704]	2.17 (.558) [711]	3.12 (.619) [696]	2.93 (.592) [695]	4.41 (.664) [706]	4.68 (.656) [718]	5.53 (.763) [676]	6.33 (.757) [669]	6.72 (.766) [704]	10.0 (.853) [759]

Note: 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 4
Empirical Covariance Matrix of Within-Worker Wage Residuals (First-Differences)
(standard error)
[cell size]

	1	2	3	4	5	6	7	8	9	10	11
1	2.67 (.460) [1081]										
2	-1.15 (.457) [4292]	3.61 (.429) [4292]									
3	.128 (.343) [960]	-1.29 (.361) [4042]	3.51 (.435) [4333]								
4	-.105 (.370) [851]	.037 (.296) [3486]	-1.35 (.368) [3730]	4.03 (.478) [3945]							
5	-.105 (.331) [717]	-.163 (.290) [3082]	.056 (.287) [3140]	-1.66 (.421) [3306]	3.96 (.469) [3479]						
6	.042 (.354) [602]	.019 (.306) [2676]	-.134 (.306) [2731]	-.012 (.355) [2695]	-1.60 (.399) [2826]	4.73 (.549) [3028]					
7	-.041 (.392) [495]	-.139 (.301) [2263]	.057 (.306) [2312]	-.066 (.313) [2272]	.012 (.326) [2227]	-1.60 (.444) [2372]	4.66 (.554) [2557]				
8	.094 (.426) [376]	-.108 (.351) [1794]	-.059 (.333) [1830]	-.039 (.347) [1817]	-1.70 (.387) [1781]	-.084 (.459) [1778]	-2.01 (.527) [1910]	4.85 (.619) [2029]			
9	.024 (.454) [271]	.139 (.334) [1314]	-.168 (.347) [1348]	.092 (.366) [1330]	-1.42 (.388) [1327]	.115 (.352) [1332]	.031 (.382) [1329]	-1.70 (.505) [1407]	4.49 (.617) [1495]		
10	.291 (.392) [207]	.031 (.351) [970]	-.122 (.344) [998]	-.332 (.482) [981]	.139 (.446) [975]	-.225 (.410) [997]	-.222 (.393) [991]	-1.28 (.603) [1022]	-1.28 (.671) [1112]	4.22 (.671) [1112]	
11	.134 (.437) [130]	-.157 (.379) [594]	.171 (.393) [618]	.427 (.555) [612]	-.663 (.496) [599]	.210 (.489) [617]	-.201 (.497) [636]	.180 (.661) [617]	-1.38 (.654) [582]	-1.38 (.611) [634]	4.41 (.769) [704]

Note: Based on first differences of 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 Function
Unbalanced Sample

Variable	Mean (sd)	Estimate (se)
years of experience	5.18 (2.50)	.541 (.0741)
experience squared	33.1 (29.9)	-.0147 (.00228)
education	13.04 (2.35)	.630 (.0408)
education*experience	67.54 (35.01)	-.00498 (.00509)
AFQT residual	.237 (14.8)	.00747 (.00258)
AFQT residual *experience	1.89 (85.7)	.00209 (.00048)
Library Card Residual	-.0018 (.434)	-.0228 (.0875)
Library Card Resid *experience	-.0107 (2.48)	.0559 (.0164)
p-value of Wald test AFQT and Lib. Card interactions with exp	1.19x10-8	
R ² (GLS estimates)	.282	
number of workers	4970	
number of wage obs.	28984	

Note: The first observation for each individual is omitted from the analysis. The dependent variable is average hourly earnings on the most recent job deflated by the 1982-84=100 CPI (mean=6.92, s.d.=3.30). All regressions also include a constant, age at entry, a dummy variable for part-time, the interaction of part-time with education, year dummies, the interaction of year dummies with education, and dummy variables for collective bargaining status, race, sex, marital status, and sex*marital status. The base year is 1991. The R-squared is computed applying the GLS estimates to the untransformed data.

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

Parameter	Associated Variance	Levels	First Differences
Variance of initial unmeasured expected ability:			
α_0	σ_1^2	2.04 (.0891)	---
Variance of measurement error:			
α_1	σ_ϕ^2	1.57 (.0582)	1.44 (.0570)
Variances of wage innovations each period:			
α_2	$\sigma_{\mu 1}^2$	.863 (.0799)	.767 (.144)
α_3	$\sigma_{\mu 2}^2$	.599 (.0822)	.757 (.101)
α_4	$\sigma_{\mu 3}^2$	.875 (.0957)	.978 (.135)
α_5	$\sigma_{\mu 4}^2$	.552 (.0942)	.648 (.107)
α_6	$\sigma_{\mu 5}^2$	1.32 (.141)	1.32 (.199)
α_7	$\sigma_{\mu 6}^2$	.826 (.146)	1.22 (.175)
α_8	$\sigma_{\mu 7}^2$	.764 (.161)	.804 (.156)
α_9	$\sigma_{\mu 8}^2$	.657 (.200)	1.27 (.191)
α_{10}	$\sigma_{\mu 9}^2$	.289 (.265)	.807 (.192)
α_{11}	$\sigma_{\mu 10}^2$	.889 (.476)	.641 (.400)
χ^2 statistic, structural test:		157.0	79.0
degrees of freedom:		54	44
p-value of test statistic:		1×10^{-11}	.0009
number of workers		4998	4998

(See Note Next Page)

(Table 6 Continued)

Note: The optimal minimum distance estimator is the GLS regression of the unique elements of the covariance matrices in tables 3 and 4 on the variables implied by 3.11 and 3.11'. The covariance matrices of the covariance elements in tables 3 and 4 are 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 or 3.11' place on the unrestricted covariance matrices in tables 3 and 4.

Table 7:
Optimal Minimum Distance Estimation of Covariance Structure
AR1 Worker Quality Overlaid with Classical Measurement Error
NLSY Unbalanced Panels
(standard errors in parentheses)

Parameter	Levels	First Differences
AR1 parameter in worker quality		
ρ	.908 (.0098)	.933 (.0123)
Variance of market-observed unmeasured fixed worker characteristics:		
σ_z^2	1.12 (.0673)	---
Steady-state variance of ability:		
σ_1^2	1.30 (.0469)	1.15 (.0683)
Variance of measurement error:		
σ_ϕ^2	1.36 (.0451)	1.29 (.0586)
χ^2 statistic, structural test:	91.5	64.4
degrees of freedom:	62	52
p-value of test statistic:	0.009	.137
number of workers	4998	4998

Note: The optimal minimum distance estimator uses the unique elements of the covariance matrices in tables 3 and 4 as moments to be fit. The coefficients were computed using the Newton-Raphson algorithm to minimize the criterion function in equation 3.13 implied by the model. The covariance matrices of the covariance elements in tables 3 and 4 are 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 or 3.11' places on the unrestricted covariance matrices in tables 3 and 4.