Working Paper Series

Employer Learning and Statistical Discrimination

Joseph G. Altonji, Charles R. Pierret

Working Papers Series
Macroeconomic Issues
Research Department
Federal Reserve Bank of Chicago
December 1997 (WP-97-11)

FEDERAL RESERVE BANK
OF CHICAGO

First Draft: July 1995
This version: October 1997

Employer Learning and Statistical Discrimination

Joseph G. Altonji

Northwestern University and Federal Reserve Bank of Chicago

Charles R. Pierret
Bureau of Labor Statistics

This research was supported by the Institute for Policy Research, Northwestern University, the Bureau of Labor Statistics, U.S. Department of Labor, and the National Science Foundation. We owe a special debt to Nachum Sicherman for assisting us with the NLSY data. We thank Paul Devereux, Judith Hellerstein, Derek Neal, Bruce Weinberg, and participants in seminars at BLS, Berkeley, Boston University, Columbia, The Federal Reserve Bank of Chicago, Indiana University, University of Maryland, McMaster, NBER, Northwestern, the University College London, the UpJohn Institute and the U. of Western Ontario for helpful comments. We are responsible for all errors and omissions. The opinions stated in this paper do not necessarily represent the official position or policy of the U.S. Department of Labor, the Federal Reserve Bank of Chicago, or the Federal Reserve System.

Employer Learning and Statistical Discrimination

Abstract

We provide a test for statistical discrimination or "rational" stereotyping in environments in which agents learn over time. Our application is to the labor market. If profit maximizing firms have limited information about the general productivity of new workers, they may choose to use easily observable characteristics such as years of education to "statistically discriminate" among workers. As firms acquire more information about a worker, pay will become more dependent on actual productivity and less dependent on easily observable characteristics or credentials that predict productivity. Consider a wage equation that contains both the interaction between experience and a hard to observe variable that is positively related to productivity and the interaction between experience and a variable that firms can easily observe, such as years of education. We show that the wage coefficient on the unobservable productivity variable should rise with time in the labor market and the wage coefficient on education should fall. We investigate this proposition using panel data on education, the AFQT test, father's education, and wages for young men and their siblings from NLSY. We also examine the empirical implications of statistical discrimination on the basis of race. Our results support the hypothesis of statistical discrimination, although they are inconsistent with the hypothesis that firms fully utilize the information in race. Our analysis has wide implications for the analysis of the determinants of wage growth and productivity and the analysis of statistical discrimination in the labor market and elsewhere.

JEL Classification: D83, J31

Joseph G. Altonji
Department of Economics
Northwestern University
Evanston, IL 60208
(847) 492-8218
altonji@nwu.edu

Charles Pierret
Bureau of Labor Statistics
2 Massachusetts Ave. NE Suite 4945
Washington, D.C. 20212
(202) 606-7519
pierret_c@bls.gov

1. Introduction

People go through life making an endless stream of judgments on the basis of limited information about matters as diverse as the safety of a street, the quality of a car, the suitability of a potential spouse, and the skill and integrity of a politician. When hiring, employers must assess the value of potential workers with only the information contained in resumes, recommendations, and personal interviews. What do employers know about the productivity of young workers, and how quickly do they learn? Given lack of information about actual productivity, do employers "statistically discriminate" among young workers on the basis of easily observable variables such as education, race, and other clues to a worker's labor force preparation. Many issues in labor economics hinge on the answers, including the empirical relevance of the signaling model of education (Weiss (1995), statistical theories of discrimination (Aigner and Cain (1977), Lundberg and Startz (1983)), and the interpretation of earnings dynamics. The desirability of changes in the laws governing hiring procedures, evaluation of employees, layoff and firing costs, and the provision of references for former employees also hinge on the answers. Although labor economists typically assume wages are strongly influenced by employer beliefs about worker productivity, there is little empirical research on how much employers know about their workers, or about how this information changes with time in the labor market.[1]

In this paper we explore the implications of a hypothesis that we refer to as Statistical Discrimination with Employer Learning, or SD-EL. Our working hypothesis is that firms have only limited information about the quality of workers in the early stages of their careers. They distinguish among workers on the basis of easily observable variables that are correlated with productivity such as years of education or degree, the quality of the school the person attended, race, and gender. (To avoid misunderstanding we wish to stress that part of the relationship between wages and race and gender may reflect biased inferences on the part of employers or other forms of discrimination that have nothing to do with productivity or information.) Firms weigh this information with other information about outside activities, work experience to date, references, the job interview, and perhaps formal testing by the firm. Each period, the firm observes noisy indicators of the worker's performance. Over time, these make the information observed at the start redundant. Wages become more closely tied to actual productivity and less strongly dependent upon the information that was readily available at the beginning of a worker's career. The main contribution of the paper is to provide a way to test for whether firms

[1]There is a large empirical and theoretical literature on labor market search and on the effects of learning about the quality of the job match on wages and mobility. See Devine and Kiefer (1991) for a comprehensive survey.

statistically discriminate on the basis of readily available information such as education and race. We also provide a way to estimate the learning profile of firms and address the issue of whether firms have a stable view of the productivity of workers with many years of labor market experience.

Our research builds on some previous work, particularly Farber and Gibbons (1996).[2] Farber and Gibbons investigate three implications of employer learning. Imagine a variable s (say schooling) which firms can observe directly and a second variable, z (say AFQT test scores or sibling's wage rate) which firms cannot observe directly. They show first that employer learning does not imply that the coefficient on s in a wage regression will change with experience. This is because future observations, on average, simply validate the relationship between expected productivity and s for new entrants. Their empirical evidence is generally supportive of this result, although they note that a positive interaction could arise if schooling is complementary with training. (Positive interactions are found in a number of data sets, including the PSID.) Second, they show that the part of z that is orthogonal to information available to employers at the beginning of a worker's careers will have an increasingly large association with wages as time passes. Third, they note that wage growth will be a Martingale process, at least in the case in which productivity of the worker is constant.

In this paper we focus on a different but related proposition that allows us to examine the issue of statistical discrimination. The proposition concerns how controlling for the experience profile of the effect of z on wages alters the interaction between experience and s. We show that not only should the coefficient on z rise with time in the labor market, but the coefficient on s should fall. We investigate these propositions using data on young men from the NLSY. We also explore the implications of statistical discrimination on the basis of race, which is also easily observable to employers and is

[2] Other relevant references are Gibbons and Katz (1991) which we discuss below and Parsons (1993). Glaezer (1992) uses variances in wage innovations as a measure of learning. His work is somewhat closely related to Farber and Gibbons. However, he attempts to distinguish between information that is specific to the job match and information about general productivity. Foster and Rosenzweig (1993) use data on piece rate and time-rate workers to investigate several implications of imperfect information on the part of employers that are different from the one studied here. Their results imply that the incompleteness of employer information is an important issue. Studies following performance evaluations within firms based on the EOPP data, or studies using firm personnel files (Medoff and Abraham (1980)) are also relevant, but have a very different focus than the present paper. Parsons (1986), Weiss (1995) and Carmichael (1989) provide useful discussions of some of the theoretical issues on the link between wages and employer perceptions about productivity. Albrecht (1982) conducts a test of screening models of education based on the idea that education will have less impact on the probability a worker will be hired if the worker was referred to the firm by another worker because some of the information contained in education will be transmitted through the referral. Montgomery (1991) presents a model in which employers obtain valuable information on the productivity of new employers through referrals and is part of a large literature on labor market networks. For empirical evidence see Holzer (1988).

correlated with hard to observe background variables that influence productivity.[3] While our basic theoretical framework and most of the empirical analysis assumes that all employers have the same information about workers, we provide a preliminary discussion of the implications of models in which the current employer has an information advantage.

In Section 2 we present our basic theoretical framework in a setting in which information is public, and then informally discuss the case in which it is private. We also consider the effect that associations between s, z, and job training would have on the analysis. In Section 3 we discuss the NLSY data used in the study. In Section 4 and 5 we present our results for education and race. In section 6 we present results in which we control for job training. In section 7 we discuss the case in which employer information is private and provide some evidence on how hard to observe variables are related to the probability of a layoff and the wage losses associated with layoffs. In section 8 we point out that interpreting our estimates of the time profile of the effect of AFQT on wages as the result of employer learning implies that high ability workers would have a substantial financial incentive to take the AFQT to differentiate themselves from those who are less able in this dimension. The fact that we do not generally observe this raises additional research questions. In section 9 we close the paper with a discussion of some of the implications of our analysis for a number of standard topics in labor economics and a research agenda.

2. Implications of Statistical Discrimination and Employer Learning for Wages.

2.1 A Model of Employer Learning and Wages

In this section we show how the wage coefficients on characteristics that employers can observe directly and on characteristics they cannot observe directly will change with experience if employers statistically discriminate and become better informed about workers over time.

[3] We are using the term "statistical discrimination" as synonymous with the use of the term "rational expectations" in the economics literature. We mean that in the absence of full information, firms distinguish between individuals with different characteristics based on statistical regularities. In other words, we mean that firms form stereotypes that are rational in the sense that they are consistent with reality. Many papers that use the term statistical discrimination analyze race or gender differentials that arise because firms have trouble processing the information they receive about the performance of minority group members. This difficulty may lead to negative outcomes for minorities because it lowers their incentives to make unobservable investments that raise productivity. It also may lead to negative outcomes if the productivity of a job match depends on the fit between the worker and the job. Some papers also consider whether firms that start with incorrect beliefs about the relationship between personal characteristics and productivity (inaccurate stereotypes) would correct them, and, in models with worker investment, whether the priors held by firms may be self fulfilling. See Aigner and Cain (1977), Lundberg and Startz (1983), Lang (1986), and Coate and Loury (1993) and Oettinger (1996). In Oettinger's model productivity is match specific and productivity signals are noisier for blacks than whites. As a result the sorting process across job changes is less efficient for blacks, and a race gap develops over time.

Our model is very similar to Farber and Gibbons (1996). Let y_{it} be the log of labor market productivity of worker i with t_i years of experience. y_{it} is determined by

(1) $$y_{it} = rs_i + H(t_i) + \alpha_1 q_i + \Lambda z_i + \eta_i$$

where s_i is years of schooling, z_i is a vector of correlates of productivity that are not observed directly by employers but are available to the econometrician, and $H(t_i)$ is the experience profile of productivity. The variable η_i consists of other determinants of productivity and is not directly observed by the employers or the econometrician. The elements of z_i might be a test score, the income of an older sibling, father's education, or indicators of childhood environment such as books in the home or ownership of a library card. We normalize z_i so that all the elements of the conformable coefficient vector Λ are positive. Without loss of generality we scale η_i so that it has a unit coefficient in the productivity equation.

In addition to s_i, the employer observes a vector q_i of other information about the worker that is relevant to productivity. The elements of q_i are related to productivity by the coefficient vector α_1. For now we assume that the experience profile of productivity does not depend on s_i, z_i, q_i, or η_i. In section 2.2 we discuss the sensitivity of our analysis to this assumption. In most of the analysis we suppress the i subscript. All variables are expressed as deviations from population means. Although we use years of schooling and race as our examples of s, our analysis applies to any variable that employers can easily observe.

We assume that the conditional expectations of $E(z|s,q)$ and $E(\eta|s,q)$ are linear in q and s, so

(2) $$z = E(z|s,q) + v = \gamma_1 q + \gamma_2 s + v$$

(3) $$\eta = E(\eta|s,q) + e = \alpha_2 s + e ,$$

where the vector v and the scalar e have mean 0 and are uncorrelated with q and s by definition of an expectation.[4] The links from s to z and η may be partially due to a causal effect of s.[5] Equations (1), (2) and (3) imply that $\Lambda v + e$ is the error in the employer's belief about the log of productivity of the worker at the time the worker enters the labor market. The sum $\Lambda v+e$ is uncorrelated with s and q. We make the additional assumption that $\Lambda v + e$ is independent of q and s.

[4] The exclusion of q from the conditional mean of η is innocuous, since we are simply defining η and the coefficient vector α_1 on q in (1) so that the mean of η does not depend on q.

[5] For example, below we use the Armed Forces Qualification Test (AFQT) as z and years of education as s, and Neal and Johnson (1996) present evidence that years of education have a sizable positive effect on AFQT.

Each period that a worker is in the labor market, firms observe a noisy signal of the productivity of the worker,

(4) $\xi_t = y + \epsilon_t$

where y is y_t - H(t) and ϵ_t reflects transitory variation in the performance of worker i and the effects of variation in the firm environment that are hard for the firm to control for in evaluating the worker. (We continue to suppress the i subscripts.) The term ϵ_t is assumed to be independent of the other variables in the model. We are also implicitly assuming that the component of ϵ_t that reflects temporal variation in productivity from sources specific to worker i is serially uncorrelated, because otherwise firms would have an incentive to base compensation in t+1 on what they know about the worker specific component of ϵ_t.[6] However, ϵ_t may be serially correlated as a result of the other factors.

Since the employers know q and s, observing ξ_t is equivalent to observing

(5) $d_t = \Lambda v + e + \epsilon_t = \xi_t - E(y|s,q)$

The vector $D_t=\{d_1,d_2,...,d_t\}$ summarizes the worker's performance history. Let μ_t be the difference between Λv+e and $E(\Lambda v+e|D_t)$. By definition μ_t is uncorrelated with D_t, q and s but in addition we assume μ_t is distributed independently of D_t, q and s.

We also assume for now that q, s, and the worker's performance history (summarized by the vector $D_t=\{d_1,d_2...d_t\}$ are known to all employers, as in Farber and Gibbons. (We discuss the private information case in Section 7.) As a result of competition among firms, the worker receives a wage W_t equal to the expected value of productivity Y_t ($Y_t = \exp(y_t)$) times the multiplicative error component $\exp(\varsigma_t)$ that reflects measurement error and firm specific factors that are outside the model and are unrelated to s, z, and q. The wage model is

(6) $W_t = E(Y_t|s,q,D_t)e^{\varsigma_t}$

Using (1), (2), (3) and (6) leads to

(7) $W_t = E(Y_t|s,q,D_t)e^{\varsigma_t} = e^{rs+H(t)} \; e^{(\alpha_1+\Lambda\gamma_1)q+(\alpha_2+\Lambda\gamma_2)s} \; e^{E(\Lambda v+e|D_t)} \; E(e^{\mu_t}) \; e^{\varsigma_t}$

Taking logs and collecting terms leads to

(8) $w_t = (r + \Lambda\gamma_2 + \alpha_2)s + H^*(t) + (\Lambda\gamma_1 + \alpha_1)q + E(\Lambda v+e\,|Dt) + \varsigma_t$

---

[6] The firm's knowledge of a serially correlated productivity component would imply serially correlated transitory variation in the wage error of the type found by Farber and Gibbons (1996), but would not have much effect on our analysis.

where $w_t = \log(W_t)$ and $H^*(t) = H(t) + \log(E(e^{\mu_t}))$. We will suppress the ς_t term in the equations that follow.

In the context of the debate over signaling models of education Riley (1979) and others have noted that unless the relationship between schooling and actual productivity changes, the coefficient on s will not change. This is true regardless of <u>why</u> s is related to productivity. Farber and Gibbons also make this point by showing in a similar model that the expected value of the coefficient of an OLS regression of w_t on s does not depend on t. They estimate an equation of the form

(8a) $w_t = b_{st}s + H(t) + \alpha_1 q + E(\Lambda v + e|D_t)$

with q treated as an error component. They find that b_{st} does not depend much on t.

Farber and Gibbons also make a second point, which is that if one adds the component z' of $(\Lambda v + e)$ that is uncorrelated with the employer's initial information s and q to the wage equation and estimates

(8b) $w_t = b_{st}s + b_{z't}z' + H(t) + \alpha_1 q + E(\Lambda v + e|D_t)$,

the coefficients on s do not depend on t. This follows almost immediately from the first result, because adding a second variable to a regression model has no effect on the expected value of the first if the two are uncorrelated. They provide evidence from NLSY that b_{st} is relatively constant and $b_{z't}$ is increasing in t.

In this paper, we establish a related set of results that permit one to examine the issue of statistical discrimination. We begin with the case in which z and s are scalars and then consider the more general cases. Among those who are working the means of q, s, and z may depend on t although this will influence estimates of $H^*(t)$. However, we assume throughout that among those who are working the covariances among q, s, and z do not depend on t. Under these assumptions the variances and covariances involving q, s, and z and the regression coefficients Φ_{qs} and Φ_{qz} defined in (10) below do not vary with t.[7]

<u>Case 1: z is a scalar.</u>

[7] Estimates of the experience profile $H^*(t)$ will be affected if the means of s, q, and z depend on t but this has no bearing on our analysis.

The analysis is cleanest when z and s are scalars. Least squares regression will identify the parameters of the expectation of w_t on s, z, and experience profile $H^*(t)$.[8] Let b_{st} and b_{zt} be the coefficients on s and z in the conditional expectation function when t=0…T, with

(9) $E(w_t\,|s.z.t) = b_{st}s + b_{zt}z + H^*(t)$.

When the individual starts work (t is 0) this equation is

(9a) $E(w_0\,|s.z.0) = b_{s0}s + b_{z0}z + H^*(0)$

To simplify the algebra but without any additional assumptions we re-interpret s, z, and q as components of s, z, and q that are orthogonal to H^*. Then the wage process (7), the fact that $E(\Lambda v+e|D_0)$ is 0 (since there is no work history when t=0), and some straightforward algebra involving the least squares regression omitted bias formula implies that

$$(10)\quad \begin{bmatrix} b_{s0} \\ b_{z0} \end{bmatrix} = \begin{bmatrix} r+\Lambda\gamma_2+\alpha_2 \\ 0 \end{bmatrix} + \begin{bmatrix} \Phi_{qs} \\ \Phi_{qz} \end{bmatrix} x$$

where Φ_{qs} and Φ_{qz} are the coefficients of the auxiliary regression of $(\alpha_1 + \Lambda\gamma_1)q$ on s and z.. The parameters $\{b_{st}, b_{zt}\}$' are the sum of the $\{b_{s0}, b_{z0}\}$' and the coefficients of the regression of $E(\Lambda v+e|D_t)$ on s and z. That is,

$$(10a)\quad \begin{bmatrix} b_{st} \\ b_{zt} \end{bmatrix} = \begin{bmatrix} b_{s0} \\ b_{z0} \end{bmatrix} + \frac{1}{|var(s,z)|}\cdot \begin{bmatrix} \mathrm{var}(z) & -\mathrm{cov}(s,z) \\ -\mathrm{cov}(s,z) & \mathrm{var}(s) \end{bmatrix} \begin{bmatrix} 0 \\ cov(v,\, E(\Lambda v+e\,|\,D_t)) \end{bmatrix}$$

where $|Var(s,z)|$ is the determinant of Var(s,z) and we use the facts that $cov(s, E(\Lambda v+e|D_t))=0$ and $cov(z,E(\Lambda v+e|D_t))=cov(v,E(\Lambda v+e|D_t))$. This may be rewritten as

$$(11)\quad \begin{bmatrix} b_{st} \\ b_{zt} \end{bmatrix} = \begin{bmatrix} b_{s0} \\ b_{z0} \end{bmatrix} + \frac{1}{|var(s,z)|}\cdot \begin{bmatrix} \mathrm{var}(z) & -\mathrm{cov}(s,z) \\ -\mathrm{cov}(s,z) & \mathrm{var}(s) \end{bmatrix}\cdot\theta_t\cdot \begin{bmatrix} 0 \\ \Lambda var(v)+\mathrm{cov}(v,e) \end{bmatrix}$$

or

(12a) $b_{st} = b_{s0} + \theta_t\Phi_s$

(12b) $b_{zt} = b_{z0} + \theta_t\Phi_z$

where Φ_s and Φ_z are the coefficients of the regression of $\Lambda v+e$ on s and z and

---

[8] Technically, it identifies the coefficients of the least squares linear projection of w_t on s, z, and $H^*(t)$ if $E(\Lambda v+e|D_t)$ is not linear in the functions of s, z, and t we introduce in our regression models. We ignore this distinction.

$$\theta_t = \text{cov}(E(\Lambda v+e|D_t), z)/\text{cov}(\Lambda v+e, z) = \text{cov}(E(\Lambda v+e|D_t),v)/\text{cov}(\Lambda v+e, v)$$

is a parameter that is specific to the experience level t. Note that $\theta_t\Phi_s$ and $\theta_t\Phi_z$ are the coefficients of the regression of $E(\Lambda v+e|D_t)$ on s and z and that θ_t summarizes how much the firm knows about $\Lambda v + e$ at time t. It is easy to show (see Appendix 1) that $\Phi_s = -\Phi_{zs}\Phi_z$ where Φ_{zs} is the coefficient of the regression of z on s. (This is the basis of proposition 3 below.)

To determine the behavior of $\theta_t\Phi_s$ and $\theta_t\Phi_z$ over time, note first that $\Phi_s < 0$ and $\Phi_z > 0$ if cov (v, Λv+e) > 0 and cov(s,z) >0. The latter condition is true when s is schooling and the scalar z is AFQT, father's education, or the wage rate of an older sibling. The condition cov(v,Λv+e) > 0 simply states that the unobserved (by the firm) productivity subcomponent v and composite unobserved productivity term Λv+e have a positive covariance. This seems plausible to us for the z variables we consider.

The change over time in b_{st} and b_{zt} is determined by θ_t. Intuitively, θ_t is bounded between 0 and 1. It is 0 in period 0, because in this period employers know nothing about $\Lambda v + e$, so $E(\Lambda v+e|D_0)=0$. The coefficient is 1 if $E(\Lambda v+e|D_t)$ is Λv+e, since in this case the employer has learned what Λv+e is and thus knows productivity y. It is also intuitive that θ_t is nondecreasing in t because the additional information that arrives as the worker's career progresses permits a tighter estimate of Λv+e.[9]

The regularity conditions on the ϵ_t process that are required for the time average of ϵ_t to converge almost surely to 0 as t becomes large constitute sufficient conditions for θ_t to converge to 1 as t becomes large. (See Theorem 3.47 in White (1984) for a very general set of conditions.) These conditions limit the degree of independence among the ϵ_t and also restrict the variances. The intuition for this is that future values of ϵ_t must be sufficiently independent of the earlier ϵ's to average out, and must not be so variable that the future d_t values have no new information about Λv+e.[10]

A simple example may be helpful. If ϵ_t is iid with variance σ_ϵ^2, then θ_t has the familiar form

---

[9] To establish this note that since D_{t-1} is a subset of the information in D_t,
$[\text{Cov}(v, E(\Lambda v+e|D_t) - E(\Lambda v+e|D_{t-1}))]/\text{Cov}(v, \Lambda v+e) = \theta_t - \theta_{t-1} \geq 0$.

[10] To establish the result note that in each period, firms observe $d_t = \Lambda v + e + \epsilon_t$. In general, the form of $E(\Lambda v+e|D_t)$ will depend on the pattern of serial correlation and the relative variances of $\varepsilon_1 \ldots \epsilon_t$. However, the firm can always choose to use $E(\Lambda v+e|\bar{D}_t)$, where $\bar{D}_t$ is the time average $\sum_1^t (\Lambda v + e + \varepsilon_k) / t$, as an unbiased but perhaps inefficient estimator given D_t.

If as t goes to infinity $\bar{D}$ converges almost surely to Λv+e, then $\text{Cov}(E(\Lambda v+e|\bar{D}_t, v)/\text{Cov}(\Lambda v+e,v)$ converges to 1 as t goes to infinity. Since $E(\Lambda v+e|D_t)$ is more efficient than $E(\Lambda v+e|\bar{D}_t)|$, $E(\Lambda v+e|D_t)$ must also converge almost surely to Λv+e, which establishes that $\text{Cov}(E(\Lambda v+e|D_t), v) / \text{Cov}(\Lambda v+e, v)$ converges to 1. We conclude that θ_t converges almost surely to 1 as t becomes large.

$$\theta_t = \frac{\mathrm{var}(\Lambda v + e)}{\mathrm{var}(\Lambda v + e) + \sigma_\epsilon^2/t} \text{ for } t = 1, 2...; \quad \theta_0 = 0 \tag{13}$$

In this case, θ_t is strictly increasing in t because the independence among the ϵ_t means that each ϵ_t has some new information about θ_t. θ_t is 0 when t is 0 and converges to 1 as t goes to infinity.

There are two conclusions, which we summarize in Proposition 1 and 2:

***Proposition 1**: Under the assumptions of the above model, the regression coefficient b_{zt} is nondecreasing in t. The regression coefficient b_{st} is nonincreasing in t.*[11]

***Proposition 2**: If firms have complete information about the productivity of new workers, then $\partial b_{st}/\partial t = \partial b_{zt}/\partial t = 0$.*

These results underlie our empirical analysis below. Using AFQT and father's education as z variables, we examine the experience profile of b_{st} and b_{zt}. The intuition for the decline in b_{st} is that as employers learn the productivity of workers, s will get less of the credit for an association with productivity that arises because s is correlated with z provided that z is included in the wage equation with a time dependent coefficient and can claim the credit.[12] We also are able to estimate the time profile of θ_t up to scale. Under the assumption that employers learn about v and e at the same rate, this enables us to estimate the time profile of employer learning about productivity up to scale. In AP (1996) we examine the implications of our estimates for pure signaling models of the return to education.

The model also implies a third result, which we state in proposition 3.

***Proposition 3**: Under the assumptions of the above model, $\partial b_{st}/\partial t = -\Phi_{zs}\, \partial b_{zt}/\partial t$.*

Since Φ_{zs} is simply the regression coefficient of z on s and can be estimated, the coefficient restriction in Proposition 3 may provide leverage in differentiating between the learning/statistical discrimination model and alternative explanations for the behavior of b_{st} and b_{zt}.

Additional Empirical Implications

As noted in footnote 4, the literature on statistical discrimination as well as the literature on labor market networks has emphasized differences across groups in the amount of information that is available to firms (or the mapping between a given set of data and what the firm actually knows) may differ across

[11] The coefficients on an unfavorable z characteristic, such as criminal involvement or alcohol use, will become more negative to the extent that these reflect permanent traits. Assuming s is negatively correlated with the unfavorable z, b_{st} will rise with t. As noted earlier, we have normalized z so that $\Lambda > 0$.

[12] It might be particularly interesting to see if the "diploma effect" declines with t while the coefficients on hard to observe productivity characteristics that correlate with getting a diploma rise. (See Frazis (1993) for a recent analysis of whether there is a diploma effect.)

groups. Our model implies that these differences will lead to group differences in wage dynamics. To see this, suppose that there are two groups, 1 and 2. For group 2 the firm's initial information set is larger than for group one. Consequently, $\mathrm{Var}(\Lambda v + e \mid \text{group } 2) < \mathrm{Var}(\Lambda v + e \mid \text{group } 1)$ and $\mathrm{cov}(\Lambda v + e, v \mid \text{group } 2) < \mathrm{cov}(\Lambda v + e, v \mid \text{group } 1)$. From equation (10') or (11), it follows that b_{st} and b_{zt} vary less over time for group 2 than group 1. In the extreme case, when firms are fully informed about group 2, $\mathrm{cov}(\Lambda v + e, v \mid \text{group } 2)$ is 0 and b_{st} and b_{zt} are constant. In future work, it would be interesting to use this implication as a way of testing the hypothesis that the quality of information that employers have differs across labor force groups. Theories that stress differences in the ability of employers to evaluate the performance of members of different groups imply different amounts of noise (from the point of view of the employer) in the signals d_t and different paths of θ_t.

In standard labor data sets based on household interviews information on y_i is not available. However, it is interesting for at least two reasons to discuss the cross equation restrictions between the equation relating y_i to s and z that are implied by the model. First, data could be gathered from both firms and workers. Second, information on y_i or indicators of y_i may be available for use in other applications of our methods to study statistical discrimination. For example, in the study of mortgage lending, panel data on households might provide data on both credit records (related to y_i), success in loan applications (the counterpart to w_{it}), and hard to observe background variables (such as the income and wealth of relatives). Suppose that one has a measure y_{it}^* that is equal to y_i plus noise ζ_{it}. Assume that ζ_{it} is independent of all other variables in the model. Then the model implies that

$$\text{(13a)} \quad y_{it}^* = (b_{s0} + \Phi_s)\, s_i + (b_{z0} + \Phi_z) z_i + \text{error}$$

where the error term is orthogonal to s and z. Note that the coefficients are time invariant. This equation and (9) are heavily overidentified. By estimating the equations jointly one can identify θ_t separately from Φ_s and Φ_z. The availability of a productivity indicator would be particularly useful when one relaxes the assumption that the effect of s and z on y is time invariant.

Case 2: z Is a vector

We now consider the case in which z is a vector $z' = \{z_1.z_2,\ldots,z_k,..z_K)$. In this case,

$$\text{(14)} \quad \begin{bmatrix} b_{s0} \\ b_{z0} \end{bmatrix} = \begin{bmatrix} r + \Lambda\gamma_2 + \alpha_2 \\ 0 \end{bmatrix} + \begin{bmatrix} \Phi_{qs} \\ \Phi_{qz} \end{bmatrix}$$

where $[\Phi_{qs}, \Phi_{qz}]'$ is the 1 x (K+1) vector of coefficients from the auxiliary regression of $(\alpha_1 + \Lambda\gamma_1)q$ on s and z.

In the vector case $[b_{st}, b_{zt}]$ are:

$$(15)\qquad \begin{bmatrix} b_{st} \\ b_{zt} \end{bmatrix} = \begin{bmatrix} b_{s0} \\ b_{z0} \end{bmatrix} + \mathrm{var}(s,z)^{-1}\begin{bmatrix} 0 \\ \mathrm{cov}(z, E(\Lambda v + e|D_t)) \end{bmatrix}$$

where var(s,z) is a K+1 x K+1 matrix, $\mathrm{cov}(z, E(\Lambda v+e|D_t))$ is a K element vector and we have used the fact that $\mathrm{cov}(s, E(\Lambda v+e|D_t))=0$.

Let G_k be the kth row of the K x K matrix $G = [\mathrm{var}(s)\mathrm{var}(z) - \mathrm{cov}(z,s)\mathrm{cov}(z,s)']^{-1}$ where var(s) is the variance of the scalar s, var(z) is the variance of the vector z, and cov(z,s) is the vector of covariances between s and the elements of z. In appendix 1 we show that b_{st} and b_{zkt} can be expressed as

$$(16)\qquad b_{st} = b_{s0} - \sum_{k=1}^{n}\left(\mathrm{cov}(s, z_k) G_k \cdot \Theta^t_{kk} \cdot [\mathrm{cov}(z, \Lambda v + e)]\right)$$

$$(17)\qquad b_{z_k t} = b_{z_k 0} + \mathrm{var}(s) G_k \cdot \Theta^t_{kk} \cdot [\mathrm{cov}(z, \Lambda v + e)]$$

where $\Theta^t_{kk} = \mathrm{cov}(E(\Lambda v+e|D_t), z_k)/\mathrm{cov}(\Lambda v+e, z_k)$. Let $\Phi_{z_k s}$ be the coefficient of the regression of z_k on s, k=1,...,K. . Equations using (16) and (17) lead to Proposition 4, which is the vector analog to Proposition 3.

Proposition 4: *When z is a vector and the assumptions of the above model hold,*

$$\frac{\partial b_{st}}{\partial t} = -\sum_{k=1}^{K} \Phi_{z_k s} \frac{\partial b_{z_k t}}{\partial t}.$$

Proposition 2 generalizes to the vector case. Proposition 1 does not. With multiple z variables, one cannot in general sign $\partial b_{st}/\partial t$ and $\partial b_{z_k t}/\partial t$ even if all the elements of Λ are positive, each element of $\mathrm{cov}(z, \Lambda v + e)$ is positive, and $\Phi_{z_k s}$ is positive and Θ^t_{kk} is increasing in t for all k. However, from Proposition 4, it follows immediately that if $\partial b_{z_k t}/\partial t > 0$ and $b_{z_k t} > 0$ for all z_k used in the analysis then $\partial b_{st}/\partial t$ is < 0. We can verify the conditions for a particular s and set of z variables.[13]

If the Θ^t_{kk} are the same for each of the z_k and equal to the common value θ_t, then the time paths of b_{st} and the elements of b_{zt} will all be proportional to θ_t :

$$(17a)\quad \begin{bmatrix} b_{st} \\ b_{zt} \end{bmatrix} = \begin{bmatrix} b_{s0} \\ b_{z0} \end{bmatrix} + \theta_t \cdot \begin{bmatrix} -\Phi_{zs}\Phi_z \\ \Phi_z \end{bmatrix}$$

The Θ^t_{kk} will be the same for all k if the following two conditions on the conditional distribution f^t of D_t and the conditional mean of z_k hold:

Condition 1. $f^t(D_t, |z_k, \Lambda v + e) = f^t(D_t|\Lambda v + e)$ for all k ;

Condition 2. $E(z_k|\Lambda v + e) = \phi_k R(\Lambda v + e)$ where ϕ_k is a scalar that is specific to k and R is some function.

Basically, these conditions imply that the distribution of the signal D_t is driven by e + Λv and that the signal D_t is not more informative about particular elements of v than others.[14] The condition will hold if, for example, d_t is generated by (5) and e and the elements of v are normally distributed. The conditions rule out the possibility that the range of a particular element of v, say v_k, is either -100,000, 0, or 100,000. In this case, a very small or very large value of D_t would be very informative about v_k.

---

[13] Initially we were surprised that the conditions that $\Lambda > 0$, cov(z, Λv + e) > 0, and Θ^t_{kk} is increasing in t for all k do not guarantee that b_{st} is positive even if $\Phi_{zks} > 0$ for all k. The intuition is as follows. The OLS estimator of b_{zkt} is equivalent to regressing the wage in period t on the residuals $\widetilde{z}_k$ from the regression of z_k on s. $\widetilde{z}_k$ is the sum of v_k plus the component of the kth element of $\gamma_1 q$ that is orthogonal to s. The components $\gamma_1 q$ that are orthogonal to s are unrelated to v_k and e but are likely to be correlated across z_k. Consequently, using OLS to estimate b_{zkt} is analogous to applying OLS in a situation in which several of the regressors are measured with error, and the measurement errors are correlated. (The $\widetilde{z}_k$ may be thought of as noisy measures of v_k.) It is possible in such a situation for the probability limit of the OLS estimator to take on the wrong sign.

[14] To establish the conditions, note first that

$$\mathrm{Cov}((E(\Lambda v + e|D_t), z_k)) = \int_{\Lambda v + e}\int_{z_k}\int_{D_t} z_k \cdot E(\Lambda v + e|D_t) \cdot g(D_t, z_k|\Lambda v_k) \cdot h(v)$$

$$= \int_{\Lambda v + e}\int_{z_k}\int_{D_t} z_k E(\Lambda v + e|D_t) \cdot g(D_t|\Lambda v) \cdot f(z_k|\Lambda v) \cdot h(\Lambda v)$$

$$= \int_{\Lambda v + e}\int_{z_k} E(z_k|\Lambda v) \cdot E(E(\Lambda v + e|D_t)|\Lambda v) \cdot h(\Lambda v)$$

and that

$\mathrm{Cov}(\Lambda v + e, z_k) = \int_{\Lambda v + e} E(z_k|\Lambda v + e) \cdot (\Lambda v + e) \cdot h(\Lambda v + e)$. It is easy to verify from these equations that Cov(E(Λv + e|Dt, z_k)/Cov(z_k|Λv + e) is the same for all z_k if $(E(z_k|\Lambda v + e) = \phi_k \Psi(\Lambda v + e)$.

These conditions are quite strong. For example, if the firm obtains indicators about subcomponents of v and e as well as y, then it is likely to learn about some components of productivity faster than others. In this case, equation (16) and (17) continue to hold, but the time path of the education slope is a weighted average of the Θ^{t}_{kk} that determine the time paths of the individual z_k. The paths of the individual z_k will reflect differences across z_k in the rate at which firms learn about the productivity components that they are correlated with. This is an important result, because it states that differences in the effects of particular variables on wage growth may reflect differences in the rate at which firms learn about the variables. This provides an alternative or a complement to the standard view that the differential effects on growth rates reflect differences in the relationship between the variables and other sources of wage growth such as on-the-job training.

<u>Case 3: s and z are both vectors.</u>

Finally, we consider the case in which both s and z are vectors. In this case we reinterpret all of the of the related variables and parameters in the model, such as b_{s0}, b_{z0}, v, etc as vectors or matrices. The vectors of coefficients b_{s0} and b_{z0} on s and z in the base year satisfy (14) where the vectors Φ_{qs} and Φ_{qz} are the coefficients in the regression of q on s and z. The vectors b_{st} and b_{zt} are given by

(17b) $$\begin{bmatrix} b_{st} \\ b_{zt} \end{bmatrix} = \begin{bmatrix} b_{s0} \\ b_{z0} \end{bmatrix} + A\left[\operatorname{cov}(v, E(\Lambda v + e | D_t)\right] \quad \text{where}$$

$$A = \begin{bmatrix} -\operatorname{var}(s)^{-1}\operatorname{cov}(s,z)\left[\operatorname{var}(z) - \operatorname{cov}(z,s)\operatorname{var}(s)^{-1}\operatorname{cov}(s,z)\right] \\ \left[\operatorname{var}(z) - \operatorname{cov}(z,s)\operatorname{var}(s)^{-1}\operatorname{cov}(s,z)\right] \end{bmatrix}$$

Since $-\operatorname{var}(s)^{-1}\operatorname{cov}(s,z)$ is the matrix of coefficients from the regression of z on s we obtain the vector version of proposition 4:

$$b_{st} - b_{s0} = \Phi_{zs}(b_{zt} - b_{z0}),$$

where Φ_{zs} is redefined to be the matrix of coefficients of the regression of the vector z on the vector s. When conditions 1 and 2 are satisfied and the signal D_t is not more informative about particular elements of v than others then (17b) reduces to (17a) with both b_{st} and b_{zt} as vectors.

Statistical Discrimination on the Basis of Race

Firms observe race. If race is correlated with productivity and firms violate the law and use race as information, then race has the properties of an s variable. To see the empirical implications of this, partition s into two variables, s_1 and s_2, where s_1 is an indicator variable for membership in a particular racial group and 0 otherwise, and s_2 is schooling.[15] In this case, the model implies almost immediately that the coefficient on s_1 does not vary over time if the interaction between z and t is excluded from the model. If this interaction is included (17a) implies that the time paths of b_{s1t} and b_{s2t} are

$$b_{s1t}=b_{s10} - \Phi_{zs1}\theta_t ,$$

$$b_{s2t}=bs_{20} - \Phi_{zs2}\theta_t$$

where Φ_{zs1} and Φ_{zs2} are the coefficients on s_1 and s_2 in the regression of z on s_1 and s_2. Assuming Φ_{zs1} is negative, as it is when s_1 indicates that the person is black and z is AFQT, father's education, or the wage of an older sibling, then the wage coefficient on s_1 will rise over time.

In contrast, if firms obey the law and do not use race as information, then in the econometric model, race has the properties of a z variable. In the case in which race is the only z variable and one s variable, such as education, is included in the analysis, then the coefficient on z in equation (11) corresponds to the coefficient on race. The model implies that if (i) race is negatively related to productivity ($\Lambda < 0$), (ii) firms do not statistically discriminate on basis of race, and (iii) firms learn over time, then the race differential will widen as experience accumulates. The intuition is that with learning firms are acquiring additional information about performance that may legitimately be used to differentiate among workers. If race is negatively related to productivity, then the new information will lead to a decline in wages. If education is negatively related to race, then the coefficient on education should fall over time.

What happens if firms do not discriminate on the basis of race and one adds a second z variable with a time varying coefficient to a model that contains race and an s variable? Let z_1 denote race and z_2 denote the additional variable, and let b_{z1t} denote the coefficient on race when experience is t and z_2 is included in the model and let b_{z1t}^* denote the corresponding coefficient when

[15] The element of r corresponding to the race indicator s_1 in the productivity equation (1) is 0 unless consumer or employee tastes for discrimination reduce profitability of employing members of the minority group, as in Becker (1971). (Even if r is 0 race may be negatively related to productivity if it is correlated with elements of z, q, or η that affect productivity.) Presumably, firms that violate the law and discriminate in response to their own prejudice or the prejudice of consumers or other employees might also be willing to use race as information. Employers who harbor prejudice against certain groups may be especially unlikely to form beliefs about the productivity of those groups that are rational in the statistical sense used in this paper.

z_2 is excluded. Assume that $\Theta_{11}^t = \Theta_{22}^t = \theta_t$ where Θ_{kk}^t is defined below (17) above. In Appendix 3 we show that

$$\partial b_{z1t} / \partial t - \partial b_{z1t}^{*} / \partial t = -\partial \theta_t / \partial t \cdot [\Phi_{z_2} \Phi_{z_2 z_1}]$$

where Φ_{z2} is the coefficient on z_2 in the regression of $\Lambda v + e$ on s, z_1 and z_2 and $\Phi_{z_2 z_1}$ is the coefficient on z_1 in the regression of z_2 on z_1 and s. When z_1 indicates whether the person is black and z_2 is AFQT, father's education, or the wage of an older sibling, Φ_{z2z1} is negative. If these variables are positively related to productivity, with $\Phi_{z_2} > 0$ then $\partial b_{z1t} / \partial t - \partial b_{z1t}^{*} / \partial t > 0$. We conclude that if firms do not statistically discriminate on the basis of race and race is negatively related to productivity, then (1) the race gap will widen with experience and (2) adding a favorable z variable to the model will reduce the race difference in the experience profile. We wish to stress that other factors that influence race differences in experience profiles as well as other forms of discrimination will also influence the wage results.

2.2 Incorporating On-the-Job Training Into the Model:

The analysis so far assumes that the effects of z and s on the log of productivity do not depend on t. Human capital accumulation is included in the model through the H(t) function but is assumed to be "neutral" in the sense that it does not influence the time paths of the effects of s and z.[16] In the more general case, the time paths of z and s depend on other factors as well as learning. In this section we first consider the effect that such dependence would have on OLS estimates of the interactions between t and z and s. Then we discuss estimation of a more realistic model that includes both human capital accumulation and learning/statistical discrimination. As we shall see, there is no clean way to sort out the relative roles of these two mechanisms without data on productivity.

Suppose that s is complementary with learning by doing or enhances the productivity of investments in general skills. We return to the case of scalar z. Then the productivity equation (net of training costs) might take the form

(18) $y_t = r\,s + r_1 s\,t + H(t) + \alpha_1 q + \Lambda z + \eta$.

[16] One may easily modify the theoretical framework to allow for this form of human capital accumulation. For example, the H(t) function may reflect learning by doing in all jobs that is observable to firms, or worker financed investments in human capital that are observable to firms.

Assuming that the training activity is observed (firms know (18)) and workers pay for the general training, the wage equation (9) becomes

(19) $w_t = (b_{st} + r_1 t)\, s + b_{zt} z + H(t) + \alpha_1 q + E(\Lambda v + e|D_t)$

Most discussions of human capital and most of the empirical evidence on employer provided training suggest that education makes workers more trainable and that educated workers receive more training. In this case r_1 will be greater than 0.[17] Probit models of the probability that a worker receives training in a given year show strong positive effects of schooling, and AFQT as well as smaller but positive, statistically significant effect of father's education. (See below.)

What are the implications of this for our investigation of the hypothesis that the reliance of employers on easily observable variables to estimate productivity declines over the career? In estimating the model we identify the sum $b_{st} + r_1 st$ rather than b_{st}. If r_1 is greater than 0, then the estimated relationship between $b_{st} + r_1 st$ and t will be biased against the hypothesis that employers learn about productivity. As it turns out, we find a strong negative relationship between $b_{st} + r_1 st$ and t, which is only consistent with a training interpretation if education reduces learning by doing, the productivity of training investments, and/or the quantity of training investments.

There is also the possibility that the productivity of employer provided training and/or learning by doing depends on z and/or η. This case is harder to analyze because employers do not observe z and η directly and are learning more about them as time goes on. As a start, we consider the extreme case in which firms are fully informed about z, so that θ_t is 1 and b_{zt} in (9) is a constant in the absence of training.

Suppose that the productivity equation is

(20) $y_t = r\, s + r_1 s\, t + r_2 z\, t + H(t) + \alpha_1 q + \Lambda z + \eta, \qquad r_2 > 0$

If firm's knowledge of s and q is fully informative about z, then the presence of r_2 in the productivity equation should lead the effect of z on the wage to rise with experience even if b_{zt} does not depend on time ($\theta_t = 1$). However, the presence of $r_2 z\, t$ in the productivity equation seems unlikely to lead to a negative estimate of $\partial b_{st}/\partial t$.

It is important to point out, however, that if the effect of z on y rises with t then introducing the interaction between z and t into the wage equation could lower the estimate of the change over

[17] Earnings slopes depend on the expected productivity of the worker if the costs or returns to training depend on variables such as z or s. Altonji and Spletzer (1992) find a relationship between test scores and measures of training using the NLS72 data set, and many studies find a link between schooling and training measures. See for example, Bartel and Sicherman (1992) and Lynch (1992).

time in the wage response to s. Let B_{st} be the expectation of the OLS estimator of the effect of s on the wage in period t. Then B_{st} is $b_{s0} + r_1 t + \Phi_{zs} r_2 t$, where Φ_{zs} is the coefficient of the regression of z on s. When one adds $z \cdot t$ to the regression, B_{st} becomes $b_{s0} + r_1 t$. B_{zt}, the expectation of the OLS estimator of the effect of z in period t, becomes $b_{z0} + r_2 t$. If $\Phi_{zs} > 0$ and $r_2 > 0$, then $\Phi_{zs} r_2 > 0$. The change in the coefficient on s when z t is added is $-\Phi_{zs} r_2 t$. Consequently, in the scalar case the simple training model with full information about z implies that $[\partial B_{st}/\partial t]$ declines by $-\Phi zs\ [\partial B_{zt}/\partial t]$ when z t is added to the wage equation.

In the pure employer learning/statistical discrimination model $\partial B_{st}/\partial t$ is equal to $\partial b_{st}/\partial t$ and, according to proposition 3, the learning model also implies that $\partial B_{st}/\partial t$ declines by $-\Phi_{zs}\partial b_{zt} = -\Phi_{zs}\partial t$ when z t is added to the wage equation. However, models differ in their implications for the *level* of $\partial B_{st}/\partial t$ after z t is added. A pure human capital model with perfect information implies $\partial B_{st}/\partial t \geq 0$ unless, in contrast to the available evidence, s has a negative partial effect on the quantity or return to on-the-job training ($r_1 < 0$).

Controlling for Training.

In the absence of data on productivity, sorting out the relative importance of employer learning and non neutral (with respect to z and s) on-the-job training may require that one build a model of the quantity of training as a function of s and z and use a proxy based on the training model to control for the effects of non neutral general human capital accumulation in the wage equation. This raises a number of difficulties that we explore in the next few paragraphs.

We return to the case of scalar z. Assume that the productivity equation (net of training costs) takes the form

(21) $y_t = r\,s + \Psi(\Sigma T_{i\tau}) - C(T_{it}) + H(t) + \alpha_1 q + \eta$.

where $\Sigma T_{i\tau} = \Sigma_{\tau=1..t} T_{i\tau}$, $\Psi(.)$ is an increasing function that summarizes the effect of accumulated training on productivity, and $C(T_{it})$ is the cost in terms of the log of productivity of T_{it} units of training in period t, and the function H(t) has been redefined to accommodate the inclusion of training. Assume that T_{it} is determined by employer beliefs about productivity given D_t, q, s, and t, as well as by D_t, q, s, and experience. Then

(22) $T_{it} = h(D_t, q, s, t) = r(s,z,t) + u_t$

where $r(s,z,t)$ is $E(h(D_t, q, s, t)|s,z,t)$ and u_t is an error term that is related to q and D_t but is assumed independent of s, z, and t. Following through on a series of substitutions that parallels those leading to (8), and assuming that the worker pays for and receives the returns to the general training yields the wage equation

$$(23) \quad w_t = (r + \gamma_2 + \alpha_2)s + \Psi(\Sigma T_{i\tau}) - C(T_{it}) + H^*(t) + \Lambda(\gamma_1 + \alpha_1)q + E(\Lambda v{+}e\,|Dt) + \varsigma_t$$

Suppose that up to an irrelevant constant $\Psi(\Sigma_{\tau=1..t} T_{i\tau})) = \psi_1 \Sigma_{\tau=1..t} T_{i\tau}$ and $C(T_{it}) = c_1 \; T_{it}$.

Then the regression function relating w_t to s, z , $\Sigma T_{i\tau}$, and T_{it} in period t may be written as

$$(24) \quad w_t = (r + a_{1t})s + (\psi_1 + a_{2t})\Sigma T_{i\tau} + (a_{3t} - c_1)T_{it} + a_{4t} z + H^*(t) + \epsilon_{wt}$$

where a_{1t}, a_{2t}, a_{3t}, and a_{4t} are the coefficients of the linear least squares projection of $\Lambda(\gamma_1 + \alpha_1)q$ + $E(\Lambda v{+}e\,|D_t)$ onto s, $\Sigma T_{i\tau}$, T_{it} , and z, and the error term ϵ_{wt} is unrelated to the variables in the model by definition of a_{1t} a_{2t}, a_{3t}, and a_{4t}. The time path of a_{1t} and a_{4t} will be influenced by changes over time in the correlations of s, z, and $\Lambda(\gamma_1 + \alpha_1)q$ with $\Sigma T_{i\tau}$ and T_{it} as well as changes over time in the correlations of z, $\Sigma T_{i\tau}$, and T_{it} with $E(\Lambda v{+}e\,|Dt)$. (The coefficients of the experience profile $H^*(t)$ will be influenced as well.)

Two implications follow from (23) and (24). First, even if training depends only on information that is known to the firm at the start, the relationship between q and T_{it} and $\Sigma T_{i\tau}$ may change with t, leading to changes over time in the coefficient on s even if there is no learning. The second point follows from the fact that training depends on D_t and so will be correlated with it. The least squares estimates of the coefficients on the training variables will reflect both the direct effect of training and a relationship between the time path of T_{it} and $E(\Lambda v{+}e\,|D_t)$. As a result, the effect of adding the training variables to the model on b_{st} and b_{zt} is complicated in a mixed human capital/employer learning model. In particular, one might expect the addition of functions of T_{it} and $\Sigma T_{i\tau}$ to the model to change and quite possibly reduce the rate of increase of b_{zt}, for two reasons. First, the training variables change over time and are positively correlated with z. Second, they will absorb part of the trend in $E(\Lambda v{+}e\,|D_t)$, and it is changes in this term that induce the variation with t in b_{st} and b_{zt}. Furthermore, the introduction of the training terms alters the partial correlation between z and s, which changes the effect on the path of b_{st} of introducing z with a time varying coefficient. Unfortunately, we have do not have a way to isolate the effects of training from the effects of statistical discrimination with learning if, as seems plausible, the quantity of training is influenced by the employer beliefs about productivity. Consider the null hypothesis that (1) learning is important, (2) variation with s and z in the rate of skill accumulation is not, and (3) variation in our measure of training is driven by worker performance (which leads to promotion into jobs that offer training) rather than by exogenous differences in the level of human capital investment. Even under this hypothesis one would expect the introduction of the training measures to lead to a reduction in the growth over time in the coefficient on z and a reduction in the impact of z on the time path of the coefficient on s. With an indicator of y_{it}, that problem is easily solved, but we lack such an indicator.

Despite the absence of a clear structural interpretation of the results we think it is important in this initial study to see how introducing measures of training alters b_{st} and b_{zt}. Consequently, below we report estimates of (24). There are two additional problems in using the training data. First, the measure T^*_{it} of T_{it} is almost certain to contain measurement error. Second, the quality of the training data prior to 1988 is too poor to be used, which means that the data needed to form the measure $\Sigma T_{i\tau}$ is missing for persons who left school prior to that year. We do not have a solution for the first problem but deal with the latter problem by estimating a flexible model relating T^*_{it} to s, z, and t using data from 1988-1993 and using the model to impute values in the earlier years.[18] We estimate variants of (24) below. Our preferred specification is a wage growth model based on the first difference of (24). The growth specification has the advantage of only requiring data on T_{it} and T_{it-1}. Perhaps more importantly, this specification also eliminates bias from unobserved person specific effects that are known to firms and are correlated with both training and wages.

3. Data

The empirical analysis is based on the 1992 release of NLSY. The NLSY is a panel study of men and women who were aged 14-22 in 1978. Sample members have been surveyed annually since 1979. (In 1994 the NLSY moved to a biannual survey schedule.) The NLSY is an attractive data set for the study of employer learning and statistical discrimination. First, the sample sizes are large. Second, sample members are observed at or near the start of their work careers and are followed for several years. Third, the NLSY contains detailed employment histories, including reasons for job changes. Fourth, it contains a rich set of personal characteristics that may be related to productivity and may be hard for employers to observe, including father and mother's education

[18] Spletzer and Lowenstein (1996) provide means of dealing with measurement error in the training data but these are beyond the scope of our study.

and occupation, drug and alcohol use, criminal activity, AFQT, aspirations and motivation, and performance in school. Furthermore, the data set contains a large number of siblings. The earnings of older siblings as well as parents may be used as indicators of characteristics of younger siblings that affect productivity but are hard for employers to observe. Finally, it contains measures of training, which we need to investigate the possibility that variation with experience in the effects of schooling and our measures of hard to observe personal characteristics are due to a relationship between these variables and the quantity of training received.

We restrict the analysis to men who are white or black who have completed 8 or more years of education. We exclude labor market observations prior to the first time that a person leaves school and accumulate experience from that point. When we analyze wage changes, we further restrict the sample to persons who do not change education between successive years. Actual experience is the number of weeks in which the person worked more than 30 hours divided by 50. Potential experience is defined as age minus years of schooling minus 6. To reduce the influence of outliers, father's education (F_ED) is set to 4 if father's education is reported to be less than 4. AFQT is standardized by age.[19] The means, standard deviations, minimum and maximums of the variables used in analysis are provided in Table A1 in the Appendix , along with the variable definitions. The mean of actual experience is 4.9. The mean of potential experience is 7.3, and the mean of education is 12.7. All statistics in the paper are unweighted. Blacks are over sampled in the NLSY and contribute 28.8 percent of our observations. Table A2 reports correlation coefficients and simple regression coefficients that summarize the relationships among the key variables used in the analysis.

4. Results for Education

In Table 1-3 we report estimates of our basic wage level specification. In table 1 we use potential experience as the experience measure and use OLS to estimate the model. The equations also control for a cubic in experience, a quadratic time trend, residence in an urban area, and dummy

---

[19] The age of the sample members at the time the AFQT was administered varies somewhat in the NLSY sample. This induces some variation in schooling levels at the time the AFQT is taken. To calculate standardized AFQT, we adjust the raw AFQT score by subtracting the mean score for each age and dividing by the standard deviation for that age. For individuals with siblings in the sample, the coefficients of the regression of the unadjusted test score of the older sibling on the test score of the younger sibling and the regression of the test score of the younger sibling on the the score of the older sibling are very similar after one also controls for age, suggesting that the information in the test is not very sensitive to age over the range in the sample.

variables for whether father's education is missing and whether AFQT is missing. We add interactions between the dummy variables for missing data and experience when interactions between father's education and experience and AFQT and experience are added to the model. These variables are not reported in the tables. Standard errors are White/Huber standard errors computed accounting for the fact that there are multiple observations for each worker.

In column 4 we present an equation that includes s, Black, and s×t. This corresponds to (7a) with b_{st} restricted to $b_{st} = b_{s0} + b_{s1}\times t$. The coefficient on s×t/10 is -.0077 (.0062), suggesting that the effect of education on wages declines slightly with experience. In column 5 we add AFQT and F_ED, where F_ED is years of father's education. As had been well documented, AFQT has a powerful association with earnings even after controlling for education. A shift in AFQT from one standard deviation below the mean to one standard deviation above is associated with an increase in the log wage of .157. The coefficient on education declines to .080 and b_{s1} becomes more negative.

In column 6 we add linear interactions between t and two different z variables, AFQT and F_ED, to the equation. The resulting equation corresponds to (9) with the restriction that $b_{st} = b_{s0} + b_{s1}\times t$ and $b_{zt} = b_{z0} + b_{z1}\times t$, except that we introduce two z variables rather than 1. The estimates imply that the effect of AFQT on the wage increases greatly with experience t. b_{AFQT1}, which is the coefficient on AFQT×t/10, is .0820 (.0125). b_{AFQTt}, which is $\partial wt/\partial AFQT$, rises from only .0179 when experience is 0 to .0999 when experience is 10. The results imply that when experience is 10 and education is held constant, persons with AFQT scores one standard deviation above the mean have a log wage that is .200 larger than persons with AFQT scores one standard deviation below the mean, while the difference is only .036 when experience is 0. The effect of father's education also increases with experience. The main effect is actually slightly negative (but not significant). However, the interaction term is positive, though not statistically significant.

Our results for AFQT and F_ED are consistent with Farber and Gibbon's results in which they use the components of AFQT and an indicator for whether the family had a library card when the person is 14 that are orthogonal to the wage on the first job and education. The key result in the table is that the coefficient on s×t/10 declines sharply (to -.0351 (.0069)) when AFQT×t and F_ED×t are added. The implied effect of an extra year of education for a person with 10 years of experience is only .0633. Strikingly, the coefficient on s rises to .0984 which is almost exactly what we obtain when we exclude all terms involving F_ED and AFQT from the model (columns 1 and 4).

These results provide support for the hypothesis that employers have limited information about the productivity of labor force entrants and statistically discriminate on the basis of education. Early wages are based on expected productivity conditional on easily observable variables such as education. As experience accumulates, wages become more strongly related to variables that are likely to be correlated with productivity but hard for the employer to observe directly. When we condition the experience profile of earnings on both easy to observe variables, such as education, and hard to observe variables, such as AFQT and father's education, we find the partial effect of the easy to observe variables declines substantially with experience. While one might argue that the positive coefficients on AFQT×t and F_ED×t are due to an association between these variables and training intensity, it is hard to reconcile this view with the negative coefficient on s×t. While measurement error in schooling may enhance the effect of F_ED and AFQT and may partially explain the decline in s between columns 1 and 3, it does not provide a simple explanation for the behavior of the interaction terms with experience.

In Table 2 we present OLS results using actual experience in place of potential experience as the experience measure t. The main difference between this table and table 1 is that the return to education is lower and the s×t interaction is positive and fairly large in the equations that exclude AFQT×t and F_ED×t. However, the coefficient on s×t/10 declines from .0200 in column 5 to -.0056 when the interaction terms are added in column 6 of Table 2. This decline is similar to the decline that we obtain in column 3.

The results in Table 2 are difficult to interpret, because the intensity of work experience may be conveying information to employers about worker quality. It is an outcome measure itself. The implications of employer learning for the wage equation are changed if one conditions on information that becomes available to employers as the worker's career unfolds and may reflect the productivity of the worker. Conditioning on actual work experience raises some of the issues that would arise if we conditioned on wages in t-1 or on training received. On the other hand, the results based on potential experience are likely to be biased by the fact that potential experience mismeasures actual. For this reason, in Table 3 we report the results of re-estimating the models by instrumental variables (IV), treating all terms involving actual experience as endogenous with corresponding terms involving potential experience as the instruments. The results in columns 5 and 6 of Table 3 are basically consistent with those in Table 1. The coefficient on AFQT is .0177 (.0096) and the coefficient on AFQT·t/10 is .1148 (.0164). These estimates imply that conditional on years of schooling, AFQT has only a small effect on initial wages, but when t is 10 a two standard deviation shift in AFQT is associated with a wage differential of .247. The coefficient on $s \cdot t/10$ declines from -.0181 when the interactions are excluded in column 5 to -.0561 in column 6.

Controlling for Secular change in the Return to Education

In column 9 of Tables1, 2, and 3 we add the interaction between s and calendar time to the model containing father's education and AFQT. [20] In the case of potential experience in Table 1, the education slope is reduced by .02 per year, and the interaction between education and experience/10 drops to -.051, but otherwise the results change little. In column 10 we add the interactions between calendar time and s, F_ED, and AFQT to the model containing the interactions between t and all three variables. In column 10 the interactions between F_ED and AFQT and calendar time have positive coefficients, indicating that the effects of these variables rose during the 1980s. Adding the time interactions reduces the size of the experience interactions with F_ED and AFQT, but the qualitative pattern of the results does not change.

Controlling for Occupation

One objection to the theoretical framework underlying the estimates in Tables 1-3 is that it assumes that the flow of information to employers is independent of the type of job the worker begins in. This is contrary to the idea that some jobs are "dead end" jobs. Perhaps education (and high AFQT) enables a worker to gain access to jobs in which firms have the ability to observe whether the worker has higher level skills that are strongly related to productivity. As a simple check on this possibility, we present a series of equations in Table 4 that control for the 2-digit occupation of the first job. The results are very similar to what we obtain when occupation is excluded.[21].

The Effects of the Wage of a Sibling

[20] Murphy and Welch (1992), Katz and Murphy (1992), Taber (1996) and Chay and Lee (1997) are among a large number of recent study of changes in the structure of wages in the U.S.. Since calendar time is positively correlated with experience t in a panel data set, the learning/statistical discrimination model implies that estimates of secular changes in the return to education and AFQT will be biased in opposite directions if one fails to add the interaction between these variables and t to the model.

[21] An interesting project for future research would be to use information from the Dictionary of Occupational Titles on skill requirements of occupations and trace how easy to observe and hard to observe productivity characteristics are related to changes over a career in the skill requirements of the job a worker holds. It would also be interesting to examine how the slopes are influenced by the skill requirements of the initial occupation held by the individual.

In Table 5, we use the wages of siblings with 5 to 8 years of experience as a hard to observe background characteristic. The coefficient on $s \times t/10$ is -.0097 (.0089) in column 4, which includes the log of the wage of the oldest sibling. The learning model does not provide an explanation for the negative interaction term, nor does the conventional view of how education is related to on-the-job training. However, when we add the interaction between the sibling wage and t in column 5, the coefficient on the education interaction falls to -.0146, and the coefficient on the interaction between the sibling wage and t/10 is .086 (.0327).[22] The effect of the sibling wage rises from .127 upon labor force entry to .213 after 10 years of experience---a very large increase. The point estimate of the interaction between education and experience result is essentially unchanged when we allow the effect of sibling wage. In Table 5, columns 5 and 6, we show that these results are robust to allowing the effects of education and the sibling wage to depend on calendar time. Our interpretation of these results begins with the premise that the labor market productivity of siblings are correlated. As a worker acquires experience this correlation is reflected in the performance record D_t and in wage rates. The sibling wage is correlated with education, and so the effect of education on the wage declines with experience because firms are estimating productivity with a bigger information set than at the time of labor force entry.[23]

The Experience Profile of the Effects of AFQT and Education on Wages

In this section we take a more detailed look at how the effects of AFQT and s vary with experience by estimating models of the form

$$w_t = f(z,t;b_z) + h(s,t;b_s) + H(t) + \epsilon_{it}$$

where b_z and b_s are now vectors of parameters. Table 6 is based on models in which $f(z,t;b_z)$ and $h(s,t;b_s)$ are quartic polynomials in t. In the top panel, the experience measure is potential

---

[22] The corresponding point estimates are -.022 and .080 when we allow the effects of education and the sibling wage to depend on calendar time.

[23] Farber and Gibbons (1996) use men and women, include Hispanics, and restrict their sample to persons who have worked at least three consecutive years since attending school. Using this sample the coefficients on $AFQT \times t$ and the effect on $s \times t$ of adding $AFQT \times t$ are similar to those reported above. We also obtain qualitatively similar results when we follow Farber and Gibbons and use the level of wages rather than the log. We experimented with an indicator for whether any person in the respondent's household had a library card at the time the respondent was 14, a variable which Farber and Gibbons also used. We confirm Farber and Gibbons' finding that the coefficient on the residual from a regression of this variable on the initial real wage, education, part-time status, an interaction between education and part-time status, race, sex, age, and calendar year increases with experience, as well as their finding that the results for library card and AFQT are weakened substantially when these variables are interacted with calendar time. However, when we use the library card variable itself the effect of the library card variable falls rather than rises with experience. We thank Henry Farber for assisting us in reconstructing the Farber and Gibbons sample.

experience; in the bottom panel we use actual experience instrumented by potential experience. All of the models in the tables contain the other control variables discussed above. They also include F_ED and F_ED×t.

The columns report $\partial w_t/\partial AFQT$, $\partial^2 w_t/\partial AFQT, \partial t$ $\partial w_t/\partial s$, and $\partial^2 w_t/\partial s\partial t$ at various experience levels. The first column of the table shows that $\partial w_t/\partial AFQT$ increases steadily from .0197 when t is 0 to .121 when t is 12. (We only go out to t=12 because sample information becomes thin at higher values.) The specification that we use in most of the paper, in which $f(z,t;b_z)$ and $h(s,t;b_s)$ are linear in t (column 6 in tables 1-3), suggests an increase in $\partial w_t/\partial AFQT$ from .0179 to .116 as t goes from 0 to 12.

As noted earlier, employer learning implies that $\partial w_t/\partial AFQT$ is nondecreasing in t (i.e., $\partial^2 w_t/\partial AFQT,\partial t \geq 0$), with a strict inequality likely if some new information arrives each period on y. If the noise in observations of y_t are iid, then the rate of increase $\partial^2 w_t/\partial AFQT,\partial t$ should decline with t, as shown in expression (12c) for θ_t above. The rate of increase must decline eventually because the amount of additional information in observations of labor market performance is declining. (θ_t is bounded at 1.) However, it is possible that the first two or three observations on a worker are particularly noisy because of factors that we have left out of the model. For example job specific or occupation specific match quality may be more variable for new workers than more experienced ones.

In column 2 we report $\partial^2 w_t/\partial AFQT,\partial t$ for various experience levels. The values increase from .0025 when t is 0 to .0104 when t is 5, remains at about this level until t is 8 (the maximum is .0108 at t = 6.5) and then decline to .0048 when t is 12. These results are reasonably consistent with a decline in the amount of new information with experience after an initial period of noisy observations.[24]

In panel B we replace potential experience with actual experience, and treat actual experience as endogenous. The 99th percentile value for this variable is only 13.33, so there not much sample information on t beyond this point. In column 1 we see that the effect of AFQT increases with experience. The rate of increase $\partial^2 w_t/\partial AFQT,\partial t$ rises at first from .0092 when t = 0 to .0138 when t=5, but declines to -.0012 when t = 12. However, the standard errors on these

[24] We used two other non-linear specifications. The first used spline functions with break points at t=2, t=4, t=7,and t=10. In the second we restricted $f(z,t;b_z)$ so that $\partial^2 w_t/\partial AFQT\ \partial t = 0$ when t is 25 and $h(s,t;b_s)$ so that $\partial^2 w_t/\partial s\ \partial t = 0$ when t is 25. The idea is that the information about productivity that is contained in AFQT is fully revealed by the time t is 25. Both of these specifications yielded results similar to the reported model in which $\partial^2 w_t/\partial AFQT\ \partial t$ is flat or increasing and then definitively decreasing after about 7 years.

derivatives are quite large. These results are also loosely consistent with the proposition that the rate at which new information about initial productivity arrives declines with experience, but the estimates are not sufficiently precise to say much about this. As the NLSY sample ages, it will be interesting to revisit the issue.

In the model with potential experience, the return to education increases slightly between t=0 to t=3, and then declines sharply. In the model with actual experience, the decline is constant throughout from .0881 at no experience to .0299 at 12 years of experience.

Testing the restrictions on the experience profiles of the effects of s and z on the wage.

It is interesting to see how well the experience profiles of the education and AFQT coefficients satisfy the restrictions in propositions 3 and 4. One complication in performing these tests is the place of race within our model --- should we treat race an s variable or a z variable? The answer to this question hinges on the extent to which employers violate the law and use race as an indicator of productivity. We discuss this at length in section 5 below. For now we will side step the issue by running separate tests on the white and black samples. Consider first a specification in which s and z are both scalars, education and AFQT score, respectively. Proposition 3 says that the product of -cov(s,z)/var(s) --- the negative of the coefficient of the regression of z on s --- times the coefficient on the interaction between AFQT and experience ($z \times t$) should equal the coefficient on the interaction between education and experience ($s \times t$). In the white sample, the product is -.00162 and the coefficient on $s \times t$ is -.00232. A Wald test does not reject the proposition. In the black sample the corresponding numbers are -.00196 and -.00498 and the proposition is rejected.[25]

We might also want to test whether the entire profile of the interactions between s and t and between z and t are in accordance with proposition 3. One way to do this is to estimate the model in which the interactions are specified as fourth-order polynomials and jointly test whether the coefficients on the four polynomial interactions are related by the coefficient of the regression of z on s. This seems a bit restrictive in that we only expect the relationship to hold over the range of observed data and polynomials that have very different coefficients can be fairly similar over a short range. However, we performed these tests on models in which the interactions of AFQT and education with experience are modeled as fourth order polynomials. Once again, we fail to reject the proposition for whites but reject for blacks.

[25] It should be noted that the standard errors for these tests do not account for possible heteroscedasticity in the data.

We also tested proposition 4, the vector analog of proposition 3, on models which include both AFQT and father's education. We also considered as z variables the dummy variables indicating whether these quantities were known. This test amounts to a t-test of whether sum of the products of -cov(s,z)/var(s) and the coefficient on z×t for each z variable is equal to the coefficient on s×t. For whites, the sum of the products equals -.00193, the coefficient on s×t is -.00254, and the proposition is not rejected. For blacks, we obtain -.00166 and -.00456 and reject the proposition.

Wage Growth Equations

In Table 7 we estimate (9) in first difference form. We restrict b_{st} to be $b_{s0} + b_{s1}$ t and b_{zt} to be $b_{z0} + b_{z1}$ t. The usual reason for working in first differences is to eliminate correlation between the regressors and a fixed error component. This motivation is not compelling in the present case. However, it is possible that the first difference specification may be less sensitive to errors in identifying when individuals start their careers.

Columns 1-4 report OLS estimates with potential experience. The coefficient on the s×Δt will pick up the effects of secular changes in the return to education as well as the changes with experience in the return to education. The upward secular trend in the return to schooling may partially explain the fact that the s×Δt has a positive coefficient in the basic model in column 1 while it is negative for the corresponding level specification in Table 1, column 4.[26] (A secular trend in the return to education or AFQT matters less when estimating the equations in levels because much of the variation in experience is across persons of different ages). Also, the estimates are much less precise when we estimate in first difference form. However, the key results are qualitatively similar to the level specifications. In particular, the coefficient on s/10 declines from .0148 (.0094) in column 1 to -.0092 (.0110) when we add the AFQT and F_ED interaction terms in column 2. The size of the decline in this coefficient is very similar to the drop in the coefficient on s×t when we add AFQT×t and F_ED×t to the level specifications. (See columns 5 and 6 of Table 1). The AFQT interaction term is positive with a t value of 3.4. The F_ED interaction is also positive and similar in magnitude to the result obtained in levels, but it is not statistically significant.

Columns 5-8 reports IV estimates of wage growth equations using actual experience as the experience measure. The coefficient on AFQT×Δt/10 is .0905 (.0197), which compares to the value

[26] See Murphy and Welch (1991) and many subsequent studies. Murnane et al (1995) provide evidence of an increase in the return to aptitude and achievement, as measured by tests.

of .1148 in Table 3, column 5. The coefficient on $s \times \Delta t/10$ declines from .0295 (.0079) to -.0030 (.0100) when $AFQT \times \Delta t/10$ and $F\_ED \times \Delta t/10$ are added.

5. Do Employers Statistically Discriminate on the Basis of Race?

Thus far we have focused the discussion on employers' use of education as an indicator of labor market productivity. In this section we examine the role of race. By almost any measure, young black men are disadvantaged relative to whites in the U.S.. On average, black males have poorer, less educated parents, are more likely to grow up in a single parent household, live in more troubled neighborhoods, attend schools with fewer resources, and have fewer opportunities for teenage employment than white males. Many of these factors are correlated with educational attainment and labor market success. They are likely to lead to a black/white differential in the average skills of young workers. Discrimination in various forms may further hinder the development of human capital in black children, and add to a gap in skills that is due to the race difference in socioeconomic background. The gap in some indicators of skill are very large. In our sample, the mean percentile score on the AFQT for the black sample is 23.78 while the mean for whites is 53.27. Neal and Johnson (1996) and a number of earlier papers have shown that in the NLSY sample of men a substantial part of the race gap in wages is associated with the race gap in AFQT.

If pre-market discrimination is an important factor in a gap between the average skills of black and white workers, then it seems likely that various forms of current labor market discrimination contribute to race differences in wages that are unrelated to skill. However, it is nevertheless interesting to examine the possibility that a correlation between race and skill might lead a rational, profit maximizing employer to use race as a cheap source of information about skills and statistically discriminate on the basis of race. Such statistical discrimination along racial lines can have very negative social consequences and is against the law. However, such discrimination would be difficult to detect.

A statistically discriminating firm might use race, along with education and other information to predict the productivity of new workers. With time, the productivity of the worker would become apparent and compensation would be based on the larger information available rather than the limited information available at the time of hire. Consequently, if statistical discrimination on the basis of race is important, then adding interactions between t and z variables such as AFQT and father's education to the wage equations should lead to a positive (or less negative) coefficient on black×t and should lead to an increase in the race intercept. As noted in section 2, if firms use race as information then it behaves as an s variable in the model and the logic is the same as in our analysis of the effect of education. On the other hand, if firms do not use or only partially use race as information, then a race indicator behaves as a z variable. As discussed in Section 2, in this case the race gap should widen with experience if race is negatively related to productivity, and adding a second z variable that is negatively related to race will reduce the race gap in experience slopes.[27]

The race differential in our basic specification in column 1 of Table 1 is -.1801. This drops to -.0969 when AFQT, F_ED, and education×t are added to the equation (column 5). When Black×t/10 is added in column 6, it enters with a coefficient of -.1456 (.0216). This coefficient is consistent with the hypothesis of no or very limited statistical discrimination on the basis of race and inconsistent with the hypothesis that firms make full use of race as information. The coefficient on Black is insignificantly different from 0, although the models do not provide a clear prediction about the sign of this variable, since race may be correlated with information in q that can legally be used. The fact that coefficient on Black×t/10 rises to -.0816 when F_ED×t and AFQT×t are added to the equation (column 8) is not informative about whether or not firms make full use of race as information. [28]

We obtain similar results using alternative experience measures in Tables 2 and 3. In Table 4, columns 7 and 8 we obtain similar results after controlling for initial 2-digit occupation. We obtain similar results using growth equations in Table 7, which should be robust to the presence of an economy wide time trend affecting the return to education, race, and AFQT. However, in the level equations we find that the results for race are sensitive to treatment of economy wide time trends. When we use potential experience as the measure of t the coefficient on Black×t declines only slightly (from -.0146 to -.0144) when we adding time trend interactions involving race and AFQT to the wage level equation corresponding to Table 1, column 7, but the race-experience interaction no longer drops when AFQT and experience is added. (Not reported.)

[27] The learning model in section 2 implies that differences across groups in the association between s and the z variable will lead to group differences in the b_{st} and b_{zt} coefficients. We have not explored this empirically. An obstacle to doing so is that the results might be sensitive to the linearity assumptions that we have made.

[28] Japanese and Chinese Americans score higher on aptitude and achievement tests than whites. Our analysis predicts that if firms statistically discriminate on the basis of race and ethnic background then the addition of AFQT and AFQT*t to an equation containing a dummy and experience interaction term for these groups will lead to an increase in the dummy variable and a reduction in the experience interaction. Sample sizes do not permit an analysis of these groups. While one could differentiate among whites based on ethnicity (see Borjas (1992), it is not clear that these ethnic differences are observable to employers. Our methods could be used to investigate statistical discrimination on the basis of attending prestigous colleges or particular college majors.

We wish to stress that the simple model of statistical discrimination cannot explain the negative coefficient on Black×t unless firms do not make full use of race as information. The accumulation of additional information during a career that can legally be used to differentiate among workers is fully consistent with our results. However, there are several other explanations of the race differences in the experience slope in the literature that may be at work here. It is also important to point out that the results for Black and Black×t alone (i.e., ignoring the behavior coefficients of the coefficients on education and education×t) are potentially consistent with a story in which firms are fully informed, AFQT is positively associated with on-the-job training, and the race difference in AFQT is partially responsible for a race differential in wage growth. Adding AFQT×t would reduce a negative bias in Black×t associated with differential training levels. The increase in Black×t when AFQT×t is added to the model would lead to a fall in the coefficient on Black. As we report below, we obtain qualitatively similar results when we add controls for employer training, but these controls reduce the magnitude of the coefficient on Black×t and the effect of adding AFQT×t on the coefficient on Black×t.

Another potential test of whether race is used to statistically discriminate or not is to see whether proposition 4 holds either when race is treated as an s variable or when it is treated as a z variable. To do this, we use the model in column 8 of table 1. With race treated as an s variable, we regress the z variables (AFQT, father's education, and the dummies for not knowing these quantities) on the two s variables. We sum the product of these coefficients and the coefficients on the z×t interactions in the main regression and compare them to the coefficients on the s×t interactions. We can then conduct a joint test of whether these two quantities are equal. For the education interaction the sum of the products equals -.00183 while the model coefficient is -.00301. For the race interaction, the two terms have opposite signs; the sum is .00644 while the model coefficient is -.00816. Not surprisingly, the proposition is soundly rejected.

When we treat race as a z variable, we begin our test by regressing the 5 z variables on education, our s variable. Here, we have only one restriction to test. The sum of the products equals -.00215 while the model coefficient equals -.00301. The proposition can be rejected at conventional levels of significance (the P-value is .027) but with corrected standard errors this will probably not be the case. This is a further indication that employers are not treating race as information, or at least not fully.

6. Models with Training

In Table 8 we report estimates of equation (24) along with models that exclude the training variables. In these models we have excluded father's education. In the basic model in column 1 the coefficient on $s \times t/10$ is -.0102. In column 2 we add T_t and $\Sigma T_{i\tau}$ to the equation. The variable T_t has the expected negative sign of -.1044 (.0179), while $\Sigma T_{i\tau}$ has a coefficient of .1864 (.0114). The coefficient on $s \times t/10$ falls to -.0346. The coefficient on AFQT falls from .0828 to .0582 while the coefficient on education rises slightly. The substantial negative experience slope on education might be consistent with a human capital story in which knowledge obtained in school depreciates over time unless one receives training. In column (3) AFQT×t/10 enters with a coefficient of .0502 (.0125), and the coefficient on and $s \times t/10$ drops from -.0358 to -.00427. These changes are consistent with employer learning/statistical discrimination. If we reverse the order in which the variables are added by adding AFQT×t before the training measures, the marginal effect of the training measures on education×t is much smaller. We have also estimated separate models for blacks and for whites and obtain a similar pattern.

In Columns 4-6 we investigate the effect of introducing the training measure on the race gap in wage slopes. The coefficient on black×t/10 declines from -.1467 to -.1048 when we add the training measures. Adding AFQT×t/10 leads to a further decline to -.0777.

To reduce the difficulties associated with the lack of data on training in the early years of the study and individual heterogeneity that is correlated with both training and wages, we turn to a first differenced version of (24). In the first difference version the current and lagged values of T_t enter. These results are in Table 9. The coefficient on education×t/10 declines from .0126 (.0094) to .0073 when the training measures are added. The coefficient on Black rises from -.0995 (.0351) to -.0923 (.0353). However, the coefficient on T_t is positive while the coefficient on T_{t-1} is negative. These signs are inconsistent with a simple human capital model but are consistent with an EL-SD model in which training opportunities are given to more productive workers and learning about productivity occurs over time. Adding the training variables to a model that contains AFQT and F_ED has little impact on the coefficients on these variables. (Compare columns 2 and 4.) Imprecision in the training measures may partially explain this fact, but does not provide an explanation for the sign pattern in the training coefficients. The coefficients on $s \times t$ and Black×t decline in absolute value when AFQT and F_ED are added, as is predicted by the EL-SD. Overall, the wage change results are quite consistent with an important role for EL-SD

We view the evidence as consistent with a role for both human capital and EL-SD, but cannot make a precise statement about the relative contribution of these factors because, as discussed above, training will be influenced by new information about employee performance and the quality of the training data is suspect.

7. Information Transmission Across Firms:

The formal model that we have used to interpret the results assumes that employers have the same information about workers. The results suggest that information about productivity does eventually get reflected in wages. However, they do not identify whether these adjustments occur primarily in the current firm, presumably in response to outside pressure from competitors who have information about the worker, or through moves to other employers with associated wage increases for workers who do not move.[29] In this section we briefly examine the issue of information transmission across firms.

A number of theoretical papers discuss whether information about productivity will be reflected in promotion paths and wage increases within firms, as well as the strategies firms might use to try to hide information about good workers (e.g. Greenwald (1986), Waldman (1984), Lazear (1986), Gibbons and Katz (1991)).[30] Unfortunately, the theory is ambiguous about whether a firm's private information concerning the worker will be reflected in wages offered by that firm to incumbent workers and about the mechanism that induces the firm to adjust wages. In some private information models in which only wages and perhaps position within the firm are observable to outside firms, the employer's information is not reflected in wages until the worker gets an outside offer. In Waldman (1984) it is reflected in wages after the firm reassigns the worker to a position in which output is more sensitive to ability. In Gibbons and Katz it is reflected in wages if the firm chooses not to lay off the worker. The firm lays off low productivity workers, who are hired by other firms at lower wages. Outside firms infer that the remaining workers are of higher quality,

---

[29] Although we do not know of systematic evidence on this, casual empiricism suggests that changes in the legal system have led some firms to adopt the explicit policy of not providing references for former employees. Also, increased firing costs and concern about litigation may have made employers more reluctant to discharge workers for poor performance. Statistical discrimination may become a more serious problem if information flows are restricted. This may lead firms to relate compensation to performance more explicitly, with more turnover being a "voluntary" response to below average wage increases. On the other hand, difference in wages across groups may be attenuated because firms may be reluctant to open up large wage differentials between persons with similar education, seniority, and experience. It is possible that the balance between these two considerations has changed over time.

[30] None of this literature considers the implications of the possibility that employers and co-workers acquire reputations for how positive they are in promoting the careers of individuals or that the incentives of co-workers and even supervisors to keep favorable information about a colleague private or in concealing unfavorable information from associates outside the firm may be quite different from those of the employer. These factors would undermine the case that firms would want to and be able to keep inside information inside the firm.

which forces the employer to raise wages of those who stay with the firm. Both models have the implication that hard to observe variables like AFQT, F_ED, and the wage of an older sibling should be positively related to wage growth if one does not condition on whether a person was laid off or not. This is what we found above.

Gibbons and Katz (1991) provide empirical support for the hypothesis that layoffs should be negatively related to wage growth. But there are a number of other reasons why layoffs should be negatively related to wage growth (labor market conditions, lost seniority, for example). To obtain more focused tests, we interact personal characteristics that are hard for employers to observe directly with indicators for layoffs and discharges. The coefficients on these variables should differ from the coefficients on characteristics that affect productivity and are easy for employers to observe, such as years of schooling if (1) layoffs occur for multiple reasons, some of which have nothing to do with the worker, (2) the probability that a layoff reflects low worker specific productivity relative to the wage is related to z variables, and (3) outside employers have information about the nature of the layoff or obtain information (through references, for example) about productivity.

This suggests an equation of the form

$$w_t - w_{t-1} = \beta_0 + \text{Layoff}_t\beta_1 + z\beta_2 + z[\text{Layoff}_t]\,\beta_3 + z[\text{Layoff}_t]t\,\beta_4 + \text{other controls.}$$

If knowledge acquired by firms is reflected in wages, then β_2 should be nonzero, and β_3 and β_4 should be near zero. If knowledge acquired by firms is not reflected in wages, then β_2 should be small and β_3 and β_4 should be nonzero. Given sample size limitations we have estimated a simplified version of the above equation on the sample of layoffs only, with $z[\text{Layoff}_t\,]t$ excluded:

$$w_t - w_{t-1} = (\beta_0 + \beta_1) + z\,\beta_3 + \text{other controls.}$$

Our evidence on whether hard to observe variables such are positively associated with layoff losses is weak at best. In fact, we find that losses are larger for persons with high AFQT. We have not controlled for labor market conditions, and among the sample of layoffs they may be correlated with AFQT.[31]

[31] We investigated whether the finding that wage losses rise with AFQT is driven by a positive correlation between AFQT and employment in a white collar, non union job, where layoffs are least likely to be influenced by seniority rules. Gibbons and Katz note that layoffs are likely to be a particularly negative signal for white collar workers and restrict their analysis to them. However, splitting the leads to an even more negative coefficient for white collar workers than for blue collar workers.

In Table 10 we report estimates of the effect of AFQT and F_ED on employer initiated separations. These include layoffs, firings, and plant closings. Our results were not very sensitive to distinguishing among these three types of job loss. We find that AFQT has a weak negative effect on the probability of losing one's job, even after conditioning on seniority in the firm. However, when seniority is controlled for a swing of two standard deviations in AFQT changes this probability by .02, which is only 1/5 of the mean layoff rate of .1. We obtain similar results when the seniority control is dropped.

Our results suggest that only a small part of the rise with t in the effect of AFQT on wages operates through an association between AFQT and layoffs and the wage losses experienced by those who are laid off.

Lazear (1986) presents a model in which both the current firm and outside firms observe indicators of the productivity of the worker. His model predicts that workers with favorable productivity traits that are hard to observe directly will be more likely to receive outside offers and more likely to quit than workers whose hard to observe characteristics make them less productive. In results not reported we find that F_ED is positively related to the quit rate conditional on education and experience and tenure. AFQT does not have a significant effect. Neither AFQT nor F_ED is significantly related to wage growth among those who quit. (Not reported).

These results tentatively suggest that information flows in the labor market are sufficient to force a firm to differentiate among workers as the firm obtains better information about their productivity. A careful investigation will require a separate paper..

8. The Potential for Testing Services to Certify Skill

Our estimates provide information about the rate at which employers learn about worker quality. In Altonji and Pierret (1996) we use our empirical estimates to explore the implications of the rate at which employers learn about worker quality for the empirical relevance of the educational screening hypothesis. We show that even if employers learn relatively slowly about the productivity of new workers, the portion of the return to education that could reflect signaling of ability is quite limited. While education may be too expensive to serve as a means for able workers to certify themselves to employers, perhaps other mechanisms could perform this function, at least for some determinants of productivity. Here we point out that interpreting our estimates of the time profile of the effect of AFQT on wages as the result of employer learning implies that high ability workers would have a substantial financial incentive to take the AFQT to differentiate themselves from those who are less able in this dimension.

Suppose that a third party were to administer the AFQT and certify the results to outside employers, in much the same way that the Educational Testing Service administers the SAT exams. Using our estimates of the learning profile and assuming that firms know all of the information contained in AFQT by the time experience is 15, we have computed how much a person who believes that he is 1 standard deviation above the mean for the AFQT would pay to take the test at the time he enters the workforce.[32] The OLS estimates using potential experience underlying Table 6, panel A, column 3) imply that if firms become fully informed about productivity by the time experience is 15 and the interest rate is .1, then the person would be willing to pay .559 of the first year's salary for the test.[33] The corresponding value when we use potential experience as an instrument for actual experience (panel B, column 3) is .330.

These calculations raise the issue of why such a testing service has not emerged if information is initially imperfect. One answer is that firms are not aware that the AFQT captures characteristics that have a strong association with productivity. It is only recently, with the availability of the NLSY, that labor economists have become aware of this. Another is that it would be difficult for a testing firm to become established at a national level. A third is that, given race differences in the distribution of AFQT scores, firms who make use of AFQT information in hiring for a specific job would have the burden of establishing that they are relevant to productivity in that job or run the risk of violating discrimination laws. This would be true even if individuals provided firms with the test results. However, we do not find these answers to be fully satisfactory.[34] Analyses based on variables such as the wage rates of siblings or father's education may be less vulnerable to this objection. In any event, we should also point out that our estimates of the AFQT-experience profile are sensitive to treatment of time trends and training, so that financial return to being certified as high AFQT is probably substantially less than the above numbers imply.

[32] If a worker did not know his ability, he could take a practice test on his own. Presumably, this would not raise the total cost of the test very much.

[33] Here we are assuming that only 1 worker takes the test and ignoring the fact that the composition of the pool of workers who choose to take the test in equilibrium would influence return for a particular type of worker.

[34] Note also that in the absence of an institution such as the Educational Testing Service, a firm might provide the test. Some firms perform their own testing.. However, if the results were available to the employees or other firms know that a particular firm tests its employees, then the firm would not be able to capture the full return to testing.

9. Conclusion

This paper explores the implication of the premise that firms use the information they have available to them to form judgments or about the productivity of workers and then revise these beliefs as additional information becomes available. This a premise that seems natural to us and receives some strong empirical support in Farber and Gibbons (1996). If profit maximizing firms have limited information about the general productivity of new workers, then they may use easily observable characteristics such as years of education or race to statistically discriminate among workers. We show that as firms acquire more information about a worker, pay may become more dependent on productivity and less dependent on easily observable characteristics or credentials. This basic idea is quite general and provides a way to test for statistical discrimination in the labor market and elsewhere in situations in which agents learn, such as credit markets.

We investigate it empirically by estimating a wage equation that contains interactions between experience and hard to observe characteristics such as AFQT and father's education along with the interaction between experience and a variable that firms can easily observe, such as years of education. We assume that all three variables are related to productivity. We find the wage effect of the unobservable productivity variables rise with time in the labor market and the wage effect of education falls. These results match the predictions of our model of statistical discrimination with learning.

We use a similar methodology to investigate whether employers statistically discriminate on the basis of race. If our model is taken literally, the small race differentials for new workers and the spread in the race gap with experience is most consistent with the view that race is negatively correlated with productivity and the productivity gap becomes reflected in wages as firms acquire additional information that can legally be used to differentiate among workers. We wish to stress however, that other factors are probably as or more important in differences between whites and blacks in wage profiles, and race differences in human capital accumulation accounts for at least part of our findings. Also, our empirical results for race are sensitive to treatment of economy wide changes in the effects of race, AFQT, and education. Future research should also address the large race gap and education gap in employment rates, particularly for young workers. In situations in which there are alternatives to the conventional labor market and employees in the alternative sector do not acquire work histories that have value or are informative to firms in the conventional sector, then statistical discrimination of the type described above may reduce participation rates of the disadvantaged group in the conventional labor market.

It is worth emphasizing that the analysis in the paper suggests alternative interpretations of empirical models of wages and other outcomes that involve experience interactions. It will be useful to re-examine the results of other studies that included interactions between experience and easy to observe variables such as schooling, race, gender, and experience in equations that also contain interactions between experience and harder to observe background measures. We have not been successful in sorting out the relative importance of differences among workers in training on one hand and statistical discrimination with learning on the other for our results. This is an important area for future research.

An important and reasonably straightforward extension of the analysis is to other easily observable and hard to observable background characteristics. For example, do firms statistically discriminate on the basis of the neighborhood one is from or on the basis of the reputation of the high school, college, or graduate school one attends? A study of whether new immigrants are judged by the average skills of their countrymen in the U.S. would be a natural step in the research by Borjas (1992) and others documenting differences among immigrants in labor market success. These issues are researchable using the approach developed in this paper. Finally, it would be useful to apply the methods of the paper to other labor market outcomes in addition to wages.

References

D. Aigner and G. Cain (1977), "Statistical Theories of Discrimination" Industrial and Labor Relations Review.

Albrecht, J. (1981), "A Procedure for Testing the Signalling Hypothesis." Journal of Public Economics, pp. 123-32.

Altonji, J. G. and C. R. Pierret (1997), "Employer Learning and the Signaling Value of Education," in Ohashi, Isao and Tachibanaki, Toshiaki, (ed.) Internal Labour Markets, Incentives and Employment, Macmillan Press Ltd.

Altonji, J.G. and J. Spletzer (1991), "Worker Characteristics, Job Characteristics, and the Receipt of On-the-Job Training." Industrial and Labor Relations Review, 45(1) pp. 58-79.

Bartel, A. P., and N. Sicherman (1993), "Technological Change and On-The-Job Training of Young Workers" unpublished paper, Columbia University.

Becker, G. S. (1971), The Economics of Discrimination. 2nd ed., Chicago: University of Chicago Press.

Borjas, G. (1992), "Ethnic Capital and Intergenerational Mobility" Quarterly Journal of Economics ,117 (1) pp.123-150.

Carmichael, H. L. (1989). "Self-Enforcing Contracts, Shirking, and Life Cycle Incentives" The Journal of Economic Perspectives 3 (Fall): 65-84.

Chay, Kenneth Y. and David S. Lee, (1997) "Changes in Relative Wages in the 1980s: Returns to Observed and Unobserved Skills and Black-White Wage Differentials", unpublished paper. February.

Coate, S. and G. Loury (1993) "Will Affirmative Action Policies Eliminate Negative Stereotypes?" C, 83 (5), pp. 1220-1240

Devine, T.J., and Kiefer, N.M. (1991). Empirical Labor Economics (New York: Oxford University Press).

Farber, H., and R. Gibbons, (1996) "Learning and Wage Dynamics" Quarterly Journal of Economics, pp.1007-47.

Frazis, H., (1993), "Selection Bias and the Degree Effect," Journal of Human Resources 28 (3): (Summer 1993): 538-554.

Gibbons, R., and L. Katz (1991) "Layoffs and Lemons" Journal of Labor Economics 9: 351-80.

Greenwald, B., (1986) "Adverse Selection in the Labor Market" Review of Economic Studies 53: 325-47

Foster, A. D. and M. R. Rosenzweig (1993), "Information, Learning, and Wage Rates in Low Income Rural Areas," Journal of Human Resources.

Holzer, H., (1988) "Search Methods Use by Unemployed Youth" Journal of Labor Economics 6: 1-20.

Jovanovic, B., (1979) "Job Matching and the Theory of Turnover" Journal of Political Economy 87: 972-90.

Katz, Lawrence F. and Kevin M. Murphy. "Changes in Relative Wages, 1963-1987: Supply and Demand Factors." Quarterly Journal of Economics. Vol 107:1, pp35-78.

Lang K., (1986) "A Language Theory of Discrimination", Quarterly Journal of Economics, 101 (May): 363-82

Lazear, E., (1986) "Raids and Offer Matching," Research in Labor Economics 8 (part A): 141-165

Lundberg, S., and R. Startz, "Private Discrimination and Social Intervention in Competitive Labor Markets," American Economic Review, 73 (June): 340-347.

L. Lynch, (1992) "Private Sector Training and the Earnings of Young Workers," American Economic Review 82 (March):299-312

Medoff, J. and K. Abraham. (1980), "Experience, Performance and Earnings." Quarterly Journal of Economics, December, pp. 703-736.

Montgomery, J. (1991). "Social Networks and Labor Market Outcomes: Toward and Economic Analysis", American Economic Review 81: 1408-18.

Murnane, Richard J., John B. Willett and Frank Levy, (1995) "The Growing Importance of Cognitive Skills in Wage Determination ", Review of Economics and Statistics, vol. lxxvii, no. 2, (May): 251-266.

Murphy, K. and F. Welch, (1992) "The Structure of Wages", Quarterly Journal of Economics, 107 (February): 285-326.

Neal, Derek A. and William R. Johnson 1996 "The Role of Premarket Factors in black-White Wage Differences." Journal of Political Economy 104 (5), 869-895.

Oettinger, G. S. (1996), "Statistical Discrimination and the Early Career Evolution of the Black-White Wage Gap". Journal of Labor Economics .

Parsons, D. O., (1986) "The Employment Relationship: Job Attachment, Worker Effort, and the Nature of Contracts" Ashenfelter and Layard, eds. Handbook of Labor Economics

Parsons, D. O., "Reputational Bonding of Job Performance: The Wage Consequences of Being Fired" unpublished paper. (July 1993)

Riley, J. G. (1979) "Testing the Educational Screening Hypothesis," Journal of Political Economy, October, pp. S227-52.

Taber, C. R., (1996) "The Rising College Premium in the Eighties: Return to College or Return to Ability," unpublished paper, Northwestern University (April).

Spletzer, J. and M. Lowenstein (1996), "Belated Training: The Relationship Between Training, Tenure and Wages" unpublished paper, Bureau of Labor Statistics.

Waldman, M. (1984), "Job Assignment, Signaling, and Efficiency," Rand Journal of Economics. 15: 255-67.

Weiss, A. (1995), "Human Capital and Sorting Models," Journal of Economic Perspectives.

White, H. (1984). Asymptotic Theory for Econometricians , Orlando, FL: Academic Press

Table 1: The Effects of Standardized AFQT, Father's Education, and Schooling on Wages
Dependent Variable: Log Wage. Experience Measure: Potential Experience.

OLS estimates (standard errors)

	Model:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
(a)	Education	0.0946 (0.0034)	0.0742 (0.0039)	0.0729 (0.0040)	0.1001 (0.0051)	0.0798 (0.0054)	0.0984 (0.0057)	0.0818 (0.0054)	0.0949 (0.0057)	0.0788 (0.0064)	0.0855 (0.0064)
(b)	Black	-0.1801 (0.0117)	-0.1039 (0.0138)	-0.0974 (0.0141)	-0.1799 (0.0117)	-0.0969 (0.0141)	-0.0956 (0.0142)	0.0153 (0.0203)	-0.0330 (0.0226)	-0.0948 (0.0142)	-0.0945 (0.0141)
(c)	Standardized AFQT		0.0807 (0.0077)	0.0783 (0.0078)		0.0785 (0.0078)	0.0179 (0.0107)	0.0790 (0.0077)	0.0328 (0.0115)	0.0187 (0.0107)	-0.0028 (0.0111)
(d)	Father's Education/10			0.0263 (0.0192)		0.0259 (0.0192)	-0.0015 (0.0031)	0.0028 (0.0019)	-0.0062 (0.0308)	-0.0163 (0.0308)	-0.0286 (0.0318)
(e)	Education * Experience/10				-0.0077 (0.0062)	-0.0098 (0.0061)	-0.0351 (0.0069)	-0.0122 (0.0061)	-0.0301 (0.0071)	-0.0510 (0.0087)	-0.0361 (0.0105)
(f)	AFQT * Experience/10						0.0820 (0.0125)		0.0622 (0.0143)	0.0817 (0.0125)	0.0316 (0.0241)
(g)	Father's Ed * Experience/100						0.0592 (0.0372)		0.0481 (0.0372)	0.0611 (0.0371)	0.0392 (0.0667)
(h)	Black * Experience/10							-0.1456 (0.0216)	-0.0816 (0.0262)		

Note: All equations control for a quadratic time trend, urban residence, and dummy variables to control for whether Father's education is missing and whether AFQT is missing, and interactions between these dummy variables and experience when Experience interactions are included. Column 9 includes the interaction between education and time/10 (the estimate is .0349 (.0078)). Column 10 includes interactions of education (.0142(.0101)), AFQT (.0688(.0228)), and Father's Education/10 (.0317(.0631)) with time/10. Standard errors are White/Huber standard errors computed accounting for the fact that there are multiple observations for each worker. The sample size is 27704 observations from 4042 individuals.

Table 2: The Effects of Standardized AFQT, Father's Education, and Schooling on Wages
Dependent Variable: Log Wage. Experience Measure: Actual Experience.

OLS estimates (standard errors)

	Model:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
(a)	Education	0.0805 (0.0026)	0.0613 (0.0031)	0.0599 (0.0032)	0.0713 (0.0033)	0.0504 (0.0038)	0.0628 (0.0040)	0.0524 (0.0038)	0.0614 (0.0040)	0.0385 (0.0059)	0.0485 (0.0064)
(b)	Black	-0.1378 (0.0113)	-0.0673 (0.0132)	-0.0624 (0.0134)	-0.1381 (0.0113)	-0.0625 (0.0134)	-0.0622 (0.0135)	-0.0025 (0.0152)	-0.0346 (0.0159)	-0.0608 (0.0135)	-0.0602 (0.0135)
(c)	Standardized AFQT		0.0754 (0.0073)	0.0730 (0.0075)		0.0731 (0.0075)	0.0366 (0.0082)	0.0726 (0.0075)	0.0430 (0.0084)	0.0373 (0.0082)	0.0041 (0.0120)
(d)	Father's Education/10			0.0324 (0.0186)		0.0321 (0.0187)	0.0005 (0.0022)	0.0032 (0.0019)	0.0085 (0.0224)	0.0042 (0.0223)	-0.0127 (0.0345)
(e)	Education * Experience/10				0.0195 (0.0055)	0.0200 (0.0054)	-0.0055 (0.0066)	0.0163 (0.0054)	-0.0025 (0.0068)	-0.0320 (0.0099)	-0.0165 (0.0114)
(f)	AFQT * Experience/10						0.0750 (0.0131)		0.0614 (0.0148)	0.0737 (0.0131)	0.0226 (0.0240)
(g)	Father's Ed * Experience/100						0.0587 (0.0367)		0.0502 (0.0370)	0.0613 (0.0365)	0.0362 (0.0678)
(h)	Black * Experience/10							-0.1267 (0.0233)	-0.0583 (0.0280)		

Note: All equations control for a quadratic time trend, urban residence, and dummy variables to control for whether Father's education is missing and whether AFQT is missing, and interactions between these dummy variables and experience when Experience interactions are included. Column 9 includes the interaction between education and time/10 (the estimate is . 0402 (.0085)). Column 10 includes interactions of education (.0195(.0104)), AFQT (.0684(.0211)), and Father's Education/10 (.0333(.0623)) with time/10. Standard errors are White/Huber standard errors computed accounting for the fact that there are multiple observations for each worker. The sample size is 27704 observations from 4042 individuals.

Table 3: IV Estimates of the Effects of Standardized AFQT, Father's Education, and Schooling on Wages
Dependent Variable: Log Wage. Experience Measure: Actual Experience with Potential Experience as Instruments .

IV estimates (standard errors)

	Model:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
(a)	Education	0.0813 (0.0028)	0.0620 (0.0034)	0.0606 (0.0035)	0.0891 (0.0050)	0.0692 (0.0054)	0.0879 (0.0056)	0.0726 (0.0054)	0.0843 (0.0057)	0.0468 (0.0065)	0.0517 (0.0069)
(b)	Black	-0.1368 (0.0116)	-0.0650 (0.0132)	-0.0601 (0.0135)	-0.1367 (0.0116)	-0.0600 (0.0136)	-0.0593 (0.0136)	0.0495 (0.0186)	0.0054 (0.0205)	-0.0531 (0.0138)	-0.0527 (0.0138)
(c)	Standardized AFQT		0.0762 (0.0074)	0.0737 (0.0075)		0.0738 (0.0076)	0.0177 (0.0096)	0.0728 (0.0075)	0.0332 (0.0102)	0.0218 (0.0096)	0.0005 (0.0127)
(d)	Father's Education/10			0.0337 (0.0188)		0.0340 (0.0188)	0.0000 (0.0028)	0.0033 (0.0019)	0.0091 (0.0282)	-0.0043 (0.0278)	0.0111 (0.0363)
(e)	Education * Experience/10				-0.0165 (0.0088)	-0.0181 (0.0087)	-0.0561 (0.0100)	-0.0242 (0.0087)	-0.0483 (0.0101)	-0.1220 (0.0188)	-0.1090 (0.0221)
(f)	AFQT * Experience/10						0.1148 (0.0164)		0.0819 (0.0188)	0.1056 (0.0163)	0.0539 (0.0399)
(g)	Father's Ed * Experience/100						0.0744 (0.0480)		0.0531 (0.0484)	0.0877 (0.0478)	0.1219 (0.1124)
(h)	Black * Experience/10							-0.2305 (0.0318)	-0.1364 (0.0387)		

Note: All equations control for a quadratic time trend, urban residence, and dummy variables to control for whether Father's education is missing and whether AFQT is missing, and interactions between these dummy variables and experience when Experience interactions are included. The instrumental variables are the corresponding terms involving potential experience and the other variables in the model. Column 9 includes the interaction between education and time/10 (the estimate is 0.0803 (.0135)). Column 10 includes interactions of education (.0670(.0166)), AFQT (.0546(.0311)), and Father's Education/10 (-.0376(.0882)) with time/10. Standard errors are White/Huber standard errors computed accounting for the fact that there are multiple observations for each worker. The sample size is 27704 observations from 4042 individuals.

Table 4: Estimates of the Effects of Standardized AFQT, Father's Education, and Schooling on Wages
Controlling for 2-digit Occupation Codes of Initial Job
Dependent Variable: Log Wage. Experience Measure: Potential Experience.
OLS estimates (standard errors)

	Model:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
(a)	Education	0.0759 (0.0042)	0.0611 (0.0045)	0.0596 (0.0046)	0.0717 (0.0058)	0.0572 (0.0061)	0.0767 (0.0064)	0.0592 (0.0061)	0.0745 (0.0064)	0.0666 (0.0067)	0.0709 (0.0068)
(b)	Black	-0.1539 (0.0131)	-0.0917 (0.0154)	-0.0829 (0.0156)	-0.1539 (0.0131)	-0.0830 (0.0156)	-0.0812 (0.0156)	0.0190 (0.0209)	-0.0413 (0.0222)	-0.0809 (0.0157)	-0.0809 (0.0156)
(c)	Standardized AFQT		0.0662 (0.0085)	0.0635 (0.0086)		0.0634 (0.0086)	-0.0036 (0.0111)	0.0638 (0.0086)	0.0061 (0.0117)	-0.0030 (0.0111)	-0.0151 (0.0119)
(d)	Father's Education/10			0.0298 (0.0206)		0.0299 (0.0207)	-0.0049 (0.0315)	0.0310 (0.0206)	0.0010 (0.0316)	-0.0051 (0.0316)	-0.0188 (0.0342)
(e)	Education * Experience/10				0.0057 (0.0066)	0.0032 (0.0065)	-0.0245 (0.0075)	0.0008 (0.0065)	-0.0212 (0.0078)	-0.0354 (0.0096)	-0.0254 (0.0119)
(f)	AFQT * Experience/10						0.0940 (0.0140)		0.0807 (0.0159)	0.0940 (0.0140)	0.0626 (0.0280)
(g)	Father's Ed * Experience/100						0.0532 (0.0411)		0.0447 (0.0414)	0.0530 (0.0411)	0.0229 (0.0752)
(h)	Black * Experience/10							-0.1377 (0.0233)	-0.0542 (0.0281)		

Note: All equations control for a quadratic time trend, urban residence, and dummy variables to control for whether Father's education is missing and whether AFQT is missing, and interactions between these dummy variables and experience when Experience interactions are included. Column 9 includes the interaction between education and time/10 (the estimate is .0206 (.0083)). Column 10 includes interactions of education (.0072(.0111)), AFQT (.0409(.0263)), and Father's Education/10 (.0405(.0697)) with time/10. Standard errors are White/Huber standard errors computed accounting for the fact that there are multiple observations for each worker. The Sample Size is 22271 observations from 3187 individuals.

Table 5: OLS Estimates of the Effects of Sibling Wage and Schooling on Wages
Dependent Variable: Log Wage. Experience Measure: Potential Experience

OLS estimates (standard errors)

	Model:	(1)	(2)	(3)	(4)	(5)	(6)	(7)
(a)	Education	0.0936 (0.0055)	0.0830 (0.0055)	0.1032 (0.0077)	0.0900 (0.0077)	0.0938 (0.0078)	0.0803 (0.0089)	0.0805 (0.0089)
(b)	Black	-0.1932 (0.0164)	-0.1620 (0.0163)	-0.1932 (0.0164)	-0.1621 (0.0163)	-0.1620 (0.0163)	-0.1619 (0.0163)	-0.1619 (0.0163)
(c)	Log Wage of Oldest Non-Missing Sibling		0.1876 (0.0191)		0.1873 (0.0191)	0.1266 (0.0276)	0.1264 (0.0276)	0.1230 (0.0323)
(d)	Sibling is Female		0.0205 (0.0155)		0.0208 (0.0155)	0.0211 (0.0155)	0.0214 (0.0155)	0.0213 (0.0155)
(e)	Education * Experience/10			-0.0133 (0.0091)	-0.0097 (0.0089)	-0.0146 (0.0090)	-0.0220 (0.0113)	-0.0216 (0.0120)
(f)	Log of Sibling Wage * Experience/10					0.0860 (0.0327)	0.0862 (0.0326)	0.0802 (0.0679)

Note: All equations control for a quadratic time trend, and urban residence. Column 6 includes the interaction between education and time/10 (the estimate is (.0204(.0113)). Column 7 includes interactions of education (.0199(.0122)), and Log Sibling Wage (.0085(.0667)) with time/10. Standard errors are White/Huber standard errors computed accounting for the fact that there are multiple observations for each worker. The Sample Size is 13,555 observations from 1881 individuals.

Table 6: The Effects of Standardized AFQT, and Schooling on Wages Over Time.

Derivatives at Selected Experience Levels

Dependent Variable: Log Wage

A) Potential Experience.

Years of Experience	$\frac{\partial w_t}{\partial AFQT}$	$\frac{\partial w_t^2}{\partial AFQT, \partial t}$	$\frac{\partial w_t}{\partial s}$	$\frac{\partial w_t^2}{\partial s, \partial t}$
0	0.0197 (0.0235)	0.0025 (0.0139)	0.0786 (0.0092)	0.0053 (0.0040)
1	0.0235 (0.0275)	0.0049 (0.0144)	0.0830 (0.0101)	0.0034 (0.0040)
3	0.0370 (0.0347)	0.0084 (0.0155)	0.0865 (0.0116)	0.0002 (0.0042)
5	0.0560 (0.0415)	0.0104 (0.0166)	0.0843 (0.0131)	-0.0023 (0.0043)
8	0.0881 (0.0512)	0.0104 (0.0181)	0.0731 (0.0152)	-0.0048 (0.0046)
12	0.1206 (0.0640)	0.0048 (0.0201)	0.0513 (0.0179)	-0.0056 (0.0048)

B) Actual Experience Instrumented with Potential Experience.

Years of Experience	$\frac{\partial w_t}{\partial AFQT}$	$\frac{\partial w_t^2}{\partial AFQT, \partial t}$	$\frac{\partial w_t}{\partial s}$	$\frac{\partial w_t^2}{\partial s, \partial t}$
0	0.0183 (0.0205)	0.0092 (0.0231)	0.0881 (0.0105)	-0.0024 (0.0086)
1	0.0278 (0.0316)	0.0099 (0.0251)	0.0843 (0.0137)	-0.0051 (0.0089)
3	0.0496 (0.0496)	0.0120 (0.0288)	0.0711 (0.0190)	-0.0074 (0.0096)
5	0.0755 (0.0658)	0.0138 (0.0323)	0.0566 (0.0236)	-0.0068 (0.0102)
8	0.1172 (0.0893)	0.0131 (0.0373)	0.0406 (0.0300)	-0.0037 (0.0112)
12	0.1475 (0.1206)	-0.0012 (0.0436)	0.0299 (0.0381)	-0.0032 (0.0123)

The equations the same variables as the equation in column (6) of table 1 except the interactions between education and experience and between AFQT and experience involve fourth-order polynomials in experience. In panel B, the instrumental variables are the corresponding terms involving potential experience and the other variables in the model.

Table7: Estimates of the Effects of AFQT, Father's Education, and Schooling on Wage Growth

Dependent Variable: Δ log Wage.
Coefficient Estimates (standard errors)

	OLS, potential experience				IV, actual experience treated as endogenous			
Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Education * ΔExperience	0.0148 (0.0094)	-0.0092 (0.0110)	-0.0080 (0.0113)	0.0126 (0.0094)	0.0295 (0.0079)	-0.0030 (0.0100)	0.0021 (0.0101)	0.0256 (0.0080)
AFQT* ΔExperience		0.0646 (0.0192)	0.0595 (0.0210)			0.0905 (0.0197)	0.0700 (0.0213)	
Father's Education * ΔExperience		0.0809 (0.0557)	0.0776 (0.0563)			0.0952 (0.0561)	0.0818 (0.0567)	
Black * ΔExperience			-0.0213 (0.0409)	-0.0995 (0.0351)			-0.0850 (0.0420)	-0.1768 (0.0372)
S.E.E	.29655	.29650	.29650	.29653	.29600	.29589	.29588	.29590

Note: All equations control for a the change in a quadratic time trend, change in urban residence, and dummy variables to control for whether father's education is missing and whether AFQT is missing, and interactions between these dummy variables and the change in experience when change in experience interactions are included. The instrumental variables are the corresponding terms involving potential experience and the other variables in the model. Standard errors are White/Huber standard errors computed accounting for the fact that there are multiple observations for each worker. The sample size is 19393 observations from 3580 individuals.

Table 8: The Effects of Standardized AFQT, Schooling, and Training on Wages
Dependent Variable: Log Wage ; Experience Measure: Potential Experience

Training Measure: Predicted before 88, Actual After

OLS estimates (standard errors)

	Model:	(1)	(2)	(3)	(4)	(5)	(6)
(a)	Education	0.0808 (0.0054)	0.0856 (0.0055)	0.0951 (0.0057)	0.0830 (0.0054)	0.0869 (0.0055)	0.0921 (0.0058)
(b)	Black	-0.1008 (0.0142)	-0.0920 (0.0143)	-0.0916 (0.0143)	0.0117 (0.0206)	-0.0131 (0.0203)	-0.0332 (0.0221)
(c)	Standardized AFQT	0.0822 (0.0078)	0.0572 (0.0079)	0.0218 (0.0104)	0.0828 (0.0078)	0.0582 (0.0078)	0.0376 (0.0114)
(e)	Education * Experience/10	-0.0102 (0.0062)	-0.0346 (0.0066)	-0.0472 (0.0073)	-0.0129 (0.0062)	-0.0358 (0.0066)	-0.0427 (0.0075)
(f)	AFQT * Experience/10			0.0502 (0.0125)			0.0288 (0.0149)
(g)	Black * Experience/10				-0.1467 (0.0221)	-0.1048 (0.0222)	-0.0777 (0.0266)
(h)	Training: T_t		-0.1044 (0.0179)	-0.0936 (0.0180)		-0.0974 (0.0179)	-0.0930 (0.0180)
(i)	Cumulative Training: ΣT_τ		0.1864 (0.0114)	0.1781 (0.0116)		0.1810 (0.0114)	0.1776 (0.0116)

Note: All equations control for a quadratic time trend, urban residence, a cubic in potential experience. In this table , T_t and T_τ are the predicted probability of training in year t if before 1987 and actual training if year t is after 1987. Predictions are based on a probit model containing: years of schooling, potential experience, Black, AFQTPCT, schooling time potential experience and potential experience squared, AFQT times potential experience and potential experience squared, and the product of AFQTPCT, schooling, and potential experience. Standard errors are White/Huber standard errors computed accounting for the fact that there are multiple observations for each worker. The sample size is 25115 observations from 3768 individuals.

Table 9: Estimates of the Effects of AFQT, Father's Education, and Schooling on Wage Growth with Controls for Training
Dependent Variable: Δ log Wage. Experience Measure: Potential Experience

Coefficient Estimates (standard errors)

Variable	(1)	(2)	(3)	(4)
Education * ΔExperience/10	0.0126 (0.0094)	-0.0080 (0.0113)	0.0073 (0.0096)	-0.0108 (0.0113)
AFQT* ΔExperience/10		0.0595 (0.0210)		0.0533 (0.0211)
Father's Education * ΔExperience/10		0.0078 (0.0056)		0.0075 (0.0056)
Black * ΔExperience/10	-0.0995 (0.0351)	-0.0213 (0.0409)	-0.0923 (0.0353)	-0.0215 (0.0408)
Lagged Training - lagged T / 10			-0.0109 (0.0950)	-0.0336 (0.0951)
Training: T / 10			0.2622 (0.0891)	0.2446 (0.0894)
S.E.E	.29653	.29650	.29649	.29647

Note: All equations control for a the change in a quadratic time trend, change in urban residence, and dummy variables to control for whether father's education is missing and whether AFQT is missing, and interactions between these dummy variables and the change in experience when change in experience interactions are included. Standard errors are White/Huber standard errors computed accounting for the fact that there are multiple observations for each worker. The sample size is 19393 observations from 3580 individuals.

Table 10: The Effects of Potential Experience, Standardized AFQT, Father's Education, and Schooling on the Probability of Employer-Initiated Separation

Linear Probability Models
Dependent Variable: Employer-Initiated Separation.

OLS estimates (standard errors)

	Model:	(1)	(2)
(a)	Potential Experience / 10	-0.0302 (0.0194)	-0.1646 (0.0540)
(b)	Potential Experience Squared / 100	-0.0143 (0.0113)	0.0148 (0.0129)
(c)	Tenure	-0.0141 (0.0006)	-0.0146 (0.0006)
(d)	Education	-0.0153 (0.0012)	-0.0206 (0.0027)
(e)	Black	0.0272 (0.0057)	0.0265 (0.0057)
(f)	Standardized AFQT	-0.0108 (0.0029)	-0.0251 (0.0061)
(g)	Father's Education / 100	0.0303 (0.0701)	0.0991 (0.1532)
(h)	Education * Experience / 10		0.0083 (0.0033)
(i)	AFQT * Experience / 10		0.0188 (0.0065)
(j)	Father's Ed * Experience /1000		-0.0910 (0.1738)

Note: An Employer-Initiated Separation includes separations because of layoffs, firings, and plant closings. All equations control for urban residence, and dummy variables to control for whether Father's education is missing and whether AFQT is missing, and interactions between these dummy variables and experience when Experience interactions are included. Standard errors are White/Huber standard errors computed accounting for the fact that there are multiple observations for each worker. The sample size is 27443 observations from 4034 individuals.

Table A1: Descriptive Statistics

Variable	Mean	Standard Deviation	Minimum	Maximum
Real Hourly Wage	8.370	4.766	2.01	96.46
Log of Real Hourly Wage (w)	2.005	0.474	0.7	4.57
Potential Experience (t)	7.349	3.665	0	21
Actual Experience (t)	4.925	3.424	0	18.26
Education (s)	12.699	2.136	8	18
Black dummy (Black)	0.290	0.454	0	1
Dummy for not knowing AFQT Score	0.038	0.191	0	1
Standardized AFQT Score (AFQT)	-0.133	1.022	-2.780	1.922
Dummy for not knowing Father's Education	0.119	0.324	0	1
Father's Education (F_ED)	11.709	3.112	4	20
Dummy for Urban Dweller	0.781	0.413	0	1
Year	86.623	81.558	79	92
Training (T_t)	0.096	0.200	0	1
Cumulative Training: (ΣT_τ)	0.462	0.549	0	5.592

Sample size = 27,704 observations except for the training measures where it is 25,115 observations.

Table A2: Relationships Among Wages, Schooling, AFQT, and Parental Education Simple Regression Coefficients (standard error) and [Correlation coefficient]

	Dependent Variable								
Right Hand Side Variable	Log Wage	Highest Grade	Father's Education	Standard. AFQT	Weeks of Company Training	Layoff	Quit	Actual Experience	Potential Experience
Highest Grade	0.0785 (0.0014) [0.3615]		0.6197 (0.0098) [0.4029]	0.2747 (0. 0027) [0.5829]	0.1189 (0.0163) [0.0514]	-0.0193 (0.0010) [-0.1259]	-0.0128 (0.0014) [-0.0589]	-0.0823 (0.0103) [-0.0329]	-0.4831 (0.0106) [-0.2923]
Father's Education	0.0298 (0.0010) [0.2092]	0.2592 (0.0041) [0.4029]		0.1341 (0.0019) [0.4362]	0.0621 (0.0106) [0.0392]	-0.0059 (0.0007) [-0.0542]	0.0014 (0.0009) [0.0112]	-0.0323 (0.0067) [-0.0331]	-0.1660 (0.0071) [-0.1538]
Standardized AFQT	0.1565 (0.0031) [0.3567]	1.2245 (0.0119) [0.5829]	1.4280 (0.0204) [0.4362]		0.3072 (0.0345) [0.0645]	-0.0377 (0.0021) [-0.1174]	-0.0121 (0.0029) [-0.0306]	-0.1036 (0.0218) [-0.0142]	-0.8138 (0.0227) [-0.2329]
Weeks of Company Training	0.0045 (0.0007) [0.0429]	0.0214 (0.0029) [0.0514]	0.0268 (0.0046) [0.0392]	0.0124 (0.0014) [0.0645]		-0.0011 (0.0004) [-0.0130]	-0.0017 (0.0006) [-0.0190]	-0.0173 (0.0044) [-0.0282]	-0.0297 (0.0047) [-0.0453]
Layoff	-0.1659 (0.0104) [-0.1094]	-0.8921 (0.0468) [-0.1259]	-0.6558 (0.0728) [-0.0542]	-0.3904 (0.0222) [-0.1174]	-0.2702 (0.1112) [-0.0130]		-0.2707 (0.0092) [-0.2080]	-1.1766 (0.0696) [-0.1391]	-0.3223 (0.0753) [-0.0629]
Quit	-0.2145 (0.0076) [-0.1909]	-0.3232 (0.0348) [-0.0589]	0.0814 (0.0539) [0.0112]	-0.0683 (0.0165) [-0.0306]	-0.2321 (0.0821) [-0.0190]	-0.1478 (0.0050) [-0.2080]		-1.2747 (0.0510) [-0.1658]	-0.8834 (0.0553) [-0.1070]
Actual Experience	0.0444 (0.0010) [0.2893]	-0.0374 (0.0047) [-0.0329]	-0.0350 (0.0072) [-0.0331]	-0.0106 (0.0022) [-0.0142]	-0.0436 (0.0110) [-0.0282]	-0.0116 (0.0007) [-0.1391]	-0.0230 (0.0009) [-0.1658]		0.8605 (0.0045) [0.7953]
Potential Experience	0.0174 (0.0010) [0.1044]	-0.1899 (0.0042) [-0.2923]	-0.1561 (0.0066) [-0.1538]	-0.0718 (0.0020) [-0. 2329]	-0.0647 (0.0103) [-0.0453]	-0.0027 (0.0006) [-0.0629]	-0.0138 (0.0009) [-0.1070]	0.7452 (0.0039) [0.7953]	

Appendix 1

From equation (11) we have $\Phi_s = -\mathrm{cov}(s,z) * \frac{\Lambda var(v) + \mathrm{cov}(v,e)}{|var(s,z)|}$ and

$\Phi_z = -\mathrm{var}(s) * \frac{\Lambda var(v) + \mathrm{cov}(v,e)}{|var(s,z)|}$. We know that $\Phi_{zs} = \frac{\mathrm{cov}(s,z)}{var(s)}$. This gives us the desired result:

$$\Phi_s = -\Phi_{zs}\,\Phi_z$$

Appendix 2: Derivation of Equation (16) and (17).

Consider equation (15). Rewriting $\mathrm{var(s,z)}^{-1}$ as a partitioned matrix leads to

$$\mathrm{var}(s,z)^{-1} = \begin{bmatrix} \mathrm{var}(s) & \mathrm{cov}(s,z) \\ \mathrm{cov}(z,s) & \mathrm{var}(z) \end{bmatrix}^{-1}$$

where var(s,z) is the (K+1)x(K+1) variance matrix.

Using the partitioned inverse formula and ignoring the first column (since it will be multiplied by 0), we have:

$$\text{(15a)}\quad E\begin{bmatrix} b_{st} \\ b_{zt} \end{bmatrix} = \begin{bmatrix} b_{s0} \\ b_{z0} \end{bmatrix} + \begin{bmatrix} -\mathrm{cov}(s,z) * G \\ \mathrm{var}(s) * G \end{bmatrix} * \left[\mathrm{cov}(z, E(\Lambda v + e | D_t))\right]$$

where $G = \left[\mathrm{var}(s) * \mathrm{var}(z) - \mathrm{cov}(z,s) * \mathrm{cov}(s,z)\right]^{-1}$

Now, consider the diagonal matrix K which has elements of cov(z, Λv+e) along the diagonal. K^{-1} is also diagonal. Thus (15a) may be rewritten as:

$$\text{(15b)}\quad E\begin{bmatrix} b_{st} \\ b_{zt} \end{bmatrix} = \begin{bmatrix} b_{s0} \\ b_{z0} \end{bmatrix} + \begin{bmatrix} -\mathrm{cov}(s,z) * G \\ \mathrm{var}(s) * G \end{bmatrix} * K * K^{-1} * \left[\mathrm{cov}(z, E(v+e | D_t))\right]$$

Manipulating this further gives us:

$$\text{(15c)}\quad E\begin{bmatrix} b_{st} \\ b_{zt} \end{bmatrix} = \begin{bmatrix} b_{s0} \\ b_{z0} \end{bmatrix} + \begin{bmatrix} -\mathrm{cov}(s,z) * G \\ \mathrm{var}(s) * G \end{bmatrix} * \Theta^t * \left[\mathrm{cov}(z, \Lambda\, v + e)\right]$$

where Θ^t is s diagonal matrix with element kk equal to $\Theta^t_{kk} = \frac{\mathrm{cov}(z_k, E(v+e|D_t))}{\mathrm{cov}(z_k, v+e)}$. Using the definition of Θ^t_{kk} and evaluating the above equation leads to (16) and (17) in the text.

Appendix 3

The regression parameters $b^*_{z_1t}$ and b_{z_2t} are

$$b^*_{z_1t} = b^*_{z_10} + \Phi^*_{z_1}\theta_t$$

$$b_{z_1t} = b_{z_10} + \Phi_{z_1}\theta_t$$

where $\Phi^*_{z_1}$ is the coefficient on z_1 in the regression of $\Lambda v + e$ on s and z_1 and Φ_{z_1} is the coefficient on z_1 in the regression of $\Lambda v + e$ on s, z_1, and z_2. By the omitted variables formula, we know that $\Phi^*_{z_1} = \Phi_{z_1} + \Phi_{z_2}\Phi_{z_2z_1}$ where Φ_{z_2} is the coefficient on z_2 in the regression of $\Lambda v + e$ on s, z_1 and z_2 and $\Phi_{z_2z_1}$ is the coefficient on z_1 in the regression of z_2 on z_1 and s. Therefore,

$$\frac{\partial b_{z_1t}}{\partial t} = \Phi_{z_1}\frac{\partial \theta_t}{\partial t} \text{ and}$$

$$\frac{\partial b^*_{z_1t}}{\partial t} = \Phi^*_{z_1}\frac{\partial \theta_t}{\partial t} = (\Phi_{z_1t} + \Phi_{z_2}\Phi_{z_2z_1})\frac{\partial \theta_t}{\partial t}$$

Taking the difference establishes that

$$\frac{\partial b_{z_1t}}{\partial t} - \frac{\partial b^*_{z_1t}}{\partial t} = -\frac{\partial \theta_t}{\partial t}\cdot[\Phi_{z_2}\Phi_{z_2z_1}]$$

where Φ_{z_2} is the coefficient on z_2 in the regression of $\Lambda v + e$ on s, z_1 and z_2 and $\Phi_{z_2z_1}$ is the coefficient on z_1 in the regression of z_2 on z_1 and s.