The Speed of Employer Learning

Fabian Lange, Yale University

The employer-learning literature finds support for statistical discrimination on the basis of schooling. How economically relevant statistical discrimination is depends on how fast employers learn about workers’ productive types. This article is the first to estimate the speed of employer learning. Employers learn quickly. Initial expectation errors decline by 50% within 3 years. This estimate places an upper bound on the contribution of signaling. This bound varies with the speed of employer learning and with discount rate. For a wide range of parameter values, the contribution of signaling to the gains from schooling is less than 25%.

I. Introduction

The relation between earnings and experience is strong, especially at the start of individuals’ careers. Earnings increase by about 80% during the first 10 years of the working life. Clearly, experience is rewarded in the labor market; the correlation between earnings and experience is large and explains a substantial fraction of the observed variation in labor earnings.

The returns to experience vary with both schooling and ability, as mea-
sured by cognitive test scores. In two landmark articles Farber and Gibbons (1996) and Altonji and Pierret (2001) demonstrate that more-able individuals enjoy much faster wage growth with experience, whereas schooling lowers the experience gradient. Furthermore, these authors introduced a novel assumption that allows researchers to interpret the variation in experience gradients with schooling and cognitive test scores as a test of statistical discrimination and employer learning. This assumption is that some correlates of individuals’ productivity are available in the analyzed data that are not available to employers. It has become standard to assume that the Armed Forces Qualification Test (AFQT) score is such a correlate. The AFQT has been administered to the respondents of the National Longitudinal Survey of Youth (NLSY), which in turn has become the most important data source in the empirical employer-learning literature.

This literature argues that if the AFQT is not available to firms, then it will initially not be priced into wages. At the beginning of the working life, the partial correlation between the AFQT and wages is therefore low. The longer individuals participate in the labor market, the more information about their true productive characteristics becomes available, and wages increasingly reflect productivity. This implies that the association of wages with the AFQT increases over the life cycle. At the same time, the returns to schooling decline with experience since employers cease to rely on schooling to predict productivity. Altonji and Pierret (2001) show that the variation in the experience gradients with the AFQT and schooling is consistent with the predictions of the model of statistical discrimination and employer learning.

In the employer-learning model, employers observe workers’ performance on the job and thus learn about workers’ unobserved ability. As they learn, they rely less on schooling and other easily observed variables to predict workers’ productive characteristics. The faster employers learn, the shorter the time period during which firms need to rely on schooling to predict productivity and the smaller the contribution of statistical discrimination to the lifetime returns to schooling. The economic relevance

---

1 Farber and Gibbons (1996) and Altonji and Pierret (2001) are not alone in reporting such patterns in the variation of experience gradients with schooling and ability. Hause (1972) presents early results consistent with their findings. Galindo-Rueda (2003) reports similar findings using UK data. Bauer and Haiksen-DeNew (2001) find the same patterns among German blue-collar, but not white-collar, workers. From developing countries there is supporting evidence from rural India and Pakistan (Foster and Rosenzweig 1993) and from Ghana (Strobl 2003).
of statistical discrimination therefore hinges on the question of how fast employers learn about workers’ productive types.¹

In this article I make two main contributions to this literature. I first show how to estimate the speed of employer learning and provide an estimate of the speed of employer learning. Employers learn fast. It takes on average 3 years for any initial expectation error on the part of employers about workers’ productivity to decline by approximately 50%.

The second main contribution is to provide a new answer to an old question: how important is the signaling motive for schooling choices relative to the productivity-augmenting effects of schooling?² It has proven difficult to find satisfactory answers to this question. The difficulties stem from the fact that the core behavior of economic agents in both job-market-signaling and human-capital models is similar. On one hand, individuals faced with the equilibrium wage-schooling relation choose schooling to maximize lifetime resources. On the other hand, firms hire workers as long as wages equal expected productivity. Workers and firms behave similarly in both models, making it difficult to separate the importance of either in determining schooling choice.

Progress in determining the contribution of signaling is therefore unlikely without making strong assumptions. Nevertheless, parsimony in the introduction of auxiliary assumptions is desirable. I therefore show that the employer-learning model alone does not allow identifying the contribution of signaling. However, progress is possible under the additional assumption that workers choose schooling in order to maximize the present value of lifetime earnings.³ This assumption, together with the estimated speed of employer learning, allows identifying an upper bound on the contribution of signaling to the gains of schooling.

This bound suggests that the signaling value of schooling does not exceed one-quarter of the total economic value of an additional year of schooling. The remainder is accounted for by productivity-augmenting effects of schooling. The size of the bound on signaling depends crucially on the nominal interest rate with which workers discount labor earnings and on the speed with which employers learn about individuals’ productive characteristics. I therefore investigate how sensitive the derived bound is to variation in these parameters. My preferred specification re-

---

¹ Altonji and Pierret (1997) recognize that the speed of employer learning is crucial for the economic significance of statistical discrimination. I build on their contribution and estimate the speed of employer learning.

² The seminal work in the signaling literature is, of course, Spence (1973).

³ It is possible to bound the contribution of signaling as long as an estimate of the costs of schooling is available. The assumption that agents maximize the present value of earnings allows me to estimate the costs of schooling using the opportunity costs of schooling. It is possible to implement the method using alternative estimates of the costs of schooling.
sults in a bound on signaling of about 10% of the total gain from an additional year of schooling. However, if the discount rate on labor earnings is low and employers learn slowly about individuals’ productive characteristics, then the bound can be as high as 45% of the gains from schooling.

The article is structured as follows: Section II informally introduces the model of statistical discrimination and employer learning as developed by Farber and Gibbons (1996) and Altonji and Pierret (2001). It examines in some detail the assumption that the AFQT score available in the data is not observed by employers. In Section III, I present the employer-learning model more formally, I show how to estimate the speed with which employers learn about workers’ productivity, and I estimate the speed of employer learning using data from the NLSY. Section IV discusses how the estimated speed of learning can be used to quantify the importance of signaling motives in the schooling decision. I begin by formally defining the parameter of interest. This parameter is the contribution of signaling to the economic gains from schooling. I demonstrate that this parameter is not identified using the model of employer learning alone. I then show how the additional assumption that individuals maximize the present value of earnings allows bounding the parameter of interest. Section V estimates this bound on the contribution of signaling. This bound depends crucially on the discount rate on future labor earnings as well as on the speed of employer learning. Section V therefore examines how sensitive the bound on the parameter of interest is to variation in these crucial parameters. Section VI concludes.

II. The Empirical Employer-Learning Literature

A. A Test for Statistical Discrimination and Employer Learning

Much of this article builds on Altonji and Pierret (2001), who in turn rely heavily on Farber and Gibbons (1996). Altonji and Pierret devise a test of the employer learning and statistical discrimination (EL-SD) hypothesis. In this section, I informally review the test Altonji and Pierret (2001) propose, their empirical findings, and the crucial assumption underlying their approach. Section III contains a more formal treatment.

The EL-SD hypothesis posits that employers cannot observe job applicants’ productive characteristics directly. Employers instead use easily observable correlates of productivity (e.g., schooling) to predict applicants’ productivity. Over time—as workers spend time in the labor force—new information on workers becomes available: employers learn.

Altonji and Pierret (2001) explore the implications of EL-SD under the assumption that all employers share the same information and that a spot market for labor services exists. Competition in labor markets ensures that wages reflect expected productivity of workers conditional on com-
monly observed information. Altonji and Pierret (2001) assume that the data contain a measure of productivity (denoted \(z\)) not available to employers. In their empirical work they rely mainly on the AFQT for this purpose.

They then derive and test the implications of EL-SD for life-cycle earnings profiles. The EL-SD implies that the linear regression coefficient on unobserved ability \(z\) is increasing with experience. Wages increasingly reflect workers’ true productivity, and therefore \(z\) is increasingly priced by firms. Thus, the return to the AFQT is expected to increase with experience, a finding confirmed in the data. Altonji and Pierret (2001) also derive implications for the observed coefficients of schooling in log earnings regressions. Schooling raises earnings early in the career because schooling raises productivity and because of statistical discrimination on the basis of schooling. The contribution of statistical discrimination to the returns to schooling declines as workers age, and firms increasingly rely on observed productivity measures rather than schooling to predict workers’ ability. This generates the second implication Altonji and Pierret (2001) test: allowing the AFQT to interact with experience lowers the schooling coefficients for experienced workers more than for less-experienced workers. Again they find support for this implication of EL-SD in the data. I will briefly review their empirical findings next.


Altonji and Pierret (2001) estimate a log earnings equation that allows for a linear interaction between years of schooling and the AFQT with experience:

\[
\log w_i = \beta_s + \beta_s s_i + \beta_z z_i + \beta_{s,z}(s_i \times x_i) + \beta_{s,x}(z_i \times x_i) + f(x_i) + \beta_i \Phi_i + \epsilon_i.
\]

Log wages \(w_i\) of individual \(i\) depend on schooling \(s_i\), the AFQT (standardized by birth cohort) \(z_i\), experience \(x_i\), and controls \(\Phi_i\). Altonji and Pierret (2001) emphasize how the relation between log wages and both the AFQT and completed schooling changes with experience. Table 1 reproduces their main results based on data from the NLSY for the years 1979–92 in columns 1 and 2. The remainder of the table reports analogous results for the longer sample (1979–98) available from the NLSY today and for different subgroups of the population.\(^5\)

\(^5\) Table 1, cols. 1 and 2, reproduce the main empirical results from Altonji and Pierret (2001), table 1. Columns 3, 4, 7, and 8 show that the result holds also for the samples used in this article. Columns 5 and 6 contain the results from a median regression after reinserting the zeros into the sample. The sample used in this study differs from that used by Altonji and Pierret in that I restrict myself to the main NLSY sample and because I use the 1979–98 waves. Their study exploits
### Table 1

The Effects of AFQT and Schooling in a Linear Specification

<table>
<thead>
<tr>
<th>Model:</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
<th>(7)</th>
<th>(8)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Education</td>
<td>0.0586**</td>
<td>0.0829**</td>
<td>0.0678**</td>
<td>0.0824**</td>
<td>0.0887**</td>
<td>0.1024**</td>
<td>0.0846**</td>
<td>0.0846**</td>
</tr>
<tr>
<td>Black</td>
<td>-0.1565**</td>
<td>-0.1553**</td>
<td>-0.0434**</td>
<td>-0.0427**</td>
<td>-0.0434**</td>
<td>-0.0415**</td>
<td>-0.2346**</td>
<td>-0.2342**</td>
</tr>
<tr>
<td>Female</td>
<td>-0.0438</td>
<td>-0.0438</td>
<td>-0.0092</td>
<td>-0.0092</td>
<td>-0.0092</td>
<td>-0.0092</td>
<td>-0.0092</td>
<td>-0.0092</td>
</tr>
<tr>
<td>Standardized AFQT</td>
<td>0.0834**</td>
<td>-0.0060</td>
<td>0.1010**</td>
<td>0.0490**</td>
<td>0.1303**</td>
<td>0.0686**</td>
<td>0.1124**</td>
<td>0.0618**</td>
</tr>
<tr>
<td>AFQT × experience/10</td>
<td>-0.0032</td>
<td>-0.0347**</td>
<td>-0.0030</td>
<td>-0.0199**</td>
<td>-0.0147**</td>
<td>-0.0311**</td>
<td>-0.0027</td>
<td>-0.0165**</td>
</tr>
<tr>
<td>R²</td>
<td>0.2861</td>
<td>0.2870</td>
<td>0.2557</td>
<td>0.2588</td>
<td>0.1528</td>
<td>0.1538</td>
<td>0.2988</td>
<td>0.3004</td>
</tr>
</tbody>
</table>

#### Sample

<table>
<thead>
<tr>
<th>No. of individuals</th>
<th>2,978</th>
<th>2,277</th>
<th>2,290</th>
<th>2,290</th>
<th>5,336</th>
<th>5,336</th>
<th>5,336</th>
<th>5,336</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of observations</td>
<td>21,058</td>
<td>24,410</td>
<td>25,778</td>
<td>25,778</td>
<td>55,181</td>
<td>55,181</td>
<td>55,181</td>
<td>55,181</td>
</tr>
</tbody>
</table>

Note.—The coefficients of regressions of log wages on schooling and Armed Forces Qualification Test (AFQT) scores, linearly interacted with the experience coefficient as well as demographic controls, are shown. Columns 1 and 2 report the results reported by Altonji and Pierret (2001) that motivate this story. The specification in Altonji and Pierret (2001) includes a cubic in experience. All specifications examined in this article allow for a full set of experience dummies. Columns 3 and 4 show that the results are found for the sample of white males from the main (nationally representative) sample of the NLSY for the period 1979–98. Columns 5 and 6 investigate whether the results are robust to reinserting the zeros into the sample and performing a median regression. In cols. 5 and 6, I report pseudo-\(R^2\)'s. Columns 7 and 8 refer to the results obtained on the full sample for the time period 1979–98. For a description of the data, see the appendix. In cols. 1–4, 7, and 8, the standard errors (in parentheses) are White/Huber standard errors accounting for potential correlation at the individual level.

* Statistical significance at the 95% level.

** Statistical significance at the 99% level.
Altonji and Pierret first estimate equation (1) while imposing the restriction \( \beta_{\varepsilon x} = 0 \). This restriction ensures that the effect of ability on log earnings is held constant over the life cycle. The returns to schooling estimated while imposing this restriction are roughly constant with experience. The coefficient in row 5 of table 1, column 1, implies that the returns to schooling decline by less than one-half of a percentage point over the first 10 years of a worker’s career. This parallel structure of log earnings often leads analysts to specify earnings equations to be separable in schooling and experience. The most famous example is, of course, the specification favored by Mincer (1958).6

Altonji and Pierret (2001) then allow the AFQT to interact with experience \( (\beta_{\varepsilon x} \neq 0) \). Two empirical findings are especially noteworthy and form the basis of their study. First, AFQT scores are increasingly associated with earnings. The return to 1 standard deviation of the AFQT is 6–8 percentage points larger with 10 years of experience than at the beginning of a worker’s career. Second, controlling for the interaction of the AFQT with experience reduces the return to schooling by about 2 percentage points during the first decade of labor-market participation. These findings, reported in Altonji and Pierret (2001) and replicated in table 1, are consistent with the implications of EL-SD sketched above; the data therefore fail to reject the hypothesis of EL-SD.

C. Is the AFQT Unobserved by Employers?

Crucial to implementing the test devised by Altonji and Pierret (2001) is that the researcher has access to a measure of applicants’ productivity that is unobserved by firms. This is an unusual assumption. If the AFQT indeed predicts individual productivity, then why don’t employers administer the AFQT or a similar test to their job applicants?

This study analyzes data from the NLSY. The history of the AFQT and the NLSY itself supports the notion that the benefits of cognitive testing were not apparent to employers at the time the respondents to the NLSY entered the labor market in the early eighties. The AFQT is a composite score of a battery of tests known as the Armed Services Vocational Aptitude Battery (ASVAB). This test battery was used throughout the military for the first time in 1976 (following the move to an all-volunteer army in 1973). By 1980 the Department of Defense had to admit to Congress that an error in scoring the ASVAB meant that about 250,000 recruits were mistakenly enlisted. Subsequent investigations

both the main and the supplemental sample but restricts itself to the 1979–92 waves. These changes in the sample selection do not affect the basic findings, as is evident when we compare cols. 1 and 2 with cols. 7 and 8. The sample selection criteria and variables used in this study are discussed in more detail in the appendix.

revealed that no link between ASVAB scores and job performance had yet been empirically established. The Department of Defense reacted to this uncertainty about the ASVAB and the AFQT by launching two major research projects. First, the AFQT was administered to the NLSY sample. Second, a decade-long project (the Joint-Service Job Performance Measurement/Enlistment Standards Project [JPM]) was put under way with the explicit goal of examining the link between the testing procedures used by the military and subsequent job performance. Only through the completion of this project in the early nineties has the empirical link between the AFQT and job performance been established (Wigdor and Green 1991). Earlier research (Ghiselli 1967) did find support for a link between cognitive-ability tests and job performance, but it is probably fair to say that substantial uncertainty still existed around this question in the late seventies and early eighties. It is only through the JPM and other research in the 1980s that the link between aptitude tests and job performance has been more firmly established (see Terpstra and Rozell [1993] for a survey of this literature).

There is also direct evidence reported in the personnel literature that cognitive-ability tests are only employed by a small fraction of firms, even though such tests have been linked to subsequent job performance (Schuler and Jackson 1987; Terpstra and Rozell 1997). Terpstra and Rozell (1997) report findings from the Bureau of National Affairs in 1983 that only 20% of surveyed employers used cognitive-ability tests in the selection of workers. They paint a picture of slow implementation of sophisticated hiring procedures (including cognitive testing) due to budget and time constraints placed upon managers. Survey evidence from human resource managers documents both uncertainty about the predictive performance of ability tests and concerns about legal repercussions. Legal concerns are cited by about a third of managers. These are founded in a belief that the use of cognitive-ability tests might lead to charges of discrimination.

Another reason that employers might be reluctant to administer tests such as the AFQT lies in the high turnover of employees. If employers can expect to lose young employees within a few months, then it might be preferable to forgo testing on young employees. There is ample evidence that worker turnover is indeed rapid during the early stages of individuals’ careers. Topel and Ward (1992), for instance, report that two-thirds of all employment spells of young workers end within the first year.

There are thus a number of reasons not to dismiss out of hand the assumption that employers do not observe the AFQT score. The economic returns to administering cognitive-ability tests are limited by the fast turnover during the early career and the speed of employer learning once workers enter the labor market. The perceived uncertainty about
The Speed of Employer Learning

the predictive power of cognitive-ability test scores is another important reason that they are not more widely used in screening applicants. This uncertainty was probably even higher in 1979 (the starting date of the NLSY) since much of the evidence on the predictive potential of cognitive-ability tests has only become available during the eighties and nineties. Furthermore, administering cognitive-ability tests could require training human resource managers and reorganizing the human resource departments of employers. The costs of incorporating testing into the selection procedures might therefore be significantly larger than simply the time costs required in administering and grading the tests. Substantial legal concerns may limit the use of cognitive testing in employee selection. All of these are reasons that the AFQT administered in 1980 to the NLSY cohort represents a measure of productivity available to researchers but plausibly not to employers.

III. The Speed of Employer Learning

A. The Employer-Learning Model

Farber and Gibbons (1996) formalized the employer-learning model in a way that has since become standard in the empirical employer-learning literature. The model specifies individual \( i \)'s log productivity \( \chi_{i,x} \) at experience \( x \) to consist of (i) a linear function \( \tilde{\chi}(s, q, \eta, z) \) of various variables \( s, q, \eta, z \) describing the information available to employers and researchers and (ii) a polynomial \( \tilde{H}(x) \) in experience \( x \):

\[
\chi_{i,x} = \tilde{\chi}(s, q, \eta, z) + \tilde{H}(x).
\]

(2)

The variables \( s \) capture the information available to both employers and researchers. Schooling is an example for such a variable. The variables \( q \) describe information available to employers but not contained in the data. An example of these variables is information obtained through job interviews. The employer-learning literature also assumes that there are correlates of productivity \( z \) that are available to researchers but not to employers. Much of the literature assumes that the AFQT score is such a correlate of productivity, and I assume that as well. Finally, the model denotes correlates of productivity neither available to employers nor in the data as \( \eta \). Log productivity \( \chi_{i,x} \) of individual \( i \) at experience level \( x \) can then be expressed as

\[
\chi_{i,x} = r s_i + \alpha q_i + \lambda z_i + \eta_i + \tilde{H}(x).
\]

(3)

The subscript \( i \) is understood and will be suppressed from now on. The function \( \tilde{H}(x) \) describes the structural relation between log productivity and experience \( x \), maybe due to a process of investment over the life cycle. The employer-learning literature focuses on the variation of the experience gradient of log earnings with schooling and ability. This variation in the
experience gradient is interpreted as the outcome of an employer-learning process about $\tilde{x}(s, q, \eta, z)$. The assumption that $\tilde{H}(x)$ does not depend on either education or the ability measure $z$ is crucial for this interpretation. Making this assumption allows us to concentrate on the problem faced by employers of predicting based on the variables $(s, q)$ and additional measures of $\tilde{x}$ that become available as workers spend time in the labor market.

To summarize, log productivity depends linearly on a set of characteristics $(s, q, z, \eta)$ and a polynomial in experience. Employers observe some of these characteristics $(s, q, z, \eta)$ initially unobserved. The empirical literature on employer learning examines the signal-extraction problem faced by employers, how this problem changes as more information becomes available, and what the implications are for experience profiles.

Assume that $(s, q, z, \eta)$ are jointly normally distributed. An implication is that the expectation of $z$ conditional on the information available to firms is linear in $(s, q)$:

$$z = E[z|s, q] + v = \gamma_1 q + \gamma_2 s + v;$$

$$\eta = E[\eta|s, q] + e = \alpha_2 s + e. \tag{5}$$

Equations (3)–(5) allow me to express log productivity as a linear function of the information available to employers at time $x = 0$:

$$\tilde{x} = (r + \lambda \gamma_2 + \alpha_s)s + (\alpha_1 + \lambda \gamma_1)q + (\lambda v + e) + \tilde{H}(x)$$

$$E[\tilde{x}|s, q] + (\lambda v + e) + \tilde{H}(x). \tag{6}$$

The process of employer learning is modeled by assuming that after each period the individual spends in the labor market, a noisy measurement $\tilde{y}_t$ of $\tilde{x}$ becomes available to all employers:

$$\tilde{y}_t = \tilde{x} + \epsilon_t. \tag{7}$$

The noise $\epsilon_t$ is uncorrelated with all other variables in the model. Assume that $\epsilon_t$ is independently, identically, and normally distributed with variance $\sigma^2$. At experience $x$, an $x$-dimensional vector of measurements, $\tilde{y}^* = (y_{t_1}, y_{t_2}, \ldots, y_{t_{x-1}})$, has become available to employers.

Because of the normality assumptions, the process of updating employer expectations about $\tilde{x}$ has a very simple structure. In each period $\tau$ the market observes a measure of individual log productivity, $y_t = \tilde{x} + \epsilon_t$. The number of additional measures available in the market is equal

$^7$ A normalization of the coefficient vector $\alpha_t$ allows suppressing $q$ in eq. (4).
The Speed of Employer Learning

to the experience of individuals. At experience $x$ the posterior distribution is normal with mean $\mu_x$ and precision $\rho_x = 1/\sigma_x^2$, where $(\mu_x, 1/\sigma_x)$ are

$$\mu_x = (1 - \theta_x)E[\chi|s, q] + \theta_x \left( \frac{1}{x} \sum_{j=2}^{x-1} y_j \right)$$  \hspace{1cm} (8)

and

$$\frac{1}{\sigma_x^2} = \frac{1}{\sigma_0^2} + \frac{x}{\sigma_x^2}. \hspace{1cm} (9)$$

Here $\sigma_x^2$ is the variance of $\chi$ conditional on $(s, q)$; it is the variance of the initial expectation error $(\lambda v + e)$. The regression coefficients $\theta_x$ at each experience level are given by

$$\theta_x = \frac{xK_1}{1 + (x - 1)K_1}, \hspace{1cm} (10)$$

where

$$K_1 = \frac{\sigma_0^2}{\sigma_0^2 + \sigma_x^2} \hspace{1cm} (11)$$

is a parameter that reflects the relative information content of initial information $(s, q)$ and subsequent measurements $y$. The regression coefficients $\theta_x$ and the parameter $K_1$ lie in the interval $[0, 1]$, and $\theta_x$ converges to one as $x$ increases. The speed at which it increases depends positively on $K_1$.

The smaller $K_1$ is, the less informative the new measures $y_i$ are, relative to the initial information available to employers. Therefore, the weight employers place on new information increases in $K_1$. We can say that the speed at which employers learn depends positively on $K_1$, and in the remainder of the article I will refer to the parameter $K_1$ as the speed of employer learning. The first major contribution of this article is to estimate $K_1$ using wage data from the NLSY.

The common employer-learning model by Farber and Gibbons (1996) assumes that all employers have access to the same information, labor markets are competitive, and a spot market for labor services exists. Therefore, wages equal the expected productivity conditional on the available information:

$$W(s, q, y^*) = E[\exp(\chi)|s, q, y^*]. \hspace{1cm} (12)$$

The distribution of $\chi$ conditional on $(s, q, y^*)$ is normal. Therefore, the conditional expected value of $\exp(\chi)$ at experience $x$ is $\exp(E[\chi|s, q, y^*] +
Note at this point that the expectation error at period $x$ is independent of $(s, z, q, \eta)$ as well as of the realizations of $y^*$. This implies that $\frac{1}{2}\sigma_x^2$ is constant with respect to $(s, z, q, \eta)$. Therefore, we can define $H(x) = \tilde{H}(x) + \frac{1}{2}\sigma_x^2$. Taking logs and using equation (8) results in the following expression for log wages:

$$\omega(s, q, y^*) = (1 - \theta_x)E[\tilde{x} | s, q] + \theta_x \left( \frac{1}{x} \sum_{x=0}^{\infty} y_i \right) + H(x).$$  \hspace{1cm} (13)

Equation (13) describes the structural relation between log wages and (i) initial information available to employers and (ii) measures of log productivity $y^*$ that become available over the life cycle of the individual.

**B. Estimating the Speed of Learning $K_t$**

Equation (13) relates log wages to the information $(s, q, y^*)$ available to employers. The empirical quantities available in the data, however, are not $(s, q, y^*)$ but $(s, z, x)$. We therefore need to describe the relation between log earnings and $(s, z, x)$ rather than $(s, q, y^*)$, which in turn requires us to determine the relation between $(q, \eta)$ and $(s, z)$.

Without loss of generality we can define the linear projections of $(q, \eta)$ on $(s, z)$:

$$q = \gamma_s s + \gamma_z z + \eta_1;$$ \hspace{1cm} (14)

$$\eta = \gamma_s s + \gamma_\eta z + \eta_2.$$ \hspace{1cm} (15)

These definitions allow us to define the empirical objects of interest: the linear projections of log wages on schooling $s$ and ability $z$ at different experience levels $x$. The following equation shows the linear projection of log wages conditional on the observed data $(s, z, x)$:

$$E^\omega[w(s, q, y^*) | s, z, x] = E^\omega[(1 - \theta_x)E[\tilde{x} | s, q]$$

$$+ \theta_x \left( \frac{1}{x} \sum_{x=0}^{\infty} y_i \right) + H(x) | s, z, x].$$  \hspace{1cm} (16)

The independence assumption on $\varepsilon_x$ allows me to write

$$E^\omega[w(s, q, y^*) | s, z, x] = (1 - \theta_x)E^\omega[E[\tilde{x} | s, q] | s, z]$$

$$+ \theta_x E^\omega[\tilde{x} | s, z] + H(x).$$  \hspace{1cm} (17)

The importance of equation (17) for the purpose of estimating the speed

---

\[ I use the formula for the expectation of a log-normal random variable Q. If Q is distributed normal with mean $\mu$ and variance $\sigma^2$, then $E[\exp(Q)] = \exp(\mu + \frac{1}{2}\sigma^2)$.\]

\[ Here I denote linear projections of A on B as $E^\omega[A | B]$.\]
of learning is that it specifies wages as a weighted average of two components that depend on \((s, z)\) as well as the experience polynomial \(H(x)\), which does not depend on \((s, z)\). We have set up the problem such that each of the two terms \(E^s[E[x|s,q]|s,z]\) and \(E^s[E[x|s,z]]\) is a linear function of schooling \(s\) and ability \(z\). The weights \(\theta\) in equation (17) depend on experience \(x\) and \(v_x\). These weights equal zero at \(x = 0\) and increase to one as \(x \to \infty\). Thus, at \(x = 0\) the regression coefficients on \((s, z)\) are entirely determined by \(E^s[E[x|s,q]|s,z]\). As \(x \to \infty\), however, the regression coefficients on \((s, z)\) are increasingly reflecting \(E^s[E[x|s,z]]\) to some limit value determined by \(E^s[E[x|s,q]|s,z]\). The speed with which this convergence process unfolds can be estimated and ultimately delivers our estimated speed of employer learning.

Before estimating this speed of employer learning, however, consider the two components of the weighted average in equation (17) in more detail. Let us begin with \(E^s[E[x|s,q]|s,z]\). The weight \(\theta_s\) equals zero at \(x = 0\). Therefore, this term describes the relation between log wages and \((s, z)\) at the beginning of individuals' careers. Equations (6) and (14) imply

\[
E^s[E[x|s,q]|s,z] = \{r + \alpha \gamma_3 + (\alpha_2 + \lambda \gamma_2 + \lambda \gamma_3)\}s + \{(\alpha_1 + \lambda \gamma_1)\}z.
\]

The effect of schooling on wages at the beginning of an individual's career is composed of three terms. The first \([A]\) captures the productivity effect of schooling. The second term \([B]\) reflects the fact that schooling covaries with \(q\). This term \([B]\) can be thought of as analogous to the ability bias term introduced into the study of schooling returns by Griliches (1977). It reflects possible correlation between schooling and other ability measures directly observed by employers. Term \([C]\) is different in nature. Employers are aware that schooling covaries with information \((z, q)\) that is unknown to them. They therefore use schooling to predict the unknown components of productivity. This gives rise to the term \([C]\). It captures the partial correlation between \(s\) and \((z, q)\) that arises due to statistical discrimination.

The dependence of log wages on \(z\) at \(x = 0\) is captured by component \([D]\) in equation (18). This term is analogous to the term \([C]\) in the schooling coefficient explained above. Firms predict productivity on the basis of \(q\), which in turn correlates with the ability measure \(z\). Firms at \(x = 0\) have no other information but \((s, q)\) on individual productivity, and therefore no other information is priced. This means that \([D]\) is the only direct link between \(z\) and log wages at \(x = 0\).

Now consider \(E^s[E[x|s,z]]\), the second term in equation (17). When
x \to \infty$, then $\theta_x \to 1$ and thus equation (17) reduces to $E^x[\hat{x}|s,z] + H(x)$. Therefore, $E^x[\hat{x}|s,z]$ expresses the relation between log wages and $(s,z)$ after all information has been revealed to employers. Rewrite $E^x[\hat{x}|s,z]$ by inserting the linear projections (eqq. [14] and [15]) in equation (3) as

$$E^x[\hat{x}|s,z] = \{r + (\alpha_x \gamma_1 + \gamma_x)\}s + \{\lambda + (\alpha_x \gamma_4 + \gamma_x)\}z.$$  

(19)

The two components $[E]$ and $[G]$ simply reflect the direct productivity effects of schooling and ability, respectively. The terms $[F]$ and $[H]$ are nonzero, and therefore the long-run coefficients do not converge to the true productivity effect of schooling or ability. Term $[F]$ is equivalent to the Griliches ability bias in schooling after true productivity has been learned by firms. This bias arises because schooling correlates with observed (by firms, not by researchers) productivity measures. It differs from $[B]$ in equation (18) because as $x$ becomes large, firms have learned about the component of productivity $\eta$. The term $[H]$ represents the bias in the returns to ability analogous to $[F]$.

We have constructed the employer-learning model in such a form that each of the two components of log wages in equation (17) is linear in schooling and ability. Rearranging terms results in

$$E^x[w(s,q,y^*)|s,z,x] = ((1 - \theta_x)b_{s0} + \theta_x b_{x0})s + ((1 - \theta_x)b_{s1} + \theta_x b_{x1})z + H(x).$$  

(20)

The weights $\theta_x = xK_x/[1 + (x - 1)K_x]$ are functions of $K_x$ and experience $x$ only. Thus, the linear projections (eq. [20]) depend only on $K_x$ and the following four parameters:

$$b_{s0} = r + \alpha_x \gamma_1 + \alpha_z + \lambda(\gamma_2 + \gamma_1);$$  

(21a)

$$b_{s1} = r + \alpha_x \gamma_1 + \gamma_z;$$  

(21b)

$$b_{x0} = (\alpha_z + \lambda \gamma_1) \gamma_4;$$  

(21c)

$$b_{x1} = \lambda + \alpha_z \gamma_4 + \gamma_z.$$  

(21d)

If we have access to data on $(s,z)$ as well as wages for experience levels up to a maximum experience $T$, then we can estimate the five parameters $(K_x, b_{s0}, b_{s1}, b_{x0}, b_{x1})$ by regressing log earnings on schooling and ability $z$ at each experience level. This delivers a coefficient estimate for schooling and ability at each experience level for a total of $2 \times T$ coefficient estimates $[\hat{\beta}_{s0}, \hat{\beta}_{s1}, \hat{\beta}_{x0}, \hat{\beta}_{x1}]^{-1}$. Consider first the parameter estimates $[\hat{\beta}_{s0}]^{-1}$. The probability limit of these estimates is given by $b(x; K_x, b_{s0}, b_{x0}) = ((1 - \theta_x)b_{s0} + \theta_x b_{x0})$. This is a function of three parameters $(K_x, b_{s0}, b_{x0})$. A natural way of estimating these three parameters is to minimize the distance between $[\hat{\beta}_{s0}]^{-1}$ and $b(x; K_x, b_{s0}, b_{x0})$ by choice of $(K_x, b_{s0}, b_{x0})$. This is...
The Speed of Employer Learning

easily done by nonlinear least squares. Analogously, I can minimize the distance between \( \hat{\beta}_s(x; K_1, \hat{b}_s, \hat{b}_z) \) and \( b_s(x; K_1, b_s, b_z) \) to obtain estimates for \( (K_1, \hat{b}_s, \hat{b}_z) \).

This method results in two estimates of \( K_1 \), one from the schooling coefficients and one from the ability coefficients. Thus, \( K_1 \) is overspecified. I can test whether the difference between \( K_1 \) estimated separately for \( s \) and \( z \) is zero. Rejection of this test indicates that the variation in schooling and ability coefficients over the life cycle is not driven by the same learning process. I will show below that the data fail to reject the equality of \( K_1 \) estimated from either schooling or ability coefficients. This leads me to estimate \( K_1 \) using the information from schooling and ability coefficients jointly.

The simple normality assumptions on \( e_s \) and \( \lambda v + \epsilon \), therefore, put sufficient structure on the learning process to summarize the speed of learning in a single parameter \( K_1 \). This parameter can be estimated based on the coefficients obtained when log wages are regressed on schooling and AFQT over the life cycle. The model also implies a specification test. The same information revelation process drives how schooling and AFQT coefficients evolve with experience. The model of employer learning therefore predicts that the parameter estimated from the schooling coefficients should not be statistically different from the estimate obtained from the coefficients on the AFQT. The remainder of this section implements the method proposed here.

C. Implementation

This study analyzes the 1979–98 waves of the cross-sectional sample from the NLSY. Details of the data are reported in the appendix.

Equation (20) relates log wages to schooling and ability over the life cycle. Within experience levels, the relation between log wages and the independent variables schooling and ability is linear. The estimating equation corresponding directly to equation (20) regresses log wages on schooling and AFQT over the life cycle:

\[
\log (w_{i,x}) = \sum_s \beta_{s,x} (sD_s) + \sum_s \beta_{z,x} (zD_s) + \beta_{s,x} \Phi_{i,x} + \epsilon_{i,x}.
\]  

(22)

The controls \( \Phi_{i,x} \) include demographic variables and year dummies. Variable \( D_s \) stands for an indicator function that takes the value one if experience is \( x \) and zero otherwise. The parameters \( \{ \beta_{s,x}, \beta_{z,x} \}_{x=0}^{\infty} \) are known functions of the structural parameters \( \{ b_{s,x}, b_{z,x}, b_{s,z}, K_1 \} \):

\[
\{ \beta_{s,x}, \beta_{z,x} \}_{x=0}^{\infty} = \{(1 - \theta_s) b_{s,0} + \theta_s b_{s,1} \} x + \{(1 - \theta_z) b_{z,0} + \theta_z b_{z,1} \} z.
\]  

(23)

The goal is to estimate the parameters \( \{ b_{s,x}, b_{z,x}, b_{s,z}, K_1 \} \). For this purpose I treat each of the estimated coefficients \( \{ \beta_{s,x}, \beta_{z,x} \}_{x=0}^{\infty} \) as an observation. I then fit (after substituting \( \theta_s = xK_1/[(1 + (x - 1)K_1)] \)) the nonlin-
ear functions $b_i(x) = (1 - \theta_i)b_{\bar{z}_i} + \theta_i b_{\bar{z}_{\bar{z}}}$ and $b_i(x) = (1 - \theta_i)b_{\bar{z}_i} + \theta_i b_{\bar{z}_{\bar{z}}}$ to \{$\hat{\beta}_{\bar{x}_i}, \hat{\beta}_{\bar{x}_{\bar{z}}}$\} by choice of \{$b_{\bar{z}_i}, b_{\bar{z}_{\bar{z}}}, b_{\bar{x}_i}, K_1$\} with the method of non-linear least squares.

Figures 1 and 2 show the coefficients \{$\hat{\beta}_{\bar{x}_i}, \hat{\beta}_{\bar{x}_{\bar{z}}}$\} obtained from estimating equation (22) and the predicted values for these coefficients implied by the estimates of \{$b_{\bar{z}_i}, b_{\bar{z}_{\bar{z}}}, K_1$\} reported in columns 1 and 2 of table 2.\(^{11}\)

The figures demonstrate that the functional form predictions on the schooling and AFQT coefficients arising from the employer-learning model match the data well.

The parameter of interest, $K_1$, is estimated twice, once using the schooling coefficients and once using the coefficients on the AFQT. The estimates of $K_1$ are 0.2891 and 0.2293 with bootstrapped standard errors of 0.1139 and 0.0860, respectively. The difference between the point estimates is 0.0598 with a standard error of 0.0996. I therefore find no evidence that the parameter estimates of $K_1$ obtained from fitting the schooling and ability coefficients as a function of experience differ from each other.

The failure to reject the equality of estimates of $K_1$ estimated from schooling and ability separately suggests using the information from both schooling and ability coefficients jointly. One method that achieves this is minimizing the joint sum of the squared errors of both around their predicted values \{$b_i(x; K_1, b_{\bar{z}_i}, b_{\bar{z}_{\bar{z}}}), b_i(x; K_1, b_{\bar{z}_i}, b_{\bar{z}_{\bar{z}}})$\}. The objective function this estimate minimizes is $\sum_{i=0}^{9} \omega_i \times ((\hat{\beta}_{\bar{x}_i} - b_i(x; K_1, b_{\bar{z}_i}, b_{\bar{z}_{\bar{z}}}))^2 + (\hat{\beta}_{\bar{x}_{\bar{z}}} - b_i(x; K_1, b_{\bar{z}_i}, b_{\bar{z}_{\bar{z}}}))^2$. The weights $\omega_i$ are inversely proportional to the number of observations at each experience level that allows estimating \{$\hat{\beta}_{\bar{x}_i}, \hat{\beta}_{\bar{x}_{\bar{z}}}$\}. Table 2, column 3, provides the estimates from minimizing this distance function by choice of $K_1, b_{\bar{z}_i}, b_{\bar{z}_{\bar{z}}}, b_{\bar{x}_i}, b_{\bar{x}_{\bar{z}}}$. As expected, the estimate of $K_1$ (0.2592) obtained from this estimation procedure is roughly in the middle of the two estimates obtained from schooling and ability coefficients separately, but the dispersion around this estimate due to sampling error is still large. The standard error for instance is 0.0922. I find that 95% of the bootstrapped estimates are larger than 0.1411.

A value of $K_1 = 0.2592$ implies that initial expectation errors by employers on average decline by 26% during the first period. After three periods the initial error will on average have fallen by 51%, and after 5 years it will have declined to 36% of its initial value. The point estimate of $K_1$ therefore suggests that a worker’s productivity is to a large extent revealed within the first few years of his career. The model of employer learning also implies that any error remaining after the first couple of years is relatively persistent. It takes 26 years (on average) to reduce the remaining expectation error to less than 10% of its initial value. Close to retirement, after 40 years of work, the expectation error will still average about 6% of the initial value.

\(^{11}\) The standard errors are obtained by bootstrapping with 5,000 repetitions.
Fig. 1—Returns to schooling over the life cycle. The scatter displays the estimated coefficients on schooling for each experience level estimated using equation (22). The line shows the predicted returns to schooling over the life cycle implied by the estimates in table 2, column 1. The estimation of these parameters is described in Section III.
Fig. 2.—Returns to ability over the life cycle. The scatter displays the estimated coefficients on the standardized AFQT score for each experience level estimated using equation (22). The line shows the predicted returns to schooling over the life cycle implied by the estimates in table 2, column 1. The estimation of these parameters is described in Section III.
Table 2
The Speed of Employer Learning

<table>
<thead>
<tr>
<th>Full Sample (Both Genders, All Races)</th>
<th>(1) Schooling AFQT Score</th>
<th>(2) Schooling AFQT Score</th>
<th>(3) Schooling AFQT Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speed of learning $K_i$</td>
<td>.2891 (.1139)</td>
<td>.2293 (.0860)</td>
<td>.2592 (.0922)</td>
</tr>
<tr>
<td>Difference in the estimated $K_i$'s</td>
<td>.0598 (.0996)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Initial value $b_0$</td>
<td>.1078 (.0152)</td>
<td>-.0044 (.0303)</td>
<td>.1043 (.0107)</td>
</tr>
<tr>
<td>Limit value $b_{95}$</td>
<td>.0538 (.0047)</td>
<td>.1772 (.0164)</td>
<td>.0525 (.0051)</td>
</tr>
</tbody>
</table>

Note.—The reported parameters are estimated by nonlinear least squares using the coefficient estimates on schooling and Armed Forces Qualification Test (AFQT) score at different experience levels obtained from eq. (20) in the text. Section III describes the link between the parameters reported here and the estimated coefficients. The standard errors are obtained by bootstrapping with 5,000 repetitions. Columns 1 and 2 report the results obtained from schooling and the AFQT coefficient separately. Column 3 shows the parameter estimates obtained from using the coefficients on schooling and AFQT jointly.

The bootstrapped distribution just rejects a lower bound of 0.1411 at the 5% significance level. With $K_i = 0.1411$, learning progresses much more slowly. It takes 6 years for the average expectation error to decline by 50%, and 18 years for it to fall to 25%. At the end of the individual’s career, the expectation error will have declined to 13% of its initial value. Thus, the lower bound value of $K_i$ allows for a substantial degree of persistence. Nevertheless, most learning takes place within the first decade of individuals’ careers, even when $K_i = 0.1411$. In this case the average expectation error after 10 years of work experience amounts to 38% of the initial expectation error.

IV. Job-Market Signaling and the Speed of Learning

The previous sections showed how to estimate the speed of employer learning. The remainder of the article examines how the employer-learning model allows learning about the importance of the signaling motive for schooling decisions. How important job-market signaling is as a motive for schooling depends on how long employers are uncertain about individuals’ productive characteristics and thus rely on schooling as a signal of productivity. Therefore, an estimate of the speed of employer learning might provide an opportunity for progress on a question that has long eluded satisfactory answers (see Lange and Topel [2006] for an overview of the existing evidence on job-market signaling).

In this section, I discuss what we can learn about the contribution of job-market signaling from the employer-learning model alone. I start by defining the quantity of interest. The quantity of interest is the percent contribution of signaling to the increase in the present value of earnings
due to an additional year of schooling. I discuss identification of this parameter using the employer-learning model from Section III.A. I consider identification using both the conditional mean function of earnings and higher moments of log earnings. As I will show, the contribution of signaling to the gains from schooling is not identified within the employer-learning model. Furthermore, there are no economically sensible restrictions on the employer-learning model that would allow me to identify a bound on the parameter of interest.

Instead, I need additional information. I show how to bound the contribution of signaling if I know the costs of schooling. The opportunity costs of schooling are plausibly the major component of schooling costs, and it is possible to estimate the opportunity costs of schooling. This additional information and an economically sensible restriction on the model allows bounding the contribution of signaling. This restriction is that the additional information available to employers allows them to reduce their reliance on schooling when predicting individuals' productivity.

I want to emphasize that the assumptions required for bounding the returns to signaling are strong. Ultimately this is a consequence of the underlying problem that makes distinguishing the human-capital and the signaling models difficult: the crucial behavioral motivations are the same in both models. In both models, workers maximize earnings faced with a wage function of schooling, and in both models, firms pay wages equal to expected productivity. The fact that the main behavioral decisions are similar makes differentiating between the job-market-signaling and human-capital models difficult. The additional assumptions made in Sections IV and V provide an opportunity to quantify the contribution of job-market signaling within the framework of the employer-learning model. There is, of course, no free lunch; the additional assumptions are not trivial.

A. Defining the Parameter of Interest

Let \( i \) denote the rate with which labor earnings are discounted. Then the expected lifetime earnings until retirement \( T \) of an individual with characteristics \((s,q,z,\eta)\) at \( x = 0\) are

\[
\int_0^T \exp(-i\tau) E[W(s,q,y)|s,q,z,\eta] d\tau
\]

\[
= \int_0^T \exp(-i\tau) E[\exp(E[\tilde{\chi}|s,q,y] + H(\tau)]|s,q,z,\eta] d\tau. \tag{24}
\]
Differentiating equation (24) with respect to \( s \) delivers the increase in the present value of earnings due to an increase in schooling:\(^{11}\)

\[
\int_0^T \exp (-ir) \frac{\partial E[W(s,q',y')|s,q,z,\eta]}{\partial s} \, dr \\
= \int_0^T \exp (-ir) \frac{\partial E[\exp ((1-\theta)E[\bar{\chi}|s,q] + \theta \bar{\chi} + H(r)]|s,q,z,\eta]}{\partial s} \, dr \\
= \int_0^T \exp (-ir)E[W(s,q,y')|s,q,z,\eta][(1-\theta)(\lambda \gamma_z + \alpha_z)] \, dr.
\]

The expected increase in log wages at \( x \) due to direct productivity effects of schooling is given by \( r \). The term \( (\lambda \gamma_z + \alpha_z) \) in equation (25) captures the increase in earnings due to signaling at \( x = 0 \). In particular, \( \lambda \gamma_z \) reflects the increase in expected productivity with schooling due to the association between schooling and \( z \). Also, \( \alpha_z \) reflects the increase in expected productivity due to the association of \( \eta \) with schooling. Thus, the component \( (\gamma_z + \alpha_z) \) reflects the increase in expected productivity at \( x = 0 \) attributable to signaling. As experience increases, this contribution of signaling to contemporaneous earnings declines. The declining weights \((1-\theta)\) reflect the decline in the contribution of signaling over the working life.

Thus, I define the signaling contribution of schooling to the present value of earnings of an individual with characteristics \((s,q,z,\eta)\) as

\[
\int_0^s \exp (-ir)E[W(s,q,y')|s,q,z,\eta][(1-\theta)(\lambda \gamma_z + \alpha_z)] \, dr.
\]

This contribution of signaling varies with \((s,q,z,\eta)\) across individuals. I aggregate across individuals by averaging among individuals with equal schooling. By law of iterated expectations, \( E[\int_0^s \exp (-ir)E[W(s,q,y')|s,q,z,\eta][(1-\theta)(\lambda \gamma_z + \alpha_z)] \, dr | s] = \int_0^s \exp (-ir)E[W(s,q,y')|s][(1-\theta)(\lambda \gamma_z + \alpha_z)] \, dr \). I therefore define the parameter of interest as

\[
F_{\text{ms}} = \frac{\int_0^s \exp (-ir)E[W(s,q,y')|s][(1-\theta)(\lambda \gamma_z + \alpha_z)] \, dr}{\int_0^s \exp (-ir)E[W(s,q,y')|s][(1-\theta)(\lambda \gamma_z + \alpha_z) + r] \, dr}.
\]

I call the parameter \( F_{\text{ms}} \) the contribution of signaling. It measures the average contribution of signaling over the average total gains from schooling. The data allow estimating the average life-cycle earnings profiles of different schooling classes. Thus, \( E[W(s,q,y')|s] \) is observed. For a given discount rate the identification question therefore reduces to identifying \((\lambda \gamma_z + \alpha_z,r)\).

The next question is therefore whether \((\lambda \gamma_z + \alpha_z,r)\) are identified in

---

\(^{11}\) To arrive at the second line of eq. (22), I use eqqs. (12) and (13), the definition \( s^* = \bar{x} + \gamma_z \), and the normality of \( \bar{x} \) conditional on \((s,q,y')\). The third line is obtained by inserting eqqs. (3), (7), and (6) and collecting terms.
the employer-learning model. The answer is no. However, as I will show below, there are ways of bounding the parameter $F_{JMS}$ from equation (26) if I assume access to an estimate of the costs of schooling.

B. Identification Based on the Employer-Learning Model

I will now show that and therefore are not identified if I assume access to an estimate of the costs of schooling.

In order to discuss the identification of, I strengthen the assumptions made in Section III. In particular, I assume that $E[u_s|s,z] = E[u_t|s,z] = 0$. This assumption implies that equations (14) and (15) represent conditional expectations rather than linear projections. A consequence of this additional assumption is that equation (20) represents a conditional-expectation function.

The conditional-mean function (eq. [20]) is a function of $\rho$, $\lambda$, $\alpha_2$, $\gamma_5$, $\gamma_6$, $\lambda$, and $K_1$. The conditional-mean function therefore provides five pieces of information. The set of equations (21a)–(21d) shows these estimates as functions of the underlying parameters of the model. I will restate this set of equations here:

\[
\begin{align*}
\beta_0 &= r + \alpha_3\gamma_3 + (\lambda\gamma_2 + \alpha_2) + \lambda\gamma_3; \\
\beta_i &= r + \alpha_i\gamma_1 + \gamma_i; \\
\beta_2 &= (\alpha_1 + \lambda\gamma_1)\gamma_1; \\
\beta_6 &= \lambda + \alpha_3\gamma_4 + \gamma_6.
\end{align*}
\] (21a')–(21d')

Counting the parameters suggests that I will have difficulty identifying the parameters of interest $(\lambda\gamma_2 + \alpha_2, r)$ using these four equations only.

The model contains 11 underlying parameters $(r, \alpha, \alpha_2, \gamma_1, \gamma_2, \gamma_3, \gamma_4, \gamma_5, \gamma_6, \lambda, \text{and } K_1)$. The parameters $\alpha_i$ and $\gamma_i$ appear only as part of $(\lambda\gamma_2 + \alpha_2)$ and cannot be separately identified. Furthermore, $K_1$ is exactly identified and does not provide any information on the remaining parameters.

There are therefore four pieces of information (eqq. [21a']–[21d']) and nine parameters $(r, \lambda\gamma_2 + \alpha_2, \alpha_1, \gamma_1, \gamma_2, \gamma_3, \gamma_4, \gamma_5, \gamma_6, \lambda \text{ and } K_1)$.

Assume that $(r^*, (\lambda\gamma_2 + \alpha_2)^*, \alpha_1^*, \gamma_1^*, \gamma_2^*, \gamma_3^*, \gamma_4^*, \gamma_5^*, \gamma_6^*, \lambda^*)$ is a solution to the set of equations (21a)–(21d). The parameters $(r, \lambda\gamma_2 + \alpha_2)$ are identified, if $(r^*, (\alpha_1 + \lambda\gamma_2)^*)$ is the unique solution for these equations. It is simple to verify that $(r^* + k, (\lambda\gamma_2 + \alpha_2)^* - k, \alpha_1^*, \gamma_1^*, \gamma_2^*, \gamma_3^*, \gamma_4^*, \gamma_5^*, \gamma_6^*, \lambda^*)$ is also a solution, where $k$ is an arbitrary constant. Therefore, $(\lambda\gamma_2 + \alpha_2, r)$ are not identified from the conditional-mean function alone.
What about higher moments? Does the distribution of log wage residuals allow identifying $(\lambda \gamma_2 + \alpha_z, r)$? The residual in log wages is

$$u_{ix} = w(s, q, y^*) - E[w(s, q, y^*)|s, z, x].$$

Equation (27) can be expressed as

$$u_{ix} = (\alpha_1 + (1 - \theta)\lambda \gamma_1)\mu_i + \theta \mu_2 + \theta \sum_{j=1}^x E_j.$$

Equation (28) shows that we can, at most, learn about the parameters $(\alpha_1, \lambda \gamma_1), K_i$, and the distribution of $(\mu_1, \mu_2, \varepsilon_i)$ from the distribution of wage residuals. These parameters do not, however, help in identifying $(r, \alpha_z + \lambda \gamma_2)$ from the set of equations (21a’)-(21d’). Therefore, the crucial parameters $(\lambda \gamma_2 + \alpha_z, r)$ cannot be identified using either the conditional mean of earnings or the distribution of earnings residuals.

What is the intuition behind this nonidentification result? Why are $(\lambda \gamma_2 + \alpha_z, r)$ not identified in the employer-learning model? The employer-learning model implies that wages ultimately converge to true productivity. Given this property, why is $r$, the productivity-augmenting effect of education, not identified from the relation between schooling and productivity that ultimately emerges? This is because schooling correlates with other determinants of productivity $(\eta, q)$ that are not observed in the data. Ultimately, the regression coefficients on schooling reflect both the correlation of schooling with $(\eta, q)$ and the productivity-augmenting effect of schooling. Regressing log wages on schooling when experience is large still only delivers a biased estimate of $r$ due to omitting $(q, \eta)$. The identification problem arises for the same reason that signaling and human-capital models are hard to distinguish in the first place: the correlation between schooling and productivity differs from the productivity-augmenting effect of schooling. The correlation between productivity and schooling does not need to arise due to a direct positive effect of schooling on productivity. Thus, even if I can measure productivity directly (for instance, by observing wages at high experience levels), I still cannot distinguish between signaling and the human-capital model.

Rearranging equations (21a’) and (21b’) yields some insights into the difficulties in identifying $(\lambda \gamma_2 + \alpha_z, r)$:

$$\lambda \gamma_2 + \alpha_z = (b_{x0} - b_{xv} + \gamma_5 - \lambda \gamma_1 \gamma_5);$$

$$r = b_{xv} + \alpha_z \gamma_5 - \gamma_5.$$  

(21a’)

(21b’)

Clearly $\gamma_5$ and $\gamma_5$ play a crucial role in preventing identification of $(r, \alpha_z + \lambda \gamma_2)$. Inspection of equation (21b’) shows that the long-run es-

\[\text{Note that } (r, \alpha_z + \lambda \gamma_2, \gamma_5, \gamma_5) \text{ do not appear in eqq. (21c’) and (21d’). We therefore cannot learn about } (r, \alpha_z + \lambda \gamma_2, \gamma_5, \gamma_5) \text{ from eqq. (21c’) and (21d’).}\]
timate of the return to schooling diverges from the true productivity effect of schooling if $\alpha \gamma_3 \neq 0$ or $\gamma_5 \neq 0$. The coefficients $(\gamma_3, \gamma_5)$ are defined in equations (14) and (15). They capture the correlation of unobserved productivity components $(q, \eta)$ with schooling (controlling for $z$). Thus, the long-run observed relation between schooling and productivity differs from the true productivity-augmenting effect of schooling because schooling correlates with other components of productivity. For the same reason, the difference in the short-run and the long-run return to schooling (see eq. [21a']) cannot be taken as an estimate of the return to signaling.

Is it possible to make economically sensible restrictions on the employer-learning model that allow identifying $(\lambda r + \alpha_2, r)$? I believe not. At the heart of the signaling model is the positive correlation between schooling and unobserved productivity. Thus, if I want to examine the contribution of signaling, I cannot restrict unobserved productivity components to be uncorrelated with schooling; I cannot restrict $\gamma_3 = 0$ and $\gamma_5 = 0$ and cannot identify $(\lambda r + \alpha_2, r)$.

C. Identification Using the Schooling Decision

Identification using only the employer-learning model is therefore not possible, but I can make progress if I am willing to introduce additional information in the form of an estimate of the costs of schooling. An estimate of the costs of schooling is useful for identification since optimal schooling decisions require that the costs of schooling equal the sum of the gains from signaling and from productivity-augmenting effects of schooling. An estimate of the costs of schooling can be obtained under the assumption that individuals maximize the present value of lifetime earnings net of tuition costs. In this case the costs of schooling are largely due to forgone earnings and tuition costs. The method proposed here does not depend on which estimate of the costs of schooling is favored by the researcher. It is possible to bound the contribution of signaling, as long as one is willing to take a stand on the costs of schooling. Denote these costs of schooling as $\phi(s)$.

Identification of an upper bound on the returns to signaling also requires $\partial E[w(s, q, 0)|s]/\partial s \geq \partial w(s, q, 0)/\partial s$. This assumption ensures that the returns to schooling observed in the data (weakly) exceed those paid by employers. Thus, I assume that employers use the information $q$ available to them (but not in the data) to rely less on schooling to set wages. A sufficient set of conditions that ensures $\partial E[w(s, q, 0)|s]/\partial s \geq \partial w(s, q, 0)/\partial s$ is that schooling is on average positively related to other correlates of productivity (and thus productivity itself). In particular, it is sufficient to assume that $\partial E[q|s]/\partial s \geq 0$ and $\partial E[z|q, s]/\partial s = \gamma_1 \geq 0$.\footnote{A weaker sufficient assumption would posit that $\gamma_1 > -\alpha_1/\lambda$. This assumption is weaker since defining $(q, z)$ as positive predictors of $\chi$ implies that both $\alpha_1$ and $\lambda$ are positive.}

What is the content of
The Speed of Employer Learning

this restriction? The signaling model implies that unobserved correlates of productivity $\eta$ are positively related with schooling. The data reveal that measures of productivity $z$ observed by researchers correlate positively with schooling. Thus, what is required is that, analogous to $(\eta, z)$, employer-observed measures of productivity $q$ are also positively related to schooling. Then, employers use the information available to them to rely less on schooling for predicting productivity.

Optimal schooling decisions require that the gains from an additional year of schooling must equal the costs from an additional year of schooling. Thus, using equation (25),

$$\int_0^T \exp(-ir)E[W(s, q, y')]s, q, z, \eta][(1 - \theta)(\lambda \gamma_2 + \alpha_z) + r)\,d\tau = \phi(s). \tag{29}$$

From equation (25), $\partial w(s, q, 0)/\partial s = (\lambda \gamma_2 + \alpha_z) + r$. Therefore, condition $\partial E[w(s, q, 0)]/\partial s \geq \partial w(s, q, 0)/\partial s$ implies

$$\int_0^T \exp(-ir)\,E[W(s, q, y')]s, q, z, \eta][(1 - \theta)\left(\frac{\partial E[w(s, q, 0)]}{\partial s} - r\right) + r]\,d\tau$$

$$\geq \phi(s). \tag{30}$$

Averaging equation (29) within schooling levels, then, results in

$$\int_0^T \exp(-ir)\,E[W(s, q, y')]s, q, z, \eta][(1 - \theta)\left(\frac{\partial E[w(s, q, 0)]}{\partial s} - r\right) + r]\,d\tau$$

$$\geq \phi(s). \tag{31}$$

The only unobserved parameter on the left-hand side of equation (31) is the productivity parameter $r$. It is important to note that the left-hand side of equation (31) is strictly increasing in $r$. There is therefore a unique $r$ that satisfies the inequality in equation (31) as an equality. This $r$ represents a lower bound on the productivity increase associated with schooling that is consistent with optimal schooling decisions. Associated with this $r$ is a value $(\partial E[w(s, q, 0)]/\partial s - r)$, which represents an upper bound on the contribution of signaling to earnings at experience $x = 0$. Multiplying by $(1 - \theta)$ results in an upper bound to the contribution of signaling at each experience level $x$. Discounting with $\exp(-ix)$ and integrating over the lifetime allows me to construct an upper bound to the contribution of signaling to the total gains from schooling. Dividing through the total costs of schooling delivers the parameter $F_{JMS}$.

Figures 3 and 4 give a pictorial representation of this approach. In figure 3 I show the expected returns from signaling if schooling does not have a productivity-augmenting effect ($r = 0$). In that case all returns are
Fig. 3.—Expected earnings in the pure signaling mode. The solid lines show average earnings profiles for high school and college graduates. The dashed lines show expected earnings for workers with the productivity of an average high school graduate who chooses to graduate from college. The figure is drawn under the assumptions that (i) schooling acts as a pure signal and has therefore no productivity-augmenting effects and (ii) employers do not observe any additional information \( q \) about individuals' productivity.
Fig. 4.—Earnings profiles if schooling has signaling and productivity effects. The bold solid lines show average earnings profiles for high school and college graduates. The thin solid line depicts average productivity of high school graduates who decide to attend college. The dashed lines show expected earnings for workers with the productivity of an average high school graduate who chooses to graduate from college. The figure is drawn under the assumption that employers do not observe any additional information $q$ about individuals' productivity.
due to signaling. The figure corresponds to the case in which firms have no other information that allows them to predict productivity. This implies that at the beginning of his career an individual with the characteristics of an average high school graduate could earn a return from going to college equal to the difference between the average high school and the average college earnings. Over time the return he can expect to earn declines since subsequent signals will reduce his ability to masquerade as a higher productivity type. Figure 3 shows the increase in log earnings an individual of average high school characteristics can expect if $K = 0.2592$ and also if $K = 0.1411$.

In figure 3 the productivity-augmenting effect of schooling ($r$) is set equal to zero. The true productivity effect of schooling, though, is not known. I propose in this article to exploit the optimizing condition for individuals’ schooling choices. I assume that I have an estimate of the costs of schooling, and I maintain, for the sake of argument, that $\frac{\partial E[w(s,q,0)]}{\partial s} = \frac{\partial E[w(s,q,0)]}{\partial s}$. Figure 4 depicts the increase in earnings from graduating from college that an individual with average high school characteristics can expect if (schooling augments productivity). A worker who decides to go to college will gain from both the increase in productivity and the signaling effect of schooling. His overall gain from schooling will be given by the areas $A + B$ (suitably discounted). Area $A$ represents the productivity-augmenting effect of schooling. Area $B$ represents the contribution of signaling to the gains from schooling. The contribution of the productivity-augmenting effect $r$ and of signaling $(\alpha_1 + \lambda_1)$ sum to $\frac{\partial E[w(s,q,0)]}{\partial s}$ at $x = 0$. Thus, holding $\frac{\partial E[w(s,q,0)]}{\partial s}$ constant, an increase in $r$ (and area $A$) reduces $(\alpha_1 + \lambda_1)$ (and area $B$).

It is important to note that the overall gains from schooling are increasing in $r$, since any productivity-augmenting effect persists (in expected value) throughout the life of the individual. The sum of $A + B$ is therefore increasing in $r$. Optimality requires that the costs of schooling equal the gains. Therefore, the productivity effect of schooling is identified as that effect that equalizes the overall returns from schooling (including both signaling and productivity effects) with the costs of schooling. There is but one productivity effect for which this condition holds. This allows me to recover the bound on the productivity effect and, consequently, the bound on the contribution of signaling to the overall gains to schooling.

V. The Bound on the Contribution of Signaling

The method outlined in the previous section relies on having an estimate of the costs of schooling available. I will arrive at this estimate by ex-
ploiting the assumption that individuals maximize the present discounted value of lifetime earnings. The objective function is then

\[
\exp(-is)\int_0^\tau \exp(-ir)E[W(s, q, y^\prime)|s, q, z, \eta]d\tau - \phi(s).
\] (32)

Here \(\phi(s)\) denotes tuition costs of schooling. The average marginal costs of schooling among individuals with equal schooling is

\[
\phi(s; i) = i\int_0^\tau \exp(-ir)E[W(s, q, y^\prime)|s]d\tau + \zeta(s).
\] (33)

An estimate of \(\phi(s; i)\) is obtained from observed life-cycle earnings profiles and the 1980 tuition costs of $2,500 reported by Heckman et al. (2003).\(^{14}\) The notation emphasizes that this estimate is conditional on the rate \(i\) used to discount labor earnings. There is considerable uncertainty about this parameter, and I will therefore provide estimates of the contribution of signaling for a wide range of plausible discount rates.

Imposing the inequality (eq. [31]) to hold as an equality and inserting the cost estimate from equation (33) provides the equation from which to solve for the lower bound \(r\) and the upper bound on the contribution of signaling:

\[
\int_0^\tau \exp(-ir)E[W(s, q, y^\prime)|s]((1 - \theta)\frac{\partial E[w(s, q, \Omega)|s]}{\partial s} - \theta) + r)d\tau = \phi(s; i).
\] (34)

The components of equation (34) that are estimated from the data are the parameter \(K\), and consequently the sequence \(\{\theta\}_{t=0}^{1}\), the return to schooling \(\frac{\partial E[w(s, q, \Omega)|s]}{\partial s}\), and the wage-experience profile \(E[W(s, q, y^\prime)|s]\)\(t=0\). Table 2 reports a point estimate of \(K = 0.2592\). Furthermore, I reject values of \(K\) less than 0.1411 at a 95% significance level. Thus, I will report the results for \(K = 0.2592\) and also for value of \(K\) in the range \([0.1411, 0.2592]\). The return to schooling \(\frac{\partial w(s, \Omega)}{\partial s}\) is estimated using an earnings regression on schooling, a polynomial (quadratic) in experience, schooling interacted with the experience polynomial, race, gender, and year fixed effects. The coefficient of schooling in such a regression evaluated at experience = 0 is equal to 8.70% with a standard deviation of 0.36%. The wage-experience profile \(E[W(s, q, y^\prime)|s]\)\(t=0\) represents the mean wage of individuals with schooling \(s\) along the life cycle. For experience of 0–18 years, this mean wage is directly available from the data. For the later years of the life cycle, I assume that the mean wage is constant.

\(^{14}\) Annual tuition costs are reported in year 2000 dollars.
Table 3
The Contribution of Signaling to the Gains from Schooling with Estimated Speed of Learning

<table>
<thead>
<tr>
<th>Interest Rate (%)</th>
<th>Costs/Gains from Schooling (000s)</th>
<th>Contribution of Signaling (%)</th>
<th>Productivity Effects of Schooling (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. $K_1 = .2592$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.00</td>
<td>22.4</td>
<td>25.66</td>
<td>3.4</td>
</tr>
<tr>
<td>4.00</td>
<td>22.0</td>
<td>19.57</td>
<td>4.3</td>
</tr>
<tr>
<td>5.00</td>
<td>21.5</td>
<td>13.84</td>
<td>5.3</td>
</tr>
<tr>
<td>6.00</td>
<td>21.1</td>
<td>8.36</td>
<td>6.5</td>
</tr>
<tr>
<td>7.00</td>
<td>20.8</td>
<td>3.10</td>
<td>7.9</td>
</tr>
<tr>
<td>8.00</td>
<td>20.4</td>
<td>&lt;.00</td>
<td>8.1</td>
</tr>
<tr>
<td>8.70</td>
<td>20.2</td>
<td>&lt;.00</td>
<td>9.3</td>
</tr>
<tr>
<td>B. $K_1 = .1411$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.00</td>
<td>22.4</td>
<td>46.77</td>
<td>2.4</td>
</tr>
<tr>
<td>4.00</td>
<td>22.0</td>
<td>35.31</td>
<td>3.5</td>
</tr>
<tr>
<td>5.00</td>
<td>21.5</td>
<td>24.91</td>
<td>4.7</td>
</tr>
<tr>
<td>6.00</td>
<td>21.1</td>
<td>15.03</td>
<td>6.0</td>
</tr>
<tr>
<td>7.00</td>
<td>20.8</td>
<td>5.53</td>
<td>7.7</td>
</tr>
<tr>
<td>8.00</td>
<td>20.4</td>
<td>&lt;.00</td>
<td>9.5</td>
</tr>
<tr>
<td>8.70</td>
<td>20.2</td>
<td>&lt;.00</td>
<td>10.8</td>
</tr>
</tbody>
</table>

Note.—Calculations are based on the data for high school graduates. The components needed for this calculation are the speed of learning, the wage profile of high school graduates, the returns to schooling at graduation, and an estimate of tuition costs. The wage profile is estimated from high school wage profiles for experience of 0–18 years and is set constant over the remainder of the lifecycle. Individuals are assumed to work for 45 years subsequent to high school graduation (44 if they attend school an additional year). The wage return at experience = 0 is estimated from a Mincer earnings equation to be 8.70%. Tuition costs are set to $1,900, in line with the numbers reported by Heckman et al. (2003).

I evaluate equation (34) using the decision of individuals with a completed high school degree to acquire additional schooling.

The component of the analysis that is not directly available from the NLSY is the interest rate at which individuals discount their lifetime earnings. Given the importance of the interest rate for this analysis, I will show how the results vary with different interest rates. The results presented in table 3 summarize the analysis when the interest rate ranges from 3.00% to 8.70%. I consider the former to be a lower bound on the rate at which individuals discount risky investments. The upper limit of 8.70% equals the return to schooling at the no-experience level.

The rate of return on physical capital might give us an indication of the appropriate rate of return for risky investments. Mulligan (2002) estimates this rate of return from national capital income and capital stock measures. He reports the rate of return to be close to 8% during the twentieth century and close to 6% once taxation has been taken into account.
The Contribution of Signaling for $K_1 = 0.2592$

Table 3, panel A, reports the upper bound on the contribution of signaling obtained using $K_1 = 0.2592$ and a range of $i$ between 3.00% and 8.70%. The lower end of this range is seen to provide a lower bound on the discount factor used to discount risky labor earnings. The upper end is equal to the return to schooling observed at $x = 0$.

A value of $K_1 = 0.2592$ implies that employers learn quickly about workers’ productive characteristics. A consequence is that the contribution of signaling is relatively low. Over the entire range of $i$, the upper bound of the contribution of signaling is below one-third. As the discount rate increases, the contribution of signaling falls quickly. A value of $i = 5\%$ implies that the rate of return on human-capital investments is close to that on stock market investments observed in the United States over the last century. For $i = 5\%$, the upper bound on signaling is approximately 14%. A value of $i = 4\%$ results in an upper bound on signaling of about 20%. While this is not negligible, it does suggest that the contribution of signaling to the gains from schooling is much smaller than the contribution resulting from productivity-augmenting effects of schooling.\(^{16}\)

The Contribution of Signaling for $K_1 = 0.1411$

The data reject a speed of learning of less than $K_1 = 0.1411$ at a 95% significance level. Table 3, panel B, shows the upper bound on the contribution of signaling implied by this slow speed of learning. I again report the bound on the for the range $\{i \in [3\%, 8.70\%]\}$. The importance of $F_{MS}$ for the overall contribution of signaling is evident. The bound on signaling is substantially higher: for $i = 3\%$, the upper bound on the contribution of signaling is close to half of the total gains. For $i = 5\%$, a value of $i = 7.55\%$ is required.

\(^{16}\) Two additional features of table 3 are noteworthy. First, the low contribution of signaling and the high speed of employer learning, $K_1 = 0.2592$, lead the lower bound on the productivity contribution $r$ to be close to the interest rate used to discount labor earnings. This is because the gains from schooling are mostly due to the productivity-augmenting effect of schooling and the opportunity costs are tightly linked to the interest rate $i$. To equalize the costs and gains from schooling requires $r$ to be close to $i$. Second, if $i$ is close to the observed schooling return $\partial E[w(s, q, 0) | s] / \partial s$, then the implied bound on the contribution of signaling is negative. The costs of schooling consist of the discounting effect and tuition costs. Therefore, as the contribution of signaling declines, the productivity-augmenting effect of schooling has to compensate both for forgone earnings and the tuition costs of schooling. This leads the productivity-augmenting effect of schooling to exceed $\partial E[w(s, q, 0) | s] / \partial s$ as $i$ approaches $\partial E[w(s, q, 0) | s] / \partial s$ from below. At this point, however, the signaling contribution $(\partial E[w(s, q, 0) | s] / \partial s) - r$ becomes negative, a result that is clearly nonsensical. In order for the signaling model to apply, we need to restrict $i$ to those values that result in positive contributions of signaling. If $K_1 = 0.2592$, then $i < 7.55\%$ is required.
the bound on $F_{jms}$ is now 25%, substantially larger than for $K_i = 0.2592$. Nevertheless, for the entire range of $i$ considered, the bound on the contribution of signaling is below that attributable to the productivity-augmenting effect of schooling. Nevertheless, if the discount rates on labor earnings are low, and the “true” $K_i$ is slow, then the contribution of signaling to the overall economic contribution of schooling might be substantially larger than that suggested by the point estimate of $K_i$.

VI. Conclusion

In this article I present a structural model of how employers learn about workers’ productivity. The additional structure imposed on the employer-learning model allows me to estimate the speed of employer learning using the coefficients on variables that are easy and hard for employers to observe. Employers learn quickly. Their initial expectation errors decline on average by one-half within the first 3 years.

In a second contribution I use the estimated speed of learning to evaluate the importance of job-market signaling for schooling decisions. I show that the contribution of signaling to the gains from schooling is not identified within the employer-learning model. I can, though, identify an upper bound on the contribution of signaling if I have access to an estimate of the costs of schooling. I estimate the costs of schooling using the opportunity costs of schooling and average tuition.

The upper bound on signaling obtained is sensitive to variation in the discount rate on labor earnings and in the estimated speed of learning. For a wide range of parameter values, the bound on the contribution of signaling is very tight—the contribution of signaling is limited to less than 15%. But if the speed of employer learning is very slow ($K_i = 0.1411$) and the discount rate is low ($i = 3\%$), then it is possible that up to 45% of the gains from schooling are due to signaling. My preferred parameter values set the speed of employer learning equal to the point estimate $K_i = 0.2592$ and the discount rate to $i = 5\%$. This discount rate would set the rate of returns to investments in human capital approximately equal to the returns on risky capital investment. With these parameter values, the contribution of signaling to the gain from an additional year of schooling is less than 15%.

Appendix

The data used in this study stem from the 1979–98 waves of the NLSY. The NLSY was administered to respondents annually from 1979 to 1992. From 1994 on, the NLSY moved to a biannual sampling scheme.

The NLSY consists of three samples. The main or cross-sectional sample is a random, nationally representative sample of 6,111 young noninstitutionalized men and women between the ages of 14 and 21 at the time
of the first interview in 1979. The supplemental sample of 5,295 youths oversamples the Hispanic, black, or disadvantaged white population. The military sample consists of youths aged 17–21 who were enlisted in the military in September 1978.

I restrict the analysis to the 5,360 nonblack respondents from the cross-sectional sample. It is not possible to construct the AFQT score for 314 individuals. I remove 244 additional individuals for which the only reported observations occur prior to graduation. Next, I drop all observation years in which individuals do not work for pay or earn wages less than $1 or more than $100. This leaves me with 4,732 individuals and 54,259 observations. With increasing experience, the sample size declines rapidly. A large part of this is due to the biannual sampling scheme after 1994 and the young age of respondents at the onset of the study. I drop all observations for those experience levels with fewer than 1,000 respondents. This limits the analysis to those observations with an experience less than 18. This results in the loss of another 31 respondents and 5,329 observations. The total remaining sample consists of 4,701 individuals with 48,930 observations. All statistics in the article are unweighted.

Table A1 contains summary statistics for the main variables used in this study. The wage is calculated as the real average hourly rate of pay (measured in cents) for the current or most recent job. The deflator is taken from the 1999 report of the president. Experience is calculated as years since graduation. The more traditional experience measure (age − years of education − 6) delivers the same results. The AFQT was administered to the sample population in 1980. Thus, different cohorts took it at different ages. To eliminate age effects, I standardize the AFQT score within each cohort.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>SD</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Highest grade completed</td>
<td>13.12</td>
<td>2.29</td>
<td>6.00</td>
<td>20.00</td>
</tr>
<tr>
<td>ln (wage)</td>
<td>6.73</td>
<td>.52</td>
<td>4.61</td>
<td>9.21</td>
</tr>
<tr>
<td>Standardized AFQT</td>
<td>.00</td>
<td>1.00</td>
<td>−2.91</td>
<td>2.91</td>
</tr>
<tr>
<td>Experience</td>
<td>8.10</td>
<td>4.51</td>
<td>.00</td>
<td>17.00</td>
</tr>
<tr>
<td>Hispanic (%)</td>
<td>7.64</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female (%)</td>
<td>50.11</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note.—AFQT = Armed Forces Qualification Test. Statistics are based on the unweighted cross-sectional sample described in the appendix. The sample consists of 4,701 individuals with 48,930 observations in the years 1979–98.

References