THE RETURNS TO SCHOOLING AND ABILITY DURING THE EARLY CAREER: EVIDENCE ON EMPLOYER LEARNING AND POST-SCHOOL INVESTMENT*)

Fabian Lange
Yale University
fabian.lange@yale.edu

1/28/2005

Abstract

The empirical employer learning literature finds support for statistical discrimination using schooling. How economically relevant statistical discrimination (and consequently job market signaling) is depends on how fast firms learn about workers' productive types. I show that firms learn quickly: on average expectations errors about productivity decline by 50 percent within three years. Based on this I find that signaling contributes at most 25 percent to the gains from schooling. Moreover, the standard human capital model has similar implications for earnings as the model of employer learning and is consistent with variation in on-the-job training by schooling and ability.

*) I would like to thank my committee chairman, Robert Topel, and the members of my committee, Gary Becker and Casey Mulligan, for their guidance and support. I would also like to thank Nate Baum-Snow, Luigi Guiso, Steve Levitt, Justin Marion, Kevin Murphy, Derek Neal, Robert Waldman, and participants at seminars at Boston University, Brown, Ente Einaudi in Rome, Tor Vergata in Rome, UCLA, University of Chicago, University of Iowa, University of Michigan, University of Wisconsin, and Yale for helpful comments and suggestions. I gratefully acknowledge the financial support of the Bradley Foundation and the European Community's Human Potential Program under contract HPRN-CT-2002-00235 [AGE] All opinions expressed herein are the author's.

The relation between log earnings and experience is strong, especially at the beginning of individuals' careers. During the first ten years of a worker's career his earnings increase on average by about 60 percent.[1] During the following decade earnings rise by another 10 percent. These are large differences. Clearly, experience is rewarded in the labor market - the variation of earnings with experience is large and explains a substantial fraction of the observed variation in labor earnings.

The returns to experience vary with both schooling and with ability. More able individuals enjoy much larger wage growth with experience, whereas schooling lowers the experience gradient. Henry S. Farber and Robert Gibbons [1996] and Joseph G. Altonji and Charles R. Pierret [2001] document these patterns using the National Longitudinal Survey of Youth 1979. These authors interpret their findings as evidence of statistical discrimination on the basis of schooling with subsequent employer learning about workers' characteristics. Their contributions have provided the literature on employer learning with renewed impetus.[2]

The employer learning literature makes the conventional assumption that firms use schooling to predict workers' productivity. The employer learning literature also makes the unconventional assumption that correlates of productivity are available in the data that are not available to employers. The assumption that these ability measures are not available to firms implies that they are initially not priced into wages. At the beginning of the working life the correlation between these ability measures 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 ability measures increases. The returns to schooling decline with experience as firms cease to rely on schooling to predict productivity. The variation in the experience gradients with ability and schooling is therefore consistent with the model of statistical discrimination on the basis of schooling and subsequent employer learning.

However, the assumption that employers learn about individuals' productive types limits the economic significance of statistical discrimination. The faster employers learn, the shorter the time-period during which firms need to rely on schooling to predict productivity. The economic relevance of statistical discrimination therefore hinges on the question how fast employers learn about workers' productive types.[3] In this paper I show how we can estimate the speed of employer learning using observed returns to schooling and ability at different experience levels.

Employers learn fast. It takes on average three years for any initial expectation error to decline by 50 percent. This limits the importance of statistical discrimination on the basis of schooling for explaining observed schooling choices and returns to schooling. I emphasize this point by using the estimated speed of employer learning to derive a bound on the contribution of Job Market Signaling[4] to the gains from schooling. I estimate that Job Market Signaling contributes at most 25 percent to the gains of an additional year of schooling over the life-cycle of an individual. The evidence on the variation in the returns to schooling and ability with experience is therefore consistent with a model of statistical discrimination. However, maintaining the assumption of the employer learning literature I find that employer learning is fast.[5]

The employer learning literature derives its appeal from a belief that a simple human capital model will not generate the observed relation between experience gradients, schooling and ability. Farber and Gibbons (1996) for instance argue (page 1017) that a human capital explanation of the observed patterns depends on an "unlikely condition" regarding worker heterogeneity with respect to education and training. In a second contribution I challenge this belief. I demonstrate that a standard formulation of the human capital model with standard assumptions about the sources of heterogeneity across individuals generates similar predictions for the earnings equation as does the employer learning literature. The human capital model is attractive since it does not rely on the assumption that more information is available to the researcher than to agents in the economy. Furthermore, I present data on on-the-job training that is consistent with the human capital model and that explains a substantial fraction of the variation in experience gradients with ability and schooling.

To summarize, this paper examines the question of whether the observed variation in the returns to schooling and ability over the life-cycle is evidence for an important role of statistical discrimination and employer learning. I first show that the importance of statistical discrimination is limited by the speed of employer learning. This speed can be estimated, and the estimated speed of learning is fast. Thus the available evidence suggests that statistical discrimination and employer learning is of limited importance, even if we maintain the assumptions of the employer learning model. Moreover, the patterns for the returns to schooling predicted by the employer learning model are also consistent with a standard formulation of the human capital model. Additional evidence from training data suggests that human capital investments indeed generate at least part of the observed variation in the returns to experience with schooling and ability.

The paper is structured as follows: Section 1 establishes the facts that form the basis of this study. I pay particular attention to functional form considerations, since estimating the speed of employer learning relies on relaxing the functional form assumptions made by Farber and Gibbons [1996] and Altonji and Pierret [2001]. In Section 2 I show how one can estimate the speed of employer learning and implement this method in Section 3. In Section 4 I bound the contribution of signaling to the gains from schooling at less than one-quarter. To arrive at this bound I exploit the first order condition of schooling choice and the previously estimated speed of learning. Section 5 demonstrates that a simple human capital model generates the same predictions for the interaction of ability and schooling with experience in the earnings equation as does the employer learning model. Section 5 also examines data on training to obtain additional evidence on the importance of on-the-job investments to explain the observed patterns. Section 6 concludes.

I Schooling, Ability and Earnings - the Early Career

A The returns to schooling and ability vary with experience

This paper studies the variation in the experience gradient of earnings with schooling and ability. It is motivated by the findings reported in the employer learning literature, most notably Altonji and Pierret [2001]. Altonji and Pierret examine earnings as a function of experience, schooling and ability using the 1979-1992 waves of the National Longitudinal Survey of Youth (NLSY). Their main results rely on the Armed Forces Qualification Test score (AFQT) as a measure of ability that is plausibly hard to observe for firms.[6] The AFQT-score has the advantage that it is available for almost all respondents to the NLSY and thus allows using a large sample. This is important since estimating the speed with which employers learn about individuals' productivity requires examining the interaction of schooling and ability with experience. Furthermore using the same survey-data with a similar ability measure allows me to maintain a high degree of comparability with Altonji and Pierre [2001] and also with Farber and Gibbons [1996].

Altonji and Pierre [2001] relax the functional form restrictions of standard earnings equations by allowing schooling and the AFQT-score to interact with experience:

$$(1) \qquad w_{i,t} = \beta_o + \beta_s s_i + \beta_z z_i + \beta_{s,x} \left(s_i * x_{i,t}\right) + \beta_{z,x} \left(z_i * x_{i,t}\right) + f\left(x_{i,t}\right) + \beta'_{\Phi} \Phi_{i,t} + \varepsilon_{i,t}$$

Log wages $w_{i,t}$ of individual i in period t depend on schooling s_i, the AFQT-score (standardized by birth-cohort) z_i, experience $x_{i,t}$ and controls $\Phi_{i,t}$. Altonji and Pierre [2001] emphasize how the relation between log wages and ability and log wages and schooling change with experience. They estimate equation (1) with and without imposing the restriction $\left(\beta_{z,x} = 0\right)$. Of particular interest is how the estimate of $\beta_{s,x}$ changes as they relax this restriction. Allowing the effects of AFQT-scores on log wages to vary with experience greatly reduces the interaction between schooling and experience. Table 1 reproduces their results based on data from the National Longitudinal Survey of Youth (NLSY) for the years 1979-1992 in Columns 1 and 2. Columns 3-8 reports analogous results for the longer sample (1979-1998) available from the NLSY today.[7]

With constant effects of AFQT-scores along the life-cycle $(\beta_{z,x} = 0)$ the estimated returns to schooling are roughly constant with experience. The coefficient in Row e of Column 1 implies that the returns to schooling decline by less than 1/2 of a percentage point over the first 10 years of a worker's career. This parallel structure of log earnings often leads analysts to impose separability between schooling and experience in the earnings regression similar to the specification favored by Mincer [1958].[8] Altonji and Pierre [2001] emphasize what happens when they allow the AFQT-score to interact with experience $(\beta_{z,x} \neq 0)$. Two empirical findings are especially noteworthy. First, AFQT-scores are increasingly associated with earnings. The return to 1 standard deviation of the AFQT-score is about 6-8 percentage points larger with 10 years of experience than at the beginning of a workers career. Second, once we control for the interaction of the AFQT-score with experience the return to schooling declines by about 2 percentage points during the first decade of labor market participation.

Columns 3-8 in Table 1 show that varying the sample, the sample period or the specification does not alter the findings by Altonji and Pierre [2001]. In Column 5 and 6 I report median regression estimates including the 'zeros'. These estimates are roughly consistent with those in Columns 1 and 2, suggesting that differential labor force participation is not responsible for the observed patterns. Another concern arises from the fact that the AFQT-score is administered to the sample in 1980 at a time when not all individuals have yet completed schooling. In a recent paper Hansen, Heckman and Mullen [2003] show that schooling in part produces the AFQT-score. The results reported here are robust for repeating the analysis on the subsample of individuals who have not yet graduated by the time the test has been given. This ensures that within a birth-cohort individuals have equal amounts of schooling at the time the test is adminstered. As in the full sample, the interaction of schooling with experience declines by 2 percentage points if the AFQT-score is interacted with experience for this sample. The results are also robust to considering alternative experience measures, allowing for industry fixed effects and performing the analysis separately by different racial groups and gender.

B Is the variation in earnings growth linear in experience?

Table 1 indicates a positive relation between earnings growth and ability. One standard deviation in the AFQT-score is associated with between 0.5-1 percentage points additional earnings growth for each year of experience. An extra year of schooling however reduces earnings growth by about 0.25 percentage points each year. The specification (1) that underlies the results in Table 1 imposes the variation in earnings growth with ability and schooling to be constant across the life-cycle. If more able individuals experience 1 percentage point higher earnings growth in the first years of their career, then the predicted additional annual earnings growth at 10 years of experience is also 1 percentage point. The employer learning literature however emphasizes that the differences in earnings growth arise because firms learn about individuals abilities. It is reasonable that the speed with which firms learn about their worker declines after the initial period of workers careers. Below I develop an explicit formulation of the learning process that captures exactly this intuition. To estimate the speed with which employers learn I will have to relax the restriction that the variation in earnings growth with ability and schooling is constant across the life-cycle. Equation (2) relaxes this restriction:

$$w_{i,t,x} = \Sigma_x \beta_{s,x} (s_i * D_x) + \Sigma_x \beta_{z,x} (z_i * D_x) + \beta'_\Phi \Phi_{i,x} + \varepsilon_{i,t} \tag{2}$$

D_x represents a full set of indicator variables for the experience of individuals and $\left\{ \{\beta_{s,x}, \beta_{z,x}\}_{x=0}^{N}, \beta_\Phi \right\}$ are the parameters to be estimated. There are systematic deviations from linearity. An F-test rejects the linear formulation in equation (30) against the formulation in equation (2). For the full sample (both genders) the p-value is less than 0.0001 and for the male sample it is 0.0697. Estimates of $\{\beta_{s,x}, \beta_{z,x}\}_{x=0}^{N}$ obtained from specification (2) are plotted against experience in Figures 1 and 2.[9] The estimated coefficients on schooling and AFQT-scores change rapidly during the initial years of individuals' careers. After a few years they stabilize and fluctuate around some long-run return. This rapid convergence is responsible for generating the finding in Section 4 that the speed of employer learning is fast.

C Relaxing the linear specification in schooling and ability

In the previous paragraphs I proposed relaxing the restriction that is imposed on the estimates by restricting the experience term in the interaction with schooling and ability to be linear. In figures 3 and 4 I show the results from relaxing the linearity in the ability and schooling variable itself. The estimating equation resulting in figures 3 and 4 is:

$$\log(w_{i,t,x}) = \Sigma_{s,x} \gamma_{s,x} D_s D_x + \Sigma_{z,x} \gamma_{z,x} D_z D_x + \beta'_\Phi \Phi_{i,t} + \varepsilon_{i,t} \tag{3}$$

where D_s denotes a set of indicator variables for completed years of schooling and D_z denotes a set of indicator variables for the deciles of the AFQT-score distribution. In figure 3 and 4 we can see that the finding that schooling returns decline in experience and ability returns increase in experience is not an artefact of the linearity in schooling and ability imposed in equation (2). Instead we see that the experience gradient increases with ability over the entire ability distribution and similarily the experience gradient declines with schooling at all schooling levels.

D Evidence from other countries and time-periods

The 1980s and 1990s were a time of rapidly increasing returns to schooling and technological innovation. It is thus possible that the empirical patterns described in Table 1 simply reflect changes in the wage structure during this time-period. This concern can not be fully addressed by controlling for year fixed effects, but requires demonstrating similar patterns for different time-periods and economies. Hause [1972] studied the relation between ability and earnings as individuals acquire experience. He showed for 4 different samples (3 in the US, 1 in Sweden) that (i) the returns to ability (as measured by test scores) increase with age, (ii) that schooling and test scores covary positively and (iii) that schooling has a positive return. These results imply that controlling for the ability-experience profile in an earnings regression will lead to a decline in the estimated returns to schooling with experience and an increase in the returns to ability with experience. Thus Hause [1972] finds the same patterns in data from 1940-1960 as we do in data from the 1980s and 1990s. Galindo-Rueda [2003] demonstrates a similar finding for UK data for approximately the same time-period as that considered by Altonji and Pierre [2001]. Bauer and Haisken-DeNew [2001] find some support in German data on blue-collar workers, but not for white-collar workers. Evidence supporting the employer learning hypothesis in developing countries is reported by Foster and Rosenzweig [1993] from rural labor markets in Pakistan and India and by Strobl [2003] for Ghanaian data. The fact that similar patterns are found in data from a variety of countries across several decades combined with the robustness checks performed on the NLSY lead me to conclude that the findings reported in Table 1 are not spurious.

II Statistical Discrimination and Employer Learning

The employer learning literature interprets the variation in experience gradients with schooling and ability as evidence of statistical discrimination and public employer learning about workers productivity. The crucial assumption of this literature is that the ability measure used (in this case the AFQT-score) is not directly observed by firms.

Firms are initially unsure about workers' productive type. They use all available information, including education to predict individuals' productivity. With each year that the worker spends in the labor market an additional measure of individual productivity become available to all firms. Firms use these measures to update their expectations about individuals' productive characteristics and thus wages increasingly reflect productivity. Therefore the association between the AFQT-score and wages increases as workers age. And, the association between schooling and earnings declines as employers cease to rely on schooling to predict productivity.

In this section I contribute to this literature by estimating the speed with which employers learn about individuals' productivity. This speed of employer learning is crucial for understanding the importance of statistical discrimination in a number of economic settings. Models of screening discrimination [Bradford Cornell and Ivo Welch [1996], Shelley J. Lundberg and Richard Startz [1998], Joshua C. Pinkston 2003] link race and gender wage gaps to variation in the ability of firms to estimate the productivity of individuals of different demographic groups. If employer learning is fast, then these differences will not persist and screening discrimination is unlikely to be a good candidate for explaining sustained differences in labor market performance by race or gender.

Models of matching (Boyan Jovanovic 1979, MacDonald 1982, Robert Gibbons, Lawrence Katz and Thomas Lemieux and Daniel Parent 2002, and many others) often rely on firms learning about the quality of the match between firms and employees. In the present paper employer learning is assumed common across firms and refers to learning about general productive characteristics. As such it is not directly analogous to the matching literature. However, if we make the additional assumption that learning about worker-firm specific matches proceeds at the same rate as learning about general productive characteristics, then the speed of employer learning estimated becomes informative for the matching literature.

Another application for which the speed of employer learning is important is Job Market Signaling. The faster firms learn about individuals' true productive type, the less the signaling motive matters for individual schooling decisions. In this section I derive a structure that allows me to estimate the speed of employer learning. The next section uses the estimated speed of employer learning to quantify the potential importance of Job Market Signaling for individual schooling decisions. It is worth pointing out that the additional assumptions necessary for this exercise are significantly stronger than those needed to estimate the speed of learning. We do for instance not need to assume asymmetric information to estimate the speed of employer learning. Indeed the assumptions made to estimate the speed of learning are only slightly stricter than those imposed by Altonji and Pierret [2001] and Farber and Gibbons [1996] in their contributions. Quantifying the importance of Job Market Signaling however requires a number of additional assumptions on workers schooling decision. I am explicit about these assumptions as I derive the upper bound on the importance of Job Market Signaling in Section 5.

A The Employer Learning Model

Farber and Gibbons [1996] introduced a formulation of the employer learning model that has since become the standard formulation in empirical studies on statistical discrimination and employer learning. It will also form the basis for this study. This model of statistical discrimination specifies log productivity χ_i of individual i to be linear in schooling s_i, information q_i available to employers but not contained in the data, information z_i available in the data but not to employers, and information η_i available neither to employers nor in the data:

$$\chi_{i,x} = rs_i + \alpha_1 q_i + \lambda z_i + \eta_i + H(x) \tag{4}$$

The subscript i is understood and will be suppressed in the exposition from now on. The function $H(x)$ describes the relation between log productivity and experience x. The assumption that $H(x)$ does not depend on either education or the ability measure z is crucial. The additional assumption $H(x) = 0$ reduces the notational burden without altering the analysis.[10]

Employers form expectations over the components of log productivity unknown to them.

Assume that these expectations are linear in the information available to firms[11]:

$$
\begin{aligned}
(5)\qquad z &= E\left[z|s,q\right]+\upsilon=\gamma_1 q+\gamma_2 s+\upsilon \\
(6)\qquad \eta &= E\left[\eta|s,q\right]+e=\alpha_2 s+e
\end{aligned}
$$

In each period a noisy measurement y_x of log-productivity χ becomes available.

$$
(7)\qquad y_x=\chi+\varepsilon_x
$$

The information set about an worker with experience x available to employers consists of s,q and the vector $y^x=\{y_1,y_2,...,y_{x-1}\}$. All information is common to all employers, labor markets are competitive and a spot market for labor services exists. Therefore the wage of an individual during the post-schooling period t is given by the expected productivity conditional on the available information: $W\left(s,q,y^x\right)=E\left[\exp\left(\chi\right)|s,q,y^x\right].$

This structure yields interesting results if we make the additional assumption that the errors (υ,e) are jointly normal and that ε_x is i.i.d. normally distributed and independent of all variables in the problem. This error ε_x can thus be thought of as true noise in the productivity measurement y_x. These simple additional assumptions on the joint distribution of the error imply that employers use Kalman Filtering to update their expectation of individuals' productivity as new measurements y_x become available.

Equations (5) and (6) allow us to express log productivity as a linear function of the information available to employers at time x=0:

$$
\begin{aligned}
(8)\qquad \chi &= \left(r+\lambda\gamma_2+\alpha_2\right)s+\left(\alpha_1+\lambda\gamma_1\right)q+\left(\lambda\upsilon+e\right) \\
(9)\qquad &= a_o s+a_1 q+\left(\lambda\upsilon+e\right)=E\left[\chi|s,q\right]+\left(\lambda\upsilon+e\right)
\end{aligned}
$$

In each employers update their expectations of χ. The normality assumptions and the independence imposed on ε_t allow us to apply the rules of Kalman Filtering. These specify that the updated expectations are a weighted average of the initial expectation and the new measure of productivity. In period 2 (after one additional measurement has become available) the expected productivity is equal to

$$E\left[\chi|s,q,y^{2}\right]=E\left[\chi|s,q\right]+K\left(1\right)*\left(y_{1}-E\left[\chi|s,q\right]\right) \tag{10}$$

where the weight $K\left(1\right)$ on the correction term $\left(y_{1}-E\left[\chi|s,q\right]\right)$ depends positively on the variance of the expectation error $\sigma_{1}^{2}=\chi-E\left[\chi|s,q\right]=Var\left(\lambda v+e\right)$ and negatively on the variance of the additional measurement y_{1}, denoted by σ_{ε}^{2} :

$$K\left(1\right)=\frac{\sigma_{1}^{2}}{\sigma_{1}^{2}+\sigma_{\varepsilon}^{2}} \tag{11}$$

The new expectation error is normal, with variance $\sigma_{2}^{2}=\frac{\sigma_{1}^{2}\sigma_{\varepsilon}^{2}}{\sigma_{\varepsilon}^{2}+\sigma_{1}^{2}}$. It is convenient that the expectation error is independent of the realization of the observed y_{1}. Iterating on equation (10) and simplifying yields:

$$\begin{aligned} E\left[\chi|s,q,y^{x+1}\right] &= \frac{1-K\left(1\right)}{1+\left(x-1\right)K\left(1\right)}E\left[\chi|s,q\right]+\frac{xK\left(1\right)}{1+\left(x-1\right)K\left(1\right)}\left(\frac{1}{x}\Sigma_{\tau=1}^{x}y_{\tau}\right) \\ &= \left(1-\theta\left(x\right)\right)E\left[\chi|s,q\right]+\theta\left(x\right)\left(\frac{1}{x}\Sigma_{\tau=1}^{x}y_{\tau}\right) \end{aligned} \tag{12}$$

Inspection of equation (12) reveals the central role of $K\left(1\right)$ for the updating process of expected log productivity. This parameter represents the noisiness in productivity expectations based on initial information $\left(s,q\right)$ relative to the noisiness of subsequent measurements. If subsequent measures are relatively noisy compared to the initial expectations of firms, then the importance firms place on new measurements for updating their expectations is small. The properties of $\theta\left(x\right)$ are that $\frac{\partial\theta(x)}{\partial x}>0$ and $\lim_{x\rightarrow\infty}\left\{\theta\left(x\right)\right\}=1$. The weight placed on information obtained during the working life of an individual increases as more time is spent in the labor market and eventually approaches 1.

The wage process follows directly from process of updating expectations. The wage at x is given by

$$W(s,q,y^x) = E[\exp(\chi)|s,q,y^x] \tag{13}$$

The distribution of χ conditional on (s,q,y^x) is normal. We can therefore apply the rules of log-normal random variables:

$$w(s,q,y^x) = (1-\theta(x))E[\chi|s,q] + \theta(x)\left(\frac{1}{x}\Sigma_{\tau=1}^x y_x\right) + u(x) \tag{14}$$

The time-effect $u(x)$ captures both the experience-profile $H(x)$ as well as the interaction of the changing distribution of the expectation error $(\chi - E[\chi|s,q,y^x])$ with the exponential function.

Expression (14) relates log wages to the information (s,q,y^x) available to employers. The empirical quantities available in the data are however not (s,q,y^x) but (s,z). In Section 3 I will project log wages on (s,z) at each experience level. To interpret these projections I need to derive the implications for the coefficients on s and z in earnings regressions that are fully interacted with experience dummies. In order to proceed we need to define the linear projections of those variables (q,η) not available in the data on (s,z) :

$$q = \gamma_3 s + \gamma_4 z + u_1 \tag{15}$$

$$\eta = \gamma_5 s + \gamma_6 z + u_2 \tag{16}$$

The empirical objects we are considering are the linear projections of log wages on schooling and AFQT-scores at different level of experience. It simplifies the notation to treat these linear projections as conditional expectations. This simplification is not necessary for any of the results in this paper. The equation of interest can be written as:

$$E[w(s,q,y^x)|s,z,x] = E\left[(1-\theta(x))E[\chi|s,q] + \theta(x)\left(\frac{1}{x}\Sigma_{\tau=1}^x y_\tau\right) + u(x)|s,z,x\right] \tag{17}$$

and using equation (14) and the independence assumptions on ε_x we get:

$$(18) \qquad E\left[w\left(s,q,y^x\right)|s,z,x\right] = \left(1-\theta\left(x\right)\right)E\left[E\left[\chi|s,q\right]|s,z\right] + \theta\left(x\right)E\left[\chi|s,z\right] + u\left(x\right)$$

Equation (18) expresses the expectation of the wage at experience level x conditional on s and z as the weighted average of 2 terms plus a term independent of s and z. The first term captures the relation between the wage and (s,z) at x=0. Since $\theta(x)$ increases with experience the importance of this term declines. The second term reflects the relation between productivity itself and (s,z). As more and more information becomes available the weight on this term increases. The objects of primary interest in estimating the speed of learning are the weights $\theta(x)$. These weights determine how the contributions of either term to log wages vary with experience.

Using equation (4), the linearity assumptions $(5),(6)$ and the definitions $(15),(16)$ allows expressing each of the 2 terms in equation (18) as a linear function of schooling and the ability measure z. Consider the relation between log wages and (s,z) at the beginning of individuals' careers. This is capture by equation (18) with x=0:

$$\begin{aligned}(19) E\left[w\left(s,q\right)|s,z\right] &= E\left[E\left[\chi|s,q\right]|s,z\right] + u\left(0\right) \\ &= \left\{\underset{[A]}{r} + \underset{[B]}{\alpha_1\gamma_3} + \underset{[C]}{\left(\alpha_2 + \lambda\left(\gamma_2 + \gamma_1\gamma_3\right)\right)}\right\}s + \left\{\underset{[D]}{\left(\alpha_1 + \lambda\gamma_1\right)\gamma_4}\right\}z + u\left(0\right)\end{aligned}$$

The initial effect of schooling on wages is composed of 3 terms. The first ($[A]$) reflects the productivity effect of schooling. The second term ($[B]$) reveals the fact that schooling covaries with information q about productivity that is known to employers at the beginning of the career. The parameter γ_3 represents the regression coefficient of q on schooling. The positive effect of q on log productivity in turn is summarized in the parameter α_1. The comovement of schooling and the productivity prediction of firms is therefore captured by the product $\gamma_3\alpha_1$. This term $[B]$ can be thought of as analogous to the ability bias term introduced into the study of schooling returns by Griliches [1977]. Term $[C]$ is different in nature. Employers are aware that schooling covaries with information (z,η) that is unknown to them. They therefore use schooling to predict the unknown components of productivity. The statistical relation between schooling and wages (controlling for ability) reflects this prediction component of wages and is summarized by $[C]$. The dependence of log wages on z at x=0 is captured by component $[D]$ in equation (19). This term is analogous to the term $[C]$ in the schooling coefficient. Firms predict wages on the basis of q (in addition to s). The component of wages predicted in this manner correlates with the ability measure z. This correlation is captured by the term $[D]$ in equation (19). 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 equation (18) with x$\rightarrow\infty$. Then $\theta(x) \rightarrow \infty$ and thus equation (18) reduces to the second term:

$$E[\chi|s,z] = \left\{ \underset{[E]}{r} + \underset{[F]}{(\alpha_1\gamma_3 + \gamma_5)} \right\} s + \left\{ \underset{[G]}{\lambda} + \underset{[H]}{(\alpha_1\gamma_4 + \gamma_6)} \right\} z \tag{20}$$

The two components $[E]$ and $[G]$ simply capture the productivity effects of schooling and ability. The presence of the term $[E]$ in equation (20) makes clear that the approach discussed here does not assume that the productivity effects of schooling are zero over the life-cycle. Thus it is possible that education increases wages even in the long run. Neither the theory nor the empirical approach restricts the returns to schooling or the estimated schooling coefficient to converge to zero in the long run. The fact that the long-run coefficients do not converge to the true productivity effect of schooling or ability respectively is reflected by $[F]$ and $[H]$ in equation (20). $[F]$ is the equivalent to the Griliches ability bias in schooling after true productivity has been learned by firms. It differs from $[B]$ in equation (19) since over time firms have learned about the component of productivity η. This knowledge is reflected in wages and the covariance of schooling with this component of productivity therefore appears in term $[F]$ above. The component $[H]$ represents the bias in ability analogous to $[F]$.

The importance of equation (18) for the purposes of estimating the speed of learning is that it specifies wages as a weighted average of 2 components. The first captures the relation between schooling, ability and log wages before learning about individual productivity has taken place and the second after it has been completed. The problem is conveniently set up in such a form that each component is linear in schooling and ability:

$$(21) \quad E\left[w\left(s,q,y^{x}\right)|s,z,x\right]=\left(1-\theta\left(x\right)\right)\left(b_{s}\left(0\right)s+b_{z}\left(0\right)z\right)+\theta\left(x\right)\left(b_{s}\left(T\right)s+b_{z}\left(T\right)z\right)+u\left(x\right)$$

The exact mixture of these 2 components is determined by the weight $\theta(x)$. The relation between $\theta(x)$ and experience is known up to a parameter $K(1)$. This parameter summarizes the speed of employer learning and can be estimated from the coefficients on schooling and ability over the life-cycle.

An additional implication of the model is captured by this formulation of the learning process. The weight $\theta(x)$ is a function of time and the parameter capturing the speed of learning $K(1)$ only. Thus the speed at which the regression coefficients for ability and schooling converge to their limit value is governed by the same parameter $K(1)$. We can estimate this parameter separately for s and z and can test whether the difference between the estimated values is 0.

Let me briefly summarize the model of the wage-setting process developed above. Employers have access to a common pool of information about workers productivity. This information includes time-invariant information (s,q) as well as repeated measurements of productivity $\{y^t\}$. Risk-neutral employers use this information to form expectations about the productivity of individuals. They rely increasingly on information revealed over the course of individuals' careers. Wages therefore become increasingly associated with the AFQT-score. Over time firms rely less and less on schooling to infer workers productivity. The association of wages with schooling therefore declines with experience. The model of employer learning developed here allows summarizing the speed of learning in a single parameter $K(1)$. I show how this speed of learning can be estimated using the coefficients obtained when log wages are regressed on schooling and AFQT-score over the life-cycle. The same information revelation process drives how schooling and AFQT-coefficients evolve with experience. The model of employer learning therefore predicts that the parameter $K(1)$ estimated from the schooling coefficients should not be statistically different from the estimate obtained from the coefficients on the AFQT-score.

Note that until now we have not imposed any restrictions on individuals' knowledge of χ.

The restrictions refer exclusively to the distribution of information available to employers and in the data. Thus the analysis of the learning process and the estimate of the speed of learning applies both to models of symmetric ignorance in which firms and workers simultaneously learn about the workers characteristics as well as asymmetric information models, in which workers have more information about their productivity than do potential employers.

III Estimating the Speed of Employer Learning

The empirical counterpart to equation (21) is the regression equation (**??**) with $\gamma_3 = \gamma_4 = 0$. For convenience I restate equation (**??**) with $\gamma_3 = \gamma_4 = 0$:

$$\text{(22)} \qquad \log(w_{i,t,x}) = \gamma_1 s_i + \gamma_2 z_i + \Sigma_x \beta_{s,x} (s_i * D_x) + \Sigma_x \beta_{z,x} (z_i * D_x) + +\beta'_\Phi \Phi_{i,t} + \varepsilon_{i,t}$$

The controls $\Phi_{i,t}$ include demographic variables and year dummies. The estimates $\left\{ \widehat{\beta}_{s,x}, \widehat{\beta}_{z,x} \right\}_{x=0}^{T}$ are consistent estimates of $\left\{ \beta_{s,x}, \beta_{z,x} \right\}_{x=0}^{T} =$
$\left\{ (1-\theta(x)) b_s(0) + \theta(x) b_s(T), (1-\theta(x)) b_Z(0) + \theta(x) b_Z(T) \right\}_{x=0}^{T}$. The goal is then to estimate the parameters $\{b_s(0),\ b_s(T),\ b_z(0),\ b_z(T), K(1)\}$. For this purpose I treat each of the estimated coefficients $\left\{ \widehat{\beta}_{s,x}, \widehat{\beta}_{z,x} \right\}_{x=0}^{T}$ as an observation and fit the non-linear functions $\beta_s(x) = (1-\theta(x)) b_S(0) + \theta(x) b_S(T)$ with $\theta(x) = x * \frac{K(1)}{1+(x-1)K(1)}$ with the method of non-linear least squares.

Figures 1 and 2 show the coefficients $\left\{ \widehat{\beta}_{s,x}, \widehat{\beta}_{z,x} \right\}$ obtained from estimating equation (22) and the predicted values for these coefficients implied by the estimates $\{b_j(0), b_j(T), K(1)\}_{j=s,z}$. Table 2 shows these estimates obtained using non-linear least squares.[12]. The figures demonstrate that the functional form predictions on the schooling and AFQT-score 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-score. Based on the schooling coefficients I obtain a value of 0.2923. From the coefficients on the AFQT-score I estimate $K(1)$ as 0.2177. The bootstrapped standard errors for both estimates are 0.1142 and 0.0713 respectively. The difference between the point estimates is 0.0746 with a standard error of 0.1038. There is therefore no evidence that the parameters fitting the coefficients as a function of experience differ from each other. In the analysis that follows I will use $K(1) = 0.25$ which is both convenient and lies between the two parameter estimates obtained from schooling and ability respectively.

A value of $K(1) = 0.25$ implies that an initial assessment error on average declines by 0.25 during the first period. After three periods the initial error will on average have fallen by 1/2 and after five years it will have declined to about 37.5% of its initial value. The estimated speed of employer learning therefore suggests that workers productivity is to a large extent revealed within the first few years of a workers career. The model of employer learning also implies that any error remaining after the first couple of years is relatively persistent. To reduce the remaining expectation error three years by a further 50% takes on average an additional 5 years. To halve the error again requires an additional 13 years. After 40 years of work the remaining expectation error on average amounts to about 7% of the initial value.

IV How Important is Job Market Signaling? An Economic Application as a Metric for the Speed of Employer Learning

In the previous section I argued that the estimated speed of employer learning is 'fast'. This judgement was based on the examining the fraction of any initial expectation error on the part of employers remaining at different levels of experience. In this section I use an economic metric for evaluating how fast the speed of employer learning is: I ask, conditional on the estimated speed of employer learning, how important can Job Market Signaling be for individual schooling decisions? Altonji and Pierret (1997) recognize that the speed of learning is crucial for quantifying the economics importance of signaling. They provide calculations that suggest that if the speed of learning is indeed fast, then the share of the internal rate of return that is attributable to signaling has to be low. The present paper builds on their economic insight by providing an estimate of the speed of learning and by using this estimated speed of learning as a parameter in the schooling decisions of individuals. This allows me to derive a bound on the contribution of signaling and to clearly show what assumptions are required to arrive at this bound.

A Deriving a Bound on the Returns to Signaling

A student who decides to attend school for another year will increase his life-time earnings because schooling increases his human capital, but also because he communicates to employers that he possesses characteristics that are attractive to firms. This mechanism resides at the core of the signaling model as developed by Spence [1973]. The model of employer learning emphasizes that uncertainty about workers productivity is resolved dynamically. The speed at which this occurs limits the contribution of signaling to the returns to education. An individual may choose to masquerade as a high productivity type by obtaining additional schooling. The gains from masquerading are limited by the speed with which his true productivity is revealed in the market. The same is true for an individual that considers a schooling level below that associated with his true productivity. This individual can expect to receive lower wages for a period of time, but over time his wages will approach his true productivity level. Thus the gain from an additional year of schooling that is associated with signaling depends crucially on the ease with which employers can ascertain the true productivity of workers.

This simple insight can be used in the context of the individual optimizing decision to derive an upper bound on the returns to schooling. The approach is illustrated in Figures 5 and 6. The solid lines in Figure 5 depict the mean log earnings for 2 schooling levels, say high school and college. The returns to schooling in a Mincer Earnings Equation are roughly constant across the life-cycle and the earnings profiles are therefore parallel. Assume for now that schooling does not have a productivity effect. Assume also that firms have no other information (denoted by q in Section (I)) that allows them to predict productivity. An individual with productivity equal to the mean high school level has the option to go to college. This individual can expect to receive a wage equal to the average college graduate in the first period after leaving college. As he spends time in the labor market firms learn about his true productivity and therefore his expected wages will slowly approach his true productivity. The worker's expected wage is shown by the broken lines in Figure 5 for 2 different speeds of learning. The upper broken line corresponds to $K(1) = 0.1$ and the lower broken line to the estimated speed of learning of $K(1) = 0.25$. The discounted area between the high school earnings profile and the broken line represents the gains from signaling under the pure signaling model. Thus if we assume that the productivity returns to schooling are 0, then we could easily obtain an upper bound of the gains from signaling by applying the estimated speed of learning and then discounting back the gains from schooling for a high school graduate. This methodology obtains an upper bound, since it does not take into consideration that the information available to employers but not in the data will limit the returns an individual can obtain from masquerading as the wrong type.

The returns to productivity are however not known. The approach followed in this paper is therefore to exploit the optimizing condition for individuals' schooling choices. Figure 6 allows for a productivity effect of schooling. A worker who decides to go to college will gain from the increase in his productivity as well as the signaling effect of schooling. His overall gain from schooling will be given by the discounted sum of areas A and B. The costs of schooling consist largely of the opportunity cost of schooling. Given a discount rate this opportunity cost can be estimated. Optimality requires that the costs of schooling equal the gains. Therefore we can identify the productivity effect of schooling as that effect that equalizes the overall returns from schooling (including both signaling and productivity effects) with the costs of schooling. This allows us to recover the contribution of signaling and of productivity effects to the overall gains and returns to schooling.

Analytically this problem presents itself as follows. Consider an individual with characteristics (s, q, η, z) . This individual faces the problem of choosing a level of education s :

$$Max_{(s)} \left\{ \exp(-rs) \int_0^{T-s} \exp(-r\tau) E\left[W\left(s, q, y^{\tau}\right) | s, q, z, \eta\right] d\tau \right\} \tag{23}$$

The first order condition of this schooling choice problem is given by:

$$\text{(24)} \quad \int_0^{T-s} \exp(-r\tau) E\left[\frac{\partial W(s,q,y^\tau)}{\partial s} | s,q,z,\eta\right] d\tau$$
$$= r \int_0^{T-s} \exp(-r\tau) E\left[W(s,q,y^\tau) | s,q,z,\eta\right] d\tau + \exp(-r(T-s)) E\left[W\left(s,q,y^{T-s}\right) | s,q,z,\eta\right]$$

The LHS of this condition captures the gains from additional schooling due to the increase in the expected wage across the life-cycle. The increase in the wage at some future age a will depend on the measurements y^a taken until then. Using the structure developed in Section (3)[13] the returns to schooling at age a if y^a has been observed are:

$$E\left[\frac{\partial W(s,q,y^\tau)}{\partial s} | s,q,z,\eta\right]$$
$$\text{(25)} \quad = E\left[W(s,q,y^a) | s,q,z,\eta\right] * \left\{(1-\theta(a)) \frac{\partial w(s,q,y^0)}{\partial s} + \theta(a) \frac{\partial \chi(s,q,z,\eta)}{\partial s}\right\}$$

Thus the first order condition reads:

$$\int_0^{T-s} \exp(-r\tau) E\left[W(s,q,y^\tau) | s,q,z,\eta\right] *$$
$$\text{(26)} \quad \left\{(1-\theta(\tau)) \frac{\partial w(s,q,y^0)}{\partial s} + \theta(\tau) \frac{\partial \chi(s,q,z,\eta)}{\partial s}\right\} d\tau$$
$$\text{(27)} \quad = r \int_0^{T-s} \exp(-r\tau) E\left[W(s,q,y^\tau) | s,q,z,\eta\right] d\tau$$
$$+ \exp(-r(T-s)) E\left[W\left(s,q,y^{T-s}\right) | s,q,z,\eta\right]$$

The right hand side captures the costs from schooling due to discounting and the reduced length of the working life. Given $K(1)$ we can calculate the weights $\{\theta(x)\}_{x=1}^{T}$. At each experience level x the wage $E\left[W(s,q,y^\tau) | s,q,z,\eta\right]$ is approximately equal to $E\left[W(s,q,y^\tau) | s\right]$. This approximation can be justified on the grounds that the mean wage by schooling level is an average over $w(s,q,z,\eta)$ and that therefore least for one type of individual (s,q,z,η) choosing s the wage profile will equal exactly the wage profile associated with that schooling level[14]. Furthermore $E\left[W(s,q,y^\tau) | s,q,z,\eta\right]$ appears on both sides of the FOC (27). This implies that any proportional approximation error that is constant across experience cancels.

An additional problem arises because the wage return $\frac{\partial w\left(s,q,y^{0}\right)}{\partial s}$ is unobserved to the econometrician since the characteristics q are not observed. Firms use q to predict workers productivity. This implies that they rely less on schooling to estimate the productivity of individuals and therefore $\frac{\partial w\left(s,q,y^{0}\right)}{\partial s}$ will be smaller than the average return to schooling $\frac{\partial w(s,0)}{\partial s}$ observed at experience x. The estimated contribution of signaling when using $\frac{\partial w(s,0)}{\partial s}$ therefore represents an upper bound on the contribution of signaling. In the extreme case the information q transmits all the information on productivity. This case of course takes us back to the traditional human capital interpretation. The returns to signaling are obviously 0 in this case. The extent to which other information might allow firms to confidently predict the productivity of their workers is not addressed in this paper. Presumabley firms have additional information on individuals' productivity, since able individuals have a strong incentive to find credible ways other than schooling to signal this information. Therefore the bound derived here might be significantly greater than the true contribution of the signaling motive to the gains from schooling.

We can now derive the bound by rewriting the optimizing condition (27) as follows:

$$\begin{aligned} & \int_0^{T-s} \exp\left(-r\tau\right) \left\{ E\left[W\left(s,y^{\tau}\right)\right] \left(\underbrace{\left(1-\theta\left(\tau\right)\right) \frac{\partial w\left(s,0\right)}{\partial s}}_{[A]} + \theta\left(\tau\right) \frac{\partial \chi}{\partial s} \right) \right\} d\tau \\ (28) \qquad = \; & \underbrace{\exp\left(-r\left(T-s\right)\right)}_{[B]} E\left[W\left(s,y^{T-s}\right)\right] + r \underbrace{\int_0^{T-s} \exp\left(-r\tau\right) E\left[W\left(s,y^{\tau}\right)\right] d\tau}_{[C]} \end{aligned}$$

This expression contains the expected wage by schooling level, the return to schooling at experience level 0, the parameters $\theta\left(a\right)$ that depend in a known fashion on the estimated learning parameter $K\left(1\right)$, the interest rate and the effect of schooling on productivity of individuals. We can calculate all these components from the data or known sources with the exception of the productivity parameter. Imposing the first order condition (28) allows us to back out the productivity enhancing effects that equalize the costs and gains of schooling. Given this productivity effect of schooling the signaling effect can be calculated directly from condition $\left(28\right)$.

B The Contribution of Signaling to Gains from Schooling

Equation (28) represents a first order condition of an individual with s years of schooling. The right hand side divides the costs of an additional year of schooling into those that arise because more schooling reduces the time-period over which the individual receives earnings in the labor market (component $[B]$) and those due to the additional discounting of the earnings stream implied by postponing working life $[C]$. The first component will typically be of an order of magnitude smaller than the second and I will therefore ignore it in the discussion that follows even though it is included in the empirical calculations. With a minor rearrangement the equation of interest reads:

$$\begin{aligned} & \int_0^{T-s} \exp(-r\tau) E[W(s,y^\tau)] \left((1-\theta(\tau)) \left(\underbrace{\frac{\partial w_s(s,0)}{\partial s}}_{[1]} - \frac{\partial \chi}{\partial s} \right) + \underbrace{\frac{\partial \chi}{\partial s}}_{[2]} \right) d\tau \\ = \; & r \int_0^{T-s} \exp(-r\tau) E[W(s,y^\tau)]\, d\tau \end{aligned} \tag{29}$$

Term [1] represents the difference in the returns to schooling at t=0 that an individual can expect to receive due to signaling. It can not be emphasized enough that this extra payment reflects true differences in productivity across levels of schooling, even though it is not generated through the production of additional human capital. The term [1] reflects the increase in productivity between schooling levels that is not matched by an equivalent production of human capital at the individual level. This gain from signaling falls with time spend in the labor market because employers learn about the true productivity of individuals. The speed at which the gains from signaling fall with labor market experience is governed by $\theta(x)$.

The components of equation (29) that are estimated from the NLSY are the parameter $K(1)$ and consequently the sequence $\{\theta(x)\}_{x=0}^{T}$, the return to schooling $\frac{\partial w_s(s,0)}{\partial s}$ and the wage-experience profile $\{E[W(s,y^x)]\}_{x=0}^{T}$. I set $K(1) = 0.25$ based on the estimates reported in Table 2. The estimation of $K(1)$ is described at length in the previous sections. The results reported in Table 2 lead me to set $K(1)$ equal to 0.25. The return to schooling $\frac{\partial w_s(s,0)}{\partial s}$ is estimated using an earnings regression of earnings 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 $\{E[W(s, y^x)]\}_{x=0}^T$ represents the mean wage of individuals with schooling s along the life-cycle. For experience 0-18 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. I evaluate equation (29) 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 life-time earnings[15]. 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 4 to 8.70 percent. I consider the former to be a lower bound on the rate at which individuals discount risky investments. The upper limit of 8.70 percent equals the return to schooling at the no-experience level.[16]

The relatively quick learning process documented in an estimated speed of learning of $K(1) = 0.25$ implies that contribution of signaling to the overall gains from an additional year of schooling is small. The proportion of gains from schooling that are attributable to signaling is however not the only possible metric one might want to consider. Table 3 also demonstrates that given this high speed of learning any productivity effect of schooling is necessarily close to the discount rate used to discount earnings. In equation (28) the financial interest rate is expressed as a weighted average of the productivity effects from schooling and the returns to signaling. If the speed of learning is high, then the weight on the productivity effect is high and therefore the productivity effect is close to the discount rate. In the extreme case learning is instantaneous and we are back in the traditional human capital model when the return to schooling equals the rate at which labor earnings are discounted.

There is a long tradition to interpret the range of the estimated coefficients on schooling in earnings regressions of six to ten percent as a plausible range for the discount rate applicable to labor earnings. Table 3 shows that if we are willing to maintain this interpretation, then the bias of the returns to schooling in Mincer earnings regressions as an estimate of the productivity effects of schooling is relatively small. Table 3, however also contains the warning that one of the main variables necessary to estimate the productivity effects of schooling is the financial interest rate used by individuals in their discounting of future earnings. This is a variable we know relatively little about.

V An Alternative Interpretation: On the Job Training

The human capital model is the standard model of earnings dynamics along the life-cycle. This makes it natural to ask whether human capital investments can explain the variation in experience gradients with schooling and the AFQT-score. The appeal of the employer learning model derives largely from the perception that the human capital model can not easily explain the variation in the experience gradient with schooling and the AFQT-score. Farber and Gibbons [1996] for instance reject lief-cycle investments as an explanation of the observed variation in experience gradients. As they put it, "a pure OJT [on-the-job training] explanation of these 2 hypotheses requires the unlikely condition that worker heterogeneity related to investment in training be independent of heterogeneity related to education".[17] In this section I argue the contrary. I show that the human capital model very naturally generates the increasing return to ability over the life-cycle as well as the observed decline in the education-experience interaction once we allow for varying effects of ability over the life-cycle. And, I provide direct evidence on OJT that supports the human capital interpretation of differential earnings growth by schooling and AFQT.

To interpret the statistical relation between earning gradients, schooling and the AFQT-scores requires that we (i) identify the AFQT-score with a theoretical concept of the human capital model, and (ii) are explicit about the heterogeneity that generates the non-degenerate distribution of AFQT-scores conditional on schooling. Regarding the first point I choose to interpret the AFQT-score as a measure of the 'ability to learn'. By this I mean that the AFQT-score corresponds to a parameter in the production function of human capital that raises the marginal product of human capital investment. This interpretation is by no means the only one that can be given to ability. Indeed the Ben-Porath specification that has become a workhorse in this literature allows for 2 interpretations of ability. One of these interprets ability as a productivity augmenting parameter in the production function and corresponds to the notion of ability employer here. The alternative interpretation casts ability as the initial stock of human capital. Individuals of different ability levels do then differ in their earnings, even if they do not differ in their human capital investment. According to this interpretation the correlation between ability and schooling is negative, since opportunity costs of schooling are higher for more able individuals. The data however reveals that the AFQT-score and schooling are positively related. If the AFQT-score is interpreted as the ability to learn, then a positive correlation between schooling and the AFQT-score arises naturally.

The second point concerns the source of heterogeneity that results in observing individuals with the same education and different levels of ability outcomes. A standard assumption about heterogeneity across agents is that they differ in the costs of human capital acquisition, maybe because of difference in the rates of interest they face. There are alternative interpretations, but for the present purpose it is sufficient (indeed desirable) to demonstrate that the standard formulation of the human capital model can generate the variation in earnings gradients reported in Section 2.

The following formulation should be familiar. Consider the problem of infinitely lived individuals choosing investment $I(t)$ along the life-cycle. The production function $F(I(t), z)$ describes the relation between investments into additional human capital production and the expansion of earnings capacity that occurs due to this investment. The production function is indexed by the ability parameter z. Assume that human capital production requires investment $\left(F(0,z) = 0, \frac{\partial F(I,z)}{\partial I} > 0\right)$ and that $\left(\frac{\partial F(I,z)}{\partial z} > 0 \, , \frac{\partial^2 F(I,z)}{\partial I \partial z} > 0\right)$. The assumption on the cross-derivative gives empirical content to the statement that the AFQT-score measures the 'ability to learn'. To ensure interior solutions I also assume that $\frac{\partial^2 F(I,z)}{\partial I^2} < 0$. Note that the level of human capital does not enter the production function. This is the neutrality assumption of Ben-Porath [1967]. Indeed the production function employed by Ben-Porath satisfies the assumptions made here. Earnings capacity is denoted by $E(t)$.

After graduation optimal investment attains an interior solution of:

$$\begin{aligned} & \underset{\{I(t)\}}{Max} \{(E(t) - I(t)) + \lambda_t E(t+1)\} \\ s.t. \;\; E(t+1) &= E(t) + F(I(t), z) \\ I(t) &> 0 \end{aligned} \tag{30}$$

the multiplier λ_t captures the value of an additional unit of human capital. Assuming no depreciation and infinitely lived agents $\lambda_t = \lambda = 1/r$ where r is the interest rate with which agents discount future earnings. The problem is then stationary and the first order condition for investment while on the job is

$$1 = \lambda \frac{\partial F(I, z)}{\partial I} \tag{31}$$

The solution $I^*(\lambda, z)$ of this problem depends both on the value of an additional unit of human capital λ and on ability z. Schooling is defined as those periods when individuals invest their entire earnings potential into additional human capital production and do not receive earnings. Denote by $E(s, z)$ the earnings capacity of an individual with ability z after s years of schooling at his graduation. Years of schooling are determined by equating the gains from additional schooling to the gains from entering the labor market:

$$E(s, z) - I^* + \lambda F(I^*, z) = \lambda F(E(s, z), z) \tag{32}$$

The value of working consists of net labor earnings as well as the value of human capital acquired through on the job investments. The value while in school consists entirely of newly produced human capital. At graduation the knowledge accumulated in school equals the additional desired investment $I^* = E(s, z)$. A standard finding in the empirical literature is that there is a strong positive covariance between test-scores and schooling. To ensure this to be indeed the case I assume that the value of remaining in school increases faster with ability than does the value of working:

$$\frac{\partial\left(E\left(s,z\right)-I^{*}+\lambda F\left(I^{*},z\right)\right)}{\partial z}<\frac{\partial F\left(E\left(s,z\right),z\right)}{\partial z} \tag{33}$$

Below we will consider the conditional expectation of net earnings, as a function of schooling s, the ability parameter z and the value of human capital λ. The first order condition (32) defines λ as a function of schooling s and the ability parameter z. Schooling is increasing in λ and z which implies that we can draw downward sloping schooling isoquands in (λ, z)-space as in Figure 7. Ability z and the value of human capital λ vary inversely, holding schooling constant.

A The Conditional Expectation of Earnings

The empirical regularities reported in this paper and earlier by Farber and Gibbons (1996) as well as Altonji and Pierret (20001) refer to variation in the experience-gradient of the conditional expectation of net earnings $E\left[Y\left(s,x,z,\lambda\right)|s,x,z\right]$. The most important result, indeed the result that is proposed as a test of the employer learning model refers to how the interaction between experience and schooling varies if we allow for an interaction between experience and ability. In proposition 1 I state the predictions of the human capital concerning these same objects of interest. The human capital model as outlined above specifies net earnings of an individual with schooling s, experience x and ability z are:

$$Y\left(s,x,z,\lambda\right)=E\left(s,z\right)-I^{*}\left(\lambda,z\right)+x*F\left(I^{*}\left(\lambda,z\right),z\right) \tag{34}$$

Of particular interest is the variation of the experience gradients with schooling and ability. The following proposition summarizes the empirical predictions:

Proposition 1

The human capital model implies:

1.1. $\frac{\partial^2 E[Y(s,x,z,\lambda)|s,x,z]}{\partial x\partial s}>0$

1.2. $\frac{\partial^2 E[Y(s,x,z,\lambda)|s,x,z]}{\partial x\partial z}>0$

1.3 $\frac{\partial^2 E[Y(s,x,z,\lambda)|s,x]}{\partial x\partial s}>\frac{\partial^2 E[Y(s,x,z,\lambda)|s,x,z]}{\partial x\partial s}$

Proposition 1 states that the experience gradient of the level of net earnings increases in schooling (1.1) and in ability (1.2.). Most important however is the prediction (1.3.) that the variation in the experience gradient with schooling is greater if we condition on experience and schooling alone than if we condition on experience, schooling *and* ability.

Proposition 1 is formulated relative to level earnings, rather than logarithmic earnings. The main finding however, the decline in the schooling-experience interaction once we interact ability with experience, implies a similar relation for the conditional expectation of log earnings (up to variation in the 2nd moments).[18]

Proposition 1 is prooved in an appendix, but I want to briefly scetch the economic intuition driving this proposition. Graduation is defined as the period in which desired investments equal earnings capacity. Proposition 1.1. follows because - holding ability constant - schooling increases earnings capacity $E(s,z)$ at graduation. Earnings capacity $E(s,z)$ however determines $I^*(\lambda,z)$. Since investments are constant over the life-cycle we have that, conditional on ability, schooling increases investments during the entire life-cycle and therefore the earnings gradient is increasing in schooling. Proposition 1.2. follows by a similar argument with reference to ability, holding schooling constant. Proposition 1.3. follows from an omitted variable bias argument. To see this one just needs to realize that - in a linear framework - the difference between the returns to schooling if ability is included or not is simply an omitted variable bias term. This term is positive since schooling and ability covary and since the returns to ability are positive. Indeed the omitted variable bias is equal to the product of the regression coefficient of ability on schooling and the linear coefficient linking ability to log wages. The former is constant over the life-cycle and the latter increases in experience x. Thus the omitted variable bias is increasing. Any formulation of the model that delivers a positive covariance between schooling and ability and also an increasing return to ability will deliver the result emphasized by the employer learning literature. The most famous example of a human capital model that generates this finding is the formulation suggested by Ben-Porath with the necessary restriction to ensure that cov(s,z)>0.

B Does the Human Capital Model fit the Data?

Individuals make human capital investments by comparing the costs of additional units of human capital with the present value of future returns to these. It is therefore a theory of levels and the predictions derived in this section refer to wage levels. Table 4 reports the results for level regressions corresponding to Columns 3-8 in Table 1. The experience gradient varies positively with the AFQT-score and with schooling. The point estimates are therefore consistent with the theory. However, we can not reject that the interaction between education and experience in column (8) is zero. Nevertheless, the 95-percent confidence is large and includes economically significant positive variation of the experience gradient with schooling. In levels we simply do not have sufficient information in the data to evaluate the prediction that the returns to experience vary positively with schooling. More important is the central prediction of the employer learning literature that the return to schooling declines significantly if we control for ability. We showed above that this prediction is likewise generated by the standard human capital model with heterogeneity in discount rates. Table 4 indicates that indeed the experience gradient is much less sensitive to schooling if we allow ability to interact with experience than if not.

In this section I showed that a standard formulation of the human capital model indeed delivers the same predictions as the employer learning model that has received substantial attention in recent years. It is maybe not surprising that there exists a human capital model that is capable of reproducing these results, but the model developed here is indeed close to the original model developed in a series of classical papers in the sixties and seventies. The empirical results presented in Section 2 are therefore consistent both with the employer learning and with the standard human capital model. The evidence from wage profiles does simply not allow us to choose between these models. What other evidence allows us to consider the merits of the employer learning and the human capital model?

C On the Job Training

The NLSY offers the possibility to directly test the implication of the human capital model, that post-schooling investments are positively related to schooling and ability. Respondents in the NLSY are repeatedly asked about on-the-job training received in seminars or through employer training in the previous year. Such formal training can serve as an indicator of post-schooling investments. Table 5) shows that formal training increases in both the AFQT-score and in years of education. Individuals with more schooling or ability receive more training, regardless if training is measured contemporaneously or cumulated over the life-cycle. Clearly the prediction of the simple model above that ability conditional on schooling (and vice versa) predicts post-schooling investment is confirmed by this data. The next obvious question is whether controlling for this indicator of post-schooling investments eliminates the variation in the experience gradient with ability and schooling. Table 6) displays the analogous results to Tables 1 and 4 once we include the training measure in the earnings regressions. I report results both for logarithmic (Columns 1-4) and level specifications (Columns 5-8). The human capital model is specified in levels and I will thus refer to Columns 5-8 in my discussion.

Training clearly carries a positive return. Individuals who report training in the last year earn on average an additional 50 cents per hour. Cumulative training over the life-cycle has about an equal effect. Thus each year during which individuals receive training is associated with an additional 1/2-dollar in hourly wages. The key findings of Table 6 are however that controlling for training (i) reduces the interaction of the AFQT-score with experience by about 1/3 (see Columns 6 and 8, row d) and that (ii) the difference in the estimated schooling-experience interaction when we allow the AFQT-score to interact with experience and when we do not allow for this interaction is substantially reduced. Consider first the specification without controlling for formal training (Columns 5 and 6, row b). Allowing the AFQT-score to interact with experience leads to a decline in the returns to schooling as individuals age relative to the specification that restricts the effect of the AFQT-score to be constant with experience. Controlling for formal training therefore succeeds in removing part of this empirical pattern. At least partially, the findings emphasized by the employer learning literature are therefore attributable to formal training. And, the observed returns to training as well as the relation between training, schooling and ability is consistent with a simple version of the human capital model.

Controlling for formal training however does not succeed in fully accounting for the observed positive interaction of ability and experience, nor does it eliminate the finding that schooling reduces the experience gradient after interacting the AFQT with experience. One explanation is of course statistical discrimination and employer learning. An alternative explanation stresses that formal training measures elicited in survey data are imperfect proxies for post-schooling investments. In that case schooling and ability predict post-school investment even after controlling for formal training. The literature on training indeed suggests that this is the case. The fraction of overall post-school investment that is undertaken through formal training serves as a preliminary indicator for how likely it is that formal training is a valid proxy for post-schooling investment. John Barron, Mark Berger and Dan Black [1997] report results from 6 different data sources that show that formal training only accounts for about 1/7 of all formal and informal training received by individuals. This small share of formal training in overall training casts doubts on the validity of formal training as a valid proxy for overall post-schooling investments.

The lack of additional measures of post-schooling investments in the NLSY however precludes testing the validity of formal training as a proxy for post-schooling investments on this data-set. The current population survey in 1991 however contains data on both formal and informal training as well as schooling. Table 7 shows that in that data-set schooling predicts informal training even after controlling for formal training. The component of schooling orthogonal to formal training is still informative about post-schooling investments. This directly contradicts the notion that formal training represents a valid proxy for all post-schooling investments. Based on this evidence we can not determine whether the remaining variation in experience gradient with schooling and AFQT-scores is evidence of employer learning or whether it is due to forms of post-schooling investment that are not captured by formal training measures.

VI Conclusion

This paper is motivated by the empirical finding that the returns to schooling decline and the returns to ability increase with experience. The employer learning literature interprets these facts as evidence for statistical discrimination and employer learning. Initially firms are assumed to use easily observable variables such as schooling to predict individuals productivity. More and more information about workers' productive types becomes available as they spend time in the labor market. Thus, variables that correlate with productivity, but are hard to observe for firms become increasingly associated with wages. At the same time the association between easily observable variables (schooling) and wages declines.

This paper shows how we can estimate the speed of learning using the coefficients on easy- and hard- to observe variables. The estimate of the speed of learning is fast. Employer's initial expectation errors decline on average by 1/2 within the first 3 years. The statistical discrimination and employer learning model is consistent with the data, but they data indicates that the learning process proceeds quickly. I use the estimated speed of learning to evaluate the importance of job market signaling for schooling decisions. Even if we accept the learning hypothesis, the speed at which it progresses limits its importance for individual optimizing decisions. I demonstrate this by evaluating the contribution of signaling to the returns to schooling given the estimated speed of learning. At most one quarter of the total gains from an additional year of schooling to an individual can be attributed to job market signaling.

The model of statistical discrimination and employer learning is consistent with the reported facts. Contrary to the opinions expressed in the literature, I demonstrate that the standard human capital model of life-cycle investments is also capable of generating these same facts. The model interprets ability as 'ability to learn'. Heterogeneity in ability by schooling enters through heterogeneity in the value agents place on human capital. Schooling increases with ability and decreases with the discount rate. The model predicts a positive interaction (in levels) of experience with ability and schooling, if both ability and schooling are allowed to interact with experience in the earnings equation. However the (level-)returns to schooling increase faster if ability is not allowed to interact with experience. The empirical evidence from the earnings equations is consistent with these implications. Consistent with the human capital explanation of the variation in experience gradients is also the finding that formal training as measured by the NLSY increases with both schooling and ability. And, controlling for formal training reduces the magnitude of the effects of schooling and ability on the experience gradients.

There are many possible alternatives of formulating the human capital model. It is therefore not surprising that a formulation of the human capital model exists that is consistent with the empirical findings emphasized in this paper. That the standard Ben-Porath model with the traditional assumption of ability and discount rate heterogeneity generates these same findings is more surprising. The human capital model in its simplest form naturally generates the same findings that are proposed by the employer learning literature as evidence of statistical discrimination. The reason is that individuals with higher ability are more productive at graduation and invest more heavily into future earnings. Data on formal training provides additional support for the human capital model. Ability and schooling both predict formal training. Again, this is consistent with the human capital model. Furthermore, controlling for formal training in the earnings equations accounts for a substantial portion of the observed variation in experience gradients. However, there remains substantial variation in experience gradients with schooling and AFQT-scores after controlling for training. Whether this variation arises because formal training is an imperfect proxy for all post-schooling investments or whether it reflects statistical discrimination with employer learning can not be determined from the data presented here.

Appendix I: Data

The data used in this study stems from the 1979-1998 waves of the National Longitudinal Sample of Youth 1979. The NLSY79 is adminstered to the respondends annually from 1979-1992. From 1994 on the NLSY79 moved to a biannual sampling scheme.

The NLSY79 consists of 3 samples. The main or cross-sectional sample is a random, nationally representative sample of 6,111 young non-institutionalized 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 ages 17-21 who were enlisted in the military in september 1978. This paper restricts the analysis to the respondents from the cross-sectional sample. All statistics in the paper are unweighted.

I restrict the analysis to individuals working for pay, with wages ranging between \$1 and \$100 in 1990 dollars. 116 individuals do not conform to this criterium at some point over the sample period. It is not possible to construct the AFQT-score for an additional 305 individuals. Removing all observations reported previous to graduation removes an additional 314 individuals. With increasing experience the sample size starts declining 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 less than 1,000 respondents and therefore limit the analysis to experience less than 18. This results in the loss of another 41 respondents. The final sample then consists of 5,336 respondents with a total of 52,243 observations.

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 for the current or most recent job. The deflator is taken from the 1999 report of the president. The wage data is measured in cents. Values below \$1 and above \$100 are dropped. Experience is calculated as years since graduation. The more traditional experience measure (age-years of education-6) delivers the same results. The AFQT has been adminstered to the sample population in 1980. Thus individuals from different cohorts took it at different ages. To eliminate age-effects for the AFQT-score I standardized the AFQT-score within each cohort. The NLSY contains a measure of formal training. Respondents report participation in training programs such as apprenticeships, business college/school, formal work training and seminars as well as several smaller programs, such as Barber/Beauty School or Flight School. The framework in which the training question is asked changes in 1988 and subsequently the reported incidences of training increase by about 3 percentage points. Overall the incidence of training in any given year is approximately 15%.

Appendix II: Proposition 1

The distribution of (s, z) does not vary with experience. Therefore the experience gradient is given by

$$\frac{\partial E\left[Y\left(s,x,z,\lambda\right)|s,x,z\right]}{\partial x}=\frac{\partial E\left[x*F\left(I^{*}\left(\lambda,z\right),z\right)|s,x,z\right]}{\partial x}=E\left[F\left(I^{*}\left(\lambda,z\right),z\right)|s,z\right] \tag{35}$$

Given (s, z) the function $\lambda(s, z)$ is not random and therefore $\frac{\partial E[Y(s,x,z,\lambda)|s,x,z]}{\partial x} = F\left(I^{*}\left(\lambda\left(s,z\right),z\right),z\right)$. To derive the variation in the experience gradient therefore boils down to differentiating $F\left(I^{*}\left(\lambda\left(s\right),z\right),z\right)$ with respect to schooling and z. Thus, conditional on ability z we get

$$\frac{\partial^{2}E\left[Y|s,x,z\right]}{\partial x\partial s}=\frac{\partial F\left(I,z\right)}{\partial I}\frac{\partial I\left(\lambda,z\right)}{\partial\lambda}\frac{\partial\lambda\left(s,z\right)}{\partial s}>0 \tag{36}$$

This expression is positive, since investment increases in λ and since, at constant ability z, schooling increases with λ. Individuals with higher schooling, holding ability fixed, have a greater value for human capital investments and therefore will continue to invest more after they graduate. Similarly we get

$$\frac{\partial^2 E\left[Y|s,x,z\right]}{\partial x \partial z} = \left(\frac{\partial F\left(I,z\right)}{\partial I}\left(\frac{\partial I\left(\lambda,z\right)}{\partial \lambda}\frac{\partial \lambda\left(s,z\right)}{\partial z} + \frac{\partial I\left(\lambda,z\right)}{\partial z}\right) + \frac{\partial F\left(I,z\right)}{\partial z}\right) \tag{37}$$

At first glance it seems as if this relation can not be signed since λ varies negatively with z. However, we imposed above the restriction that schooling increases in ability z as is customarily observed in the data. This requires that $I^*\left(\lambda\left(s,z\right),z\right) = E\left(s,z\right)$ and we know that $\frac{\partial E(s,z)}{\partial z} > 0$. Thus we get $\frac{\partial E(s,z)}{\partial z} = \frac{\partial I(\lambda,z)}{\partial \lambda}\frac{\partial \lambda(s,z)}{\partial z} + \frac{\partial I(\lambda,z)}{\partial z} > 0$. Expression (37) is strictly positive. This prooves proposition 1.2 and 1.3. Proposition 1.3. follows from an omitted variable bias argument.

Affiliation: Yale University

References

Altonji, Joseph, G., Charles Pierret, ”Employer Learning and the Signaling Value of Education” in Industrial Relations, Incentives and Employment, I. Ohashi and T. Tachibanaki, eds. (London: Macmillan Prese Ltd., 1997), pp. 159-195.

——, and ——, ”Employer Learning and Statistical Discrimination,”, Quarterly Journal of Economics, CXVI (2001), 313-350.

Autor, David, David. Scarborough, ”Does Job Testing Harm Minority Workers,” manuscript, (2004), available at http://econ-www.mit.edu/faculty/index.htm?prof_id=dautor.

Baker, Michael, "Growth-Rate Heterogeneity and the Covariance Structure of Life-Cycle Earnings", Journal of Labor Economics, XV (1997), 338-375.

Barron, John, M., Mark C. Berger, Dan A. Black, ”How Well Do We Measure Training?”, Journal of Labor Economics, XV (1997), 507-529.

Bauer, Thomas, John P.Haisken-DeNew, ”Employer Learning and the Returns to Schooling”, Labour Economics, IIX (2001), 161-180.

Becker, Gary S., Human Capital, (New York, NY: Columbia University Press for the National Bureau of Economic Research, 1964).

Bedard, Kelly, ”Human capital versus signaling models: University access and high school dropouts”, Journal of Political Economy CIX (2001), 749-775.

Card, David, ”The Causal Effect of Education on Earnings”, Handbook of Labor Economics, Vol 3A, Chapter 30, eds. Orley Ashenfelter and David Card, (Amsterdam, Netherlands: North-Holland, 1999)

Cornell, Bradford, Ivo Welch, ”Culture, Information and Screening Discrimination,” Journal of Political Economy CIV (1996), 542-571.

Farber, Henry S., Robert Gibbons, ”Learning and Wage Dynamics,” Quarterly Journal of Economics, CXI (1996), 1007-1047.

Galindo-Rueda, Fernando, ”Employer Learning and Schooling-Related Statistical Discrimination in Britain,” IZA Discussion Papers, No. 777 (2003).

Gibbons, Robert, Lawrence Katz, Thomas Lemieux, Daniel Parent, ”Comparative Advantage, Learning, and Sectoral Wage Determination,” NBER Working Paper No. W8889 (2002).

Hanoch, Giora, ”An Economic Analysis of Earnings and Schooling,” The Journal of Human Resources, II (1967), 310-329.

Hansen, Karten T., James J. Heckman, Kathleen J. Mullen, ”The Effect of Schooling and Ability on Achievement Test Scores,” NBER Working Paper w9881, (2003).

Hause, John C., ”Earnings Profile: Ability and Schooling,” The Journal of Political Economy, LXXX (1972), S108-138.

——, ”The Fine Structure of Earnings and the On-the-Job Training Hypothesis”, Econometrica, IIL (1980), 1013-1029.

Heckman, James J., Lance Lochner, Christopher Taber, ”Explaining Rising Wage Inequality: Explorations with a Dynamic General Equilibrium Model of Labor Earnings of Heterogeneous Agents,” Review of Economic Dynamics, I (1998), 1-58.

Heckman, James J., Lance Lochner, Petra Todd, ”Fifty Years of Mincer Earnings Regressions,” NBER Working Paper 9732 (2003).

Jovanovic, Boyan, ”Job Matching and the Theory of Turnover,” The Journal of Policital Economy, XCVII (1979), 972-990.

Ljungqvist Lars, Thomas J. Sargent, Recursive Macroeconomic Theory, (Cambridge, MA, MIT Press 2000).

Lundberg, Shelly J., Richard Startz, " Race, information and segregation," manuscript available at http://ssrn.com/abstract=148334 (1998).

MacDonald, Glenn M., "A Market Equilibrium Theory of Job Assignment and Sequential Accumulation of Information," American Economic Review, LXXII (1982), 1038-1055.

Mincer, Jacob, Schooling, Experience and Earnings, (New York, NY, NBER Press 1958).

Riley, John C., "Testing the Educational Screening Hypothesis," Journal of Political Economy, LXXXVII (1979), S227-252.

Spence, Michael, "Job Market Signaling," Quarterly Journal of Economics, LXXXIII (1973), 355-379.

Stiglitz, Joseph E., "The Theory of "Screening," Education and the Distribution of Income," American Economic Review, LXV (1975), 283-300.

Strobl, Eric "Is Education used as a Signaling Device for Productivity in Developing Countries? Evidence from Ghana," IZA Discussioin Paper, No. 683 (2003).

Topel, Robert H., Michael P. Ward, "Job Mobility and the Careers of Young Men," The Quarterly Journal of Economics, LVII (1992), 439-479.

Notes

[1]Estimates from a sample of white males from the NLSY79.

[2]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. Fernando Galindo-Rueda [2003] reports similar findings using UK data. Thomas Bauer and John P. Haisken-DeNew [2001] find the same patterns among German blue collar, but not white collar, workers. From developing countries there is supporting evidence by Andrew D. Foster and Mark R. Rosenzweig [1993] from rural Pakistan and India and Eric Strobl [2003] from Ghana. The findings reported in John C. Hause [1972] are also consistent with Farber and Gibbons [1996] and Altonji and Pierret [2001]. Hause however interprets these results very differently.

[3]Altonji and Pierret [1997] recognize that the speed of employer learning is crucial for the economic significance of statistical discrimination. They argue that observed coefficient patterns are consistent with a fast speed of learning and that this limits the contribution of signaling to the life-cycle returns to schooling. Sections 3-5 of this paper build on their contribution in showing how to obtain an estimate of the speed of employer learning and how to use this estimate to bound the returns to signaling.

[4]Statistical discrimination on the basis of schooling is the core assumption on firm behavior made in the Job Market Signaling model.

[5]A cautionary note is in order here. Whether a given estimated speed of learning is perceived as fast depends on the application considered. The analysis of the importance of job market signaling for explaining the gains from schooling leads me to call the estimated speed of learning fast. In other applications the same speed of learning might be judged to be slow and therefore statistical discrimination might be considered important.

[6]Wigdor and Green [1991] provide the main study linking AFQT-scores to subsequent on-the-job performance. Their main conclusion from a decade long effort of studying the link between job performance and the AFQT is that the AFQT indeed predicts future job performance within the military.

[7]Column 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 paper. 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 utilized by Altonji and Pierret in that I restrict myself to the main NLSY sample and that I use the waves 1979-1998. Their study exploits both the main and the supplemenatal sample, but restricts itself to the 1979-1992 waves. These changes in the sample selection do not affect the basic findings as is evident when we compare columns 1 and 2 with columns 7 and 8. The sample selection criteria and variables used in this study are discussed in more detail in Appendix 1.

[8]Heckman, Lochner and Todd [2003] present evidence challenging the parellel structure of log earnings with experience.

[9]The meaning of the solid lines in these figures will be explained in Section 2.

[10]Naturally I will allow for an experience profile in the empirical analysis.

[11]A normalization of the coefficient vector α_1 allows suppressing q in equation (3).

[12]The standard errors are obtained by boot-strapping with 2,000 repititions.

[13]We have that $W(s, y^a, a) = \exp(w(s, y^a, a))$. Thus:

$$E[W_s(s, y^a, a)|\alpha] = E\left[\exp(w(s, y^a, a))\frac{\partial}{\partial s}(w(s, y^a, a))|\alpha\right]$$

The analysis in Section 3 implies:

$$\frac{\partial}{\partial s}(w(s, y^a, a)) = (1 - \theta(a))w_s(s, 0) + \theta(a)\frac{1}{a}\Sigma_{\tau=1}^a\frac{\partial y_t}{\partial s}$$

Note that the measurement error in $y_t = x + \varepsilon_t$ is independent of the conditioning set and $w_s(s, 0)$ and $x_s(s, \alpha)$ do not depend on any of the unknown random measurements. Therefore we have:

$$E[W_s(s, y^a, a)|\alpha] = E[W(s, y^a, a)|\alpha]((1 - \theta(a))w_s(s, 0) + \theta(a)x_s(s, \alpha))$$

[14]This is ensured based on assumptions (2) and (3) and the additional assumption that the support of the parameter q is compact.

[15]Also not available is the length of working life, which I set to equal 45 years.

[16]The rate of return on physical capital might give us an indication on 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 20th century and close to 6% once taxation has been taken into account.

[17]Farber and Gibbons [1996, p. 1017]

[18]The employer learning literature is not consistent in specifying its prediction in logarithms or levels. Farber and Gibbons (1996) specify their learning model in levels, whereas Altonji and Pierret (2001) specify the learning model in logarithms (as I do in Section 2). Whether the model is specified in levels or logs is arbitrary for the learning model and depends simple of whether the linearity in the information set is imposed in levels or logs. This underscores that the relevant predicted implication of the employer learning model is the prediction corresponding to Proposal 1.3.