Employer Learning, Statistical Discrimination and Occupational Attainment

Joseph G. Altonji*

Statistical discrimination plays an prominent role in theoretical discussions of the return to education as well as race and gender differences.[1] Building on some earlier work by Henry Farber and Robert Gibbons (1996) (hereafter FG) and others on employer learning in the labor market, Altonji and Pierret (2001) (hereafter AP) provide a test for statistical discrimination in an environment in which firms are learning about characteristics of workers over time. However, their model assumes that the rate at which employers learn about a given worker is independent of the type of job that he or she is in. Consequently, their framework cannot readily be used to study statistical discrimination in hiring and or how statistical discrimination and employer learning influence the skill level of the initial occupation and progression over a career.

In this paper, I extend AP's analysis in three key respects. First, as in Robert Gibbons et al. (2004), the sensitivity of output to worker skill is assumed to depend on the skill level of the job. Second, the rate at which employers learn about the worker's skill is assumed to depend on the skill level of the job. Finally, the probability of being hired into a given job depends on expected productivity relative to the wage. I show that statistical discrimination influences initial employment rates, wage levels and job type, and that employers' initial estimate of productivity influences wage growth even in an environment in which access to training is not an issue. The implication is that the market may be slow to learn that a worker is highly skilled if worker's best early job opportunity given the information available to employers is a low skill level job that reveals little about the worker's talent. I close with a brief discussion of directions for empirical work suggested by the analysis.

I. Statistical Discrimination When Employer Learning Depends on Job Skill Intensity

The model draws upon FG and AP as well as a number of papers that stress the importance of comparative advantage across occupations on the basis of general skill, including Oded Galor and Nachum Sicherman (1990) and Gibbons et al. The latter two papers cite a rich literature that deals with other important factors that influence occupation choice and mobility, including preferences, abilities or training that are occupation specific, which I abstract from here.

As a result of parents, schooling, and other factors, a worker comes to the labor market with level h of general skill/human capital. Net output y of a worker depends on h and the skill level of the job, indexed by j. The higher j, the more sensitive output is to h. A production function with this key feature is

$$(1) \qquad y = c + \mu h + ajh - .5bj^2 \ ; \ \mu > 0,\ a > 0,\ b > 0,$$

where c, μ, a, and b are parameters. The interaction term ajh means that more skilled workers have a comparative advantage in jobs in which skills matter more for productivity. FG and AF abstract from matching issues and essentially assume that $y = c + \mu h$. The negative quadratic term in j captures the idea that overhead costs rise with j. One could let c depend on labor market experience without changing the analysis. I set c to 0 without loss of generality.

Following FG and AP, assume that firms and the worker have the same information about h, although an extention to the case of private information would be highly desireable. Firms and the worker are imperfectly informed about h at the start of the worker's career and learn over time. Their initial information about h consists of the vector s that is observed by both the agents and the econometrician and the vector q that is observed by the agents but not the econometrician. In addition, h depends on the vector z that that is not observed directly by agents but is observed by the econometrician and on the scalar index η, which neither the agents nor econometrician observe. Examples of s include education, race and gender. Skill is related to the four sets of factors through the equation

$$(2) \qquad h = rs + \alpha q + \Lambda z + \eta.$$

where r, α and Λ are fixed parameters. There is scope for statistical discrimination on the basis of s and q because they help predict z and η. Let $E(.|.)$ be the conditional expectation function. Denote the employers' and the worker's first period belief $E(h|s,q)$ about h by $\hat{h}_1$ and define v to be the expectational error.

Careers last two period. After the first period, all agents observe the signal $\xi = (\mu + aj_1)(\hat{h} + v) + \varepsilon$, where j_1 is worker's first period job type and ε is independent noise. Firms account for $\hat{h}_1$ when interpreting ξ, so observing ξ is equivalent to observing $D = (\mu + aj_1)v+$

ε. Assuming expectations are linear,

$$\hat{v} \equiv E(v|j_1, D) = \frac{(\mu + aj_1)var(v)}{(\mu + aj_1)^2 var(v) + var(\varepsilon)} D. \tag{3}$$

Workers are risk neutral expected lifetime income maximizers. Initially I focus on a version of the model in which firms pay expected productivity given s and q and workers choose j to maximize expected income. An alternative interpretation with the same mathematical structure is that creation of type j jobs and competition among firms establishes a market wage rate $w(\hat{h})$ for each value of $\hat{h}$ with zero profits for all j. Firms fill type j jobs with workers who have the level of $\hat{h}$ that maximizes expected output net of wages. Job queues and rationing don't arise because the choices of firms and workers agree. See Section IV, where I go on to modify the model to allow for unemployment.

II. Effects of Statistical Discrimination on the First Period Job Type and Wages.

In the first period the worker chooses the job type j_1 that maximizes wages and receives the wage w_1. j_1 and w_1 depend on s and q and are determined by

$$w_1 = \max_j (\mu \hat{h}_1 + aj\hat{h}_1 - .5bj^2). \tag{4}$$

Differentiation establishes that $j_1 = a\hat{h}_1/b$ and is increasing in $\hat{h}_1$. To simplify the analysis of the choice of j_1, I have assumed that workers ignore the fact that j influences how much employers learn about v in the first period even though such information is valuable for job choice in the second period. Such considerations would lead workers to choose a higher value of j than $\frac{a}{b}\hat{h}_1$ and receive a lower wage. I am also ruling out internal labor markets in which firms design jobs with the purpose of generating information about worker skills in mind, as analyzed in O'Flaherty and Siow (1995).

Solving (4) and replacing j_1 with $a\hat{h}_1/b$ establishes that the relationship between w_1 and $\hat{h}_1$ is

$$w_1 = \mu \hat{h}_1 + .5\frac{a^2}{b}(\hat{h}_1)^2. \tag{5}$$

Equation (5) implies the following proposition about initial occupation and wages.

Proposition 1: The skill level j_1 and first period wage w_1 are increasing in the elements of s and q that positively affect $\hat{h}_1$.

Part of the link from s and q to j_1 and w_1 is due to the fact that both s and q are directly related to h conditional on the unobserved variables z and η. However, Proposition 1 would hold even if s and q do not enter (2) (e.g.,, $r = \alpha = 0$) provided that they are helpful in predicting z and η and firms statistically discriminate on the basis of them. Note also that the link between w_1 and $\hat{h}_1$ is convex even though y is linear in $\hat{h}_1$ for a given job type. The convexity results from the fact that higher $\hat{h}_1$ workers obtain jobs in which skills are more valuable.

Shelly Lundberg and Richard Startz (1983), Kevin Lang (1986) and others have presented models in which the quality of information that employers obtain is lower and their ability to interpret that information is lower for minority group members. This has negative effects on the incentives of minority group members to make investments that are hard to observe directly. In the model above one might think of such considerations as leading to differences across groups in the response of s and q to particular actions taken by individuals or as differences across groups in the response of $\hat{h}_1$ to a particular elements of s and q, such as early work experience or education. Lundberg (1991) stresses that when firms face an assignment problem group differences in the precision of employer information lead to differences in average productivity of the group. In the above model, one may think of $\hat{h}_1$ as a random variable that is distributed in accordance with the distribution of s, q for a given population group. The group variance of $\hat{h}_1$ will be higher the greater the information content of s, q for the group. It is obvious that this implies greater dispersion of w_1, but using (5) one can easily show that the group mean of the w_1 is increasing in the group variance of $\hat{h}_1$, holding the group mean of $\hat{h}_1$ constant. This establishes that the amount of information that firms have about a group positively influences the average wage for the group even if there are no feedbacks into investment.

III. The Effect of Employers' Initial Beliefs about Productivity on Wage Growth

I now turn to wage growth and occupational mobility. After substituting $a\hat{h}_1/b$ for j_1 in (3), some simple algebra leads to

$$\hat{v} = \theta_{\hat{h}} v + \theta_{\hat{h}} (b/a^2 \hat{h}_1)\varepsilon \qquad (6)$$

where

$$\theta_{\hat{h}} = \frac{(\mu + \frac{a^2\hat{h}_1}{b})^2 var(v)}{(\mu + \frac{a^2\hat{h}_1}{b})^2 var(v) + var(\varepsilon)} \tag{7}$$

In period 2 the worker is again paid expected output and chooses j_2 accordingly given $\hat{h}_1$ and D. The optimal choice is $j_2 = a[\hat{h}_1 + \hat{v}]/b$, which means that $j_2 - j_1 = a\hat{v}/b$. The second period wage w_2 depends on the initial information s, q, and the signal D through $\hat{h}_1, \hat{v}$ in accordance with

$$w_2 = \mu(\hat{h}_1 + \hat{v}) + .5\frac{a^2}{b}(\hat{h}_1 + \hat{v}))^2. \tag{8}$$

The mean of w_2 conditional on the employers' initial skill estimate $\hat{h}_1$ is

$$E(w_2|\hat{h}_1) = \mu\hat{h}_1 + .5\frac{a^2}{b}[(\hat{h}_1)^2 + var(\hat{v})] = w_1 + .5\frac{a^2}{b}var(\hat{v}).$$

From (7) one can see that θ_h is strictly increasing in $\hat{h}_1$ and approaches 1 as $\hat{h}_1$ goes to infinity. This fact and some straightforward calculations establish that $var(\hat{v})$ and thus the second term of the above equation is strictly increasing in $\hat{h}_1$.[2] This establishes the following proposition:

PROPOSITION 3. Average Wage Growth is Increasing in Employers' Initial Belief $\hat{h}_1$ about Productivity.

Proposition 3 implies immediately that expected wage growth is a positive (negative) function of s if $\hat{h}_1$ is increasing (decreasing) in s. The result stands in contrast to AF's result that s will be unrelated to wage growth. It holds despite the fact that both the revision $\hat{v}$ in beliefs about worker skill and the average change in j between period 1 and 2 have mean 0 regardless of initial beliefs $\hat{h}_1$. The result derives from the fact that persons with larger values of $\hat{h}_1$ start out in jobs in which skills matter more for output and that in such jobs there is a greater flow of information about v. Workers with positive $\hat{v}$ move to more skilled jobs and achieve wage gains that more than offset the losses of workers with a negative value of $\hat{v}$ of the same absolute value. Altonji and Pierret (1998) and Lange (2004) provide evidence that on average employers learn about productivity quickly enough to severely restrict the scope of signalling models of the return to education, but their argument is weakened if the employer learning is rapid only if the worker acquires education.

To get a better sense of the implications of the result consider a worker with very low $\hat{h}_1$. For such an individual j_1 will be low and the update $\hat{v}$ to employer beliefs will tend to be small even if h and thus $v = h - \hat{h}_1$ are high. Consequently, high skill individuals with observable characteristics that are associated with low skill will tend to be trapped for a while in low skill jobs. With enough time, the market will learn, but the retarding effect on upward mobility of low information flow in low skill jobs would be reinforced if training or learning by doing depends positively on j.[3] By the same token, the odds of moving to a very high level position over the course of a career is influenced by $\hat{h}_1$ for a given h. Suppose that top management positions in major corporations have a skill level above some value j^*. The odds that a person ends up with $j_2 > j^*$ is equal to the odds that $\hat{h}_1 + \hat{v} > aj^*/b$. To reach such a position, one must not only have high h but must be "discovered" by the market. Holding h fixed, a lower value of $\hat{h}_1$ implies a larger value of v, which tends to increase $\hat{v}$. However, low $\hat{h}_1$ will lead to a lower value of j_1. This will reduce the information flow, reduce $\hat{v}$, and make it unlikely that $\hat{h}_1 + \hat{v}$ will be high enough to warrant a high skill position. Consequently, individuals who start with a low value of $\hat{h}_1$ because of statistical discrimination on the basis of educational credentials, family background, or race/ethnicity will be unlikely to reach the most skill intensive positions in the economy. Prior to the market workers with a low value for one element of s who know that they are high h may have an incentive to signal this to the market through another s variable, such as education.

Finally, one may show that the variance of $w_2 - w_1$ is increasing in $\hat{h}_1$. Consequently, the model provides another explanation for the common finding that the growth in dispersion of wages with experience depends positively on education.

IV. Employment

I now extend the framework to allow the odds that a worker is hired into job j to depend on productivity relative to the wage. The production function is (1). Firms base hiring decisions on $\hat{h}$. I impose that assumption that they pay a fixed wage $w(j)$ to hires regardless of $\hat{h}$. They choose $w(j)$ to be $w(j) = w(\hat{h}(j))$, where $\hat{h}(j)$ is the average value of $\hat{h}$ of workers hired in job j. Given the distribution of $\hat{h}$, firms create jobs of type j until expected profits equal 0 and so $w(\hat{h}(j)) = \mu\hat{h}(j) + aj\hat{h}(j) - .5bj^2$. Workers pick j to maximize wages, and $j = a\hat{h}/b$ as before. To find $w(\hat{h})$, substitute out for j in above the equation, which leads to

$$w(\hat{h}(j)) = \mu\hat{h}(j) + .5a \cdot (\frac{a}{b}\hat{h}(j))^2. \tag{9}$$

One may verify using the wage function and (1) that the firm's choice of $\hat{h}(j)$ that solves $dw(\hat{h}(j))/d\hat{h} = dy(\hat{h}(j), j)/d$ is $\hat{h}(j) = bj/a$, which is consistent with the worker's choice of j. By substituting out for $h(j)$ in the wage equation one obtains the $w(j)$ function

$$w(j) = \mu\frac{b}{a}j + .5bj^2. \tag{10}$$

Given $w(j)$ all workers prefer higher j jobs but are constrained by firms.

One can introduce unemployment by assuming that workers only contact one employer per period, and that the value to the firm of net output y is the product of y and a transitory, idiosyncratic component p. The firm will hire a worker with $\hat{h}$ if the $p \cdot y/w(j) = \frac{p(\mu\hat{h}+aj\hat{h}-.5bj^2)}{w(j)} > 1$. Firms make a profit when they actually hire, but I assume that this is counterbalanced by fixed costs of operating, which are not modeled. Entry by firms leads $w(j)$ to adjust to the point that profits associated with creating a job j are 0, and (10) may not hold.

Assume that $1/p$ is uniformly distributed between $1/p_{\max}$ and $1/p_{\min}$. Then the probability of being hired in job j for a worker with $\hat{h}$ is

$$(\frac{(\mu\hat{h} + aj\hat{h} - .5bj^2)}{w(j)} - 1/p_{\max})(1/p_{\min} - 1/p_{\max})^{-1}, \tag{11}$$

which is increasing in the ratio of productivity to pay in job type j. This version of the model makes explicit the idea that an entry level high school graduate is unlikely to get hired as the CEO of a major company no matter what wage he would accept.

Assume in addition that product demand and technology are such that $j \geq j_{\min}$. The worker targets search at the j that maximizes the product of $w(j)$ and the hiring probability. When the job skill constraint is not binding the first order condition is

$$(a\hat{h} - bj) - (w(j)/p_{\max})](1/p_{\min} - 1/p_{\max})^{-1} = 0.$$

Since $w'(j) > 0$, $j(\hat{h})$ and therefore $w(j(\hat{h})$ rise with $\hat{h}$. This fact and (11) means that the sign of the relationship between $\hat{h}$ and employment is ambiguous. The situation is different for workers with $\hat{h} < \frac{b}{a}j_{\min}$, for whom the minimum skill level constraint binds. These workers seek $j_{\min}$ jobs and obtain them with probability

$$(\frac{(\mu\hat{h} + aj_{\min}\hat{h} - .5bj_{\min}^2)}{(w(j_{\min})} - 1/p_{\max})(1/p_{\min} - 1/p_{\max})^{-1},$$

which is less than the employment probability for workers with $\hat{h} = \frac{b}{a}j_{\min}$. Furthermore, the employment probability is linear in $\hat{h}$ in the range below $\frac{b}{a}j_{\min}$ even though the reservation wage is 0. In practice, one would expect some downward pressure on $w(j)$ near $j_{\min}$ and for firms to adopt wage policies of the form $w(j_{\min}, \hat{h})$. Such adjustments would reduce unemployment and weaken the association with $\hat{h}$. However, heterogeneity in the value of nonwork time, efficiency wage considerations and minimum wages would limit this.

A key point is that unemployment will interact with early statistical discrimination to deter learning. For unemployed workers, there is no first period signal D, $\hat{v}$ is 0, and $\hat{h}_2$ will equal to $\hat{h}_1$. Given that the employment probability and therefore the odds of observing a signal are increasing in $\hat{h}_1$ for workers with $\hat{h}_1 < \frac{b}{a}j_{\min}$, this reinforces the earlier result that learning is increasing in $\hat{h}_1$. The market may be slow to learn the true productivity of high h individuals with poor s and q characteristics. It is also consistent with the idea that the unemployment probability will be decreasing in $\hat{h}_1$ and unemployment might further restrict the information flow that would enable high h, low $\hat{h}_1$ individuals to move up to jobs that value h.

IV. Empirical Implications for Wage Growth, Employment, and Occupation Change

AP develop a test for whether firms statistically discriminate on the basis of s variables such as education or race. Their test is based on the fact that if wages are determined by expected productivity given employer information, then as employers learn about v the relationship between wages and s and z will change. The above analysis suggests that the effect of z on wage growth increases with s. The intuition for this conjecture is that s raises j_1, which leads to larger flow of information about v. Because v is correlated with z, the increased information flow strengthens the link between z and wage growth. This line of reasoning suggests that in the regression

(12) $$w_t = \beta_0 + \beta_1 s + \beta_2 z + \beta_3 s \cdot t + \beta_4 z \cdot t + \beta_5 s \cdot z \cdot t + u_t$$

$\beta_5 > 0$. Going beyond the intuition and formally establishing the properties of the regression $E(w_t|s, z)$ using the model above is difficult because of the presence of the squared term in (5) and (8) and the fact that in general $\theta_{\hat{h}}$ is related to z as well as s is stochastic conditional on z and s but should be carefully explored in future research.

Another implication of the model is that $j_2 - j_1 = a\hat{v}/b$, which implies that the changes in occupational skill level are more strongly related to v and thus to z variables when $\hat{h}_1$ is high. One could relate proxies for the change in j to z for different values of s. Finally, the model seems to have implications for the variance of wage changes as a function of experience and for the experience profile of unemployment for low skill groups.

I have abstracted from other prominent features of models of wage growth over a career and job choice, including general training that varies with the type of job, occupation specific training, multiple skill types, and heterogenous preference over job characteristics types. Indicators of productivity y and job training as well as data on job characteristics, s variables, and z variables are probably needed if the career consequences of statistical discrimination and employer learning is to be properly assessed.

Notes

* Department of Economics, Yale University, New Haven, CT 06520. I thank seminar participants at Yale, Kerwin Charles, Fabian Lange, and Xin Yu for helpful comments and the NSF for support under grant SES-0112533.

[1] See Altonji and Blank (1999) for a survey.

[2] To establish that $\partial Var(\hat{v})/\partial \hat{h} > 0$, note that $Var(\hat{v}) = (\frac{qvar(v)}{qvar(v)+var(\varepsilon)})^2(var(v)+var(\varepsilon)/q)$ where $q = (\mu + \frac{a^2\hat{h}_1}{b})^2$. Defining $Q_v = var(v)/[var(v) + var(\varepsilon)]$, one may show that $\partial\frac{Var(\hat{v})}{(var(v)+var(\varepsilon)}/\partial q = -Q_v^2 \cdot \frac{Q_v-1}{(qQ_v+1-Q_v)^2}$, which is greater than 0 when $(Q_v - 1) < 0$. The latter condition holds given $var(\varepsilon) > 0$. (If $var(\varepsilon) = 0$, D is perfectly informative about v regardless of j_1 and thus regardless of $\hat{h}_1$.) Since q is strictly increasing in $\hat{h}_1$, $Var(\hat{v})$ is strictly increasing in $\hat{h}_1$. The result is robust to allowing $var(\varepsilon)$ to grow with j_1 at a rate that is less than proportional to the growth in $(\mu + aj)$.

[3] Heisz and Oreopoulos (2003) analyze a somewhat similar model in which $\hat{h}_1$ is positively correlated with wage growth because because statistical discrimination in the first period influences access to training opportunities. They do not consider the effects of the job type on employer learning, which is the focus here.

References

Altonji, Joseph G. and Blank, Rebecca. "Race and Gender in the Labor Market," in O. Ashenfelter and D. Card, (eds), *Handbook of Labor Economics* Volume 3c Elsevier Science B.V.,1999, pp. 3144-3259

Altonji, Joseph G. and Charles R. Pierret. "Employer Learning and the Signalling Value of Education," in I. Ohashi and T. Tachibanaki (eds), *Internal Labour Markets, Incentives and Employment*, McMillan Press, 1998, pp 159-195

Altonji, Joseph G. and Charles R. Pierret. "Employer Learning and Statistical Discrimination." *Quarterly Journal of Economics*, February 2001, 116(1), pp 313-350.

Coate, Stephen and Glenn Loury. "Will Affirmative-Action Policies Eliminate Negative Stereotypes?", American Economic Review, December 1993, 83(5), pp. 1220-1240.

Gibbons, Robert, Lawrence F. Katz, Thomas Lemieux, and Daniel Parent. "Comparative Advantage, Learning, and Sectoral Wage Determination", unpublished paper, Harvard University, revised July 2004.

Heisz, Andrew and Philip Oreopoulos. "The Importance of Statistical Discrimination in Job Placement and Promotion: Evidence Using MBAs and Lawyers from Different-Ranked Schools", mimeo, University of Toronto, 2003.

Lang, Kevin. "A Language Theory of Discrimination", *Quarterly Journal of Economics*, May 1986, 101(2)), pp. 363-82.

Lange, Fabian. "The Returns to Schooling and Ability During the Early Career: Evidence on Job Market Signaling, Employer Learning and Post-School Investments", Ph.D. dissertation, University of Chicago, 2004.

Lundberg, Shelly J. and Richard Startz. "Private Discrimination and Social Intervention in Competitive Labor Markets." *American Economic Review*, June 1983, 73(3), pp. 340-47.

O'Flaherty, Brendan and Aloysius Siow. "Up-or-Out Rules in the Market for Lawyers." *Journal of Labor Economics*, October 1995, 13(4), pp. 709-735.

Sicherman, Nachum and Oded Galor. "A Theory of Career Mobility", *The Journal of Political Economy*, February 1990, 98(1), pp. 169-192.