Signaling, Self - Selection and Employer Learning*†‡

(Job Market Paper)

Barış Kaymak‡

University of Rochester

November 28, 2006

Abstract

The purpose of this paper is to quantify the signaling function of education in the labor market. Using a tractable model of educational choice under asymmetric information, this paper shows that workers with higher ability sort into the markets where education does not play an important signaling role. Weaker incentives to signal ability lead to lower educational attainment in these markets conditional on ability. Furthermore the education-ability profiles are flatter. Using data from the NLSY, the market is grouped into two occupational categories based on the average AFQT scores, and the differences in educational profiles are examined. Allowing for differences in the ability and the productivity of human capital across these groups, the difference in educational attainment between the two groups implied by the human capital model is not compatible with that observed in the data. Allowing for a differential role for signaling reconciles the observed differences with those predicted by the model. Results suggest that signaling is particularly important for workers of low ability. The contribution of signaling to wages for these workers is 22% of the return to education estimated by OLS. For the higher ability workers, the return to signaling is much smaller.

*I am thankful to Mark Bils for invaluable advice. I also thank Ronni Pavan, Paulo Barelli, Gordon Dahl and Fatih Guvenen for helpful discussions and suggestions.
†Please visit www://troi.cc.rochester.edu/~kymk for updated versions
‡Department of Economics, University of Rochester, PO Box 270156, Rochester, NY-14627. E-mail: kymk@troi.cc.rochester.edu
1 Introduction

The positive association between wages and educational attainment is one of the most consistent empirical findings in economics. The literature provides two competing theoretical explanations for this relationship. The human capital theory pioneered by Becker (1964) and Mincer (1974) suggests that education augments the market productivity of a worker. The signaling theory initiated by Spence (1973), on the other hand, points out that the informational problems in identifying the productivity of workers before hiring them may render education as a screening device that separates productive workers from the less productive ones without actually contributing to their marketable skills. The distinction between the two explanations is crucial since their implications on education as a social investment are different. The main contribution of this paper is to quantify the relative importance of the two functions of education.

Both theories predict a similar set of empirical outcomes making it fundamentally difficult to distinguish the two models in a single market. This difficulty has been expressed several times in the literature. Lang and Kropp (1986) and Bedard (2001) provide tests of the signaling hypothesis based on the premise that in a signaling equilibrium, changes in the costs or benefits of education for a particular group affect the entire distribution of educational choices, contrary to what the human capital theory predicts. Their results indicate at least a partial signaling role for education, but the magnitude of this role remains unknown. One needs at least two markets to quantitatively differentiate between these theories. Earlier studies has focused on the comparison of markets where the signaling incentives presumably differ in strength, such as the self employed versus salaried workers (Wolpin 1977), or occupational/industrial categories (Riley 1979b). The latter categorizes sectors into two groups, unscreened and screened, assuming that sectors are not segregated with respect to ability. This assumption is hard to justify given the clear differences in measured dimensions of workers’ ability across sectors. The observed patterns of worker sorting has been analyzed within the human capital model where markets differ in their return to education and ability, but the effect of differences in the signaling role of education on the occupation choice is not examined.

This paper develops a tractable model of signaling with endogenous occupation choice, and shows that the occupations where the signaling role of education is limited attract workers with higher ability. All else equal, workers in these occupations attain lower education conditional on their ability, and the education-ability gradient within the occupation is flatter. Using data from the National Longitudinal Survey of Youth (NLSY), I then group occupational categories based on the average Armed Forces

---

1See Lange and Topel (2005) for a survey of the existing empirical literature on signaling and the social value of education

2For instance, Murphy and Topel (1990) and Blackburn and Neumark (1992) find evidence on sorting of workers by ability across industries.

3Gibbons et al. (2005) and Altonji (2005) are recent examples of such studies.
Qualification Test (AFQT) score of workers, and compare the educational profiles of workers given their test scores. I examine whether the differences in the productivity of human capital and ability alone can account for the observed differences in educational achievement. The model fails to generate these differences without differential role for signaling. The differences in educational achievement suggested by these factors is much higher than what is observed. I then simulate the model allowing for differential return to signaling, and find that this difference can be reconciled only if the signaling role of education is significantly higher in the low ability occupations.

In a related literature Farber and Gibbons (1996) and Altonji and Pierret (2001) test whether the market learns about the true productivity of their workers by observing them at work. They build on the assumption that the econometrician has access to a correlate of a worker’s market skills that is not directly available to the employers. They show that the wage distribution increasingly reflects the variation in this correlate over the workers’ experience indicating the presence of employer learning. Although this implies that the market is uncertain about the workers’ productive qualities, these empirical patterns can not be considered as direct tests for signaling per se. The results are also compatible with a market where learning is symmetric between the worker and the employer. The difference is crucial for identifying the signaling role of education, because the main underlying assumption behind signaling is that workers have an informational advantage over the employers.

Altonji and Pierret (1997) and more recently Lange (2005) point out the relevance of the presence of employer learning to the size of the informational role of education. Employer learning reduces the workers’ incentives to use educational attainment in order to signal their productivity. If the output of each worker is accurately observable on the job, then the employers would learn the true productivity of their workers in a short period of time. The faster the market finds out the productivity type of a worker, the smaller is the need to signal. Based on this observation, Lange (2005) estimates the speed with which employers learn by gauging how well the AFQT score captures the variation in wages at different experience levels. This provides him with an upper bound on the size of the signaling. However, this bound is crucially dependent on the assumed discount rate. The nature of the bound derives from the condition that the private return to education can not exceed the total observed return. If we take a stand on the discount rate, or equivalently the private return to education, then the difference between total return and the assumed private return can be used extract the signaling bias simply using the optimality condition for education choices. Lange (2005) calculates the bound on the contribution of signaling to vary from zero to 26%.

AFQT score is a widely used proxy for ability in the literature. Even if the unknown component of productivity is correlated with education and employers factor this correlation into wages over experience, it would still not be signaling unless this variable fundamentally distinguishes workers with respect to their marginal cost or benefit of education under the symmetric information environment. The literature refers to this phenomenon as statistical discrimination based on education.
of the private return for a range of 3% to 9% for the discount rate for his benchmark specification.

This result suggests that progress on the assessment of signaling is possible by modeling the education choices combined with a structure on the employer learning process. Each assumed private return implies a (an average) level of education consistent with the structure of the model\textsuperscript{6}. If we restrict the private return to match the observed educational attainment, then a precise bound on the return to signaling can be obtained, but this is still a bound as discussed next. Introduction of two sectors enables the estimation of the common discount rate, and the difference in the level of signaling\textsuperscript{7} even when the ability differences across markets are taken as exogenous. Introducing endogeneity in the occupation choice further allows the signaling role in each sector to be estimated.

The connection from the speed of employer learning to the size of signaling is also clouded when employers use other devices to screen their workers. Both Riley (1979b) and Lange (2005) base their results on the distribution of wages that can only be observed once the worker is hired. However, pre-screening via cognitive tests, interviews or reference letters reduces the amount of uncertainty about the workers’ productivity prior to entry. This is not captured in the wage regressions\textsuperscript{8}. Two markets with different speeds of learning might employ different strategies to examine their workers prior to employment. If anything, one would expect that in markets where it is hard to observe an individual worker’s output, employers would invest in other screening devices more intensively. This prevents measuring the return to signaling directly by the estimated speed of learning alone.

The model presented here incorporates the employer learning, which in turn generates an intuitive hybrid model of human capital and signaling. I estimate the speed of learning for the two occupational categories and find that the employer learning is in fact slower for workers with higher ability. Therefore the extent of uncertainty that is not resolved by the pre-screening devices prior to entry can not be similar across these groups. Otherwise, higher average return to signaling would exacerbate the already wide gap between education profiles implied by the human capital model.

I use the estimated speeds of learning in the simulations and allow the size of the initial uncertainty upon employment (after the workers are screened already by additional devices) to vary across these sectors in the simulations. Estimating the speed of learning disciplines the extent of educational differences that can be explained by signaling role of education. The results suggest that for the set of occupations with higher average AFQT scores, most of the uncertainty is unraveled prior to employment.

\textsuperscript{6}This paper assumes that agents maximize the expected value of their lifetime income.

\textsuperscript{7}The significance of the common discount rate assumption is discussed in section 4. If the discount rate is decreasing in ability, then the results presented here present a lower bound on signaling.

\textsuperscript{8}Under rather strong assumptions, some information on the relative importance of additional screening devices can be extracted using the variation in starting salaries. This is discussed in section 2.2.
For the low AFQT category only 24% of the uncertainty is eliminated by pre-screening devices. Combined with the estimated speed of learning, this implies that the return to signaling in this category is 22% of the observed OLS return. On the contrary, for the occupations with higher ability the implied level of signaling is negligibly small. This is in accordance with the endogenous sorting of workers with higher ability into occupations where the signaling role of education is smaller.

Section 2 presents a single sector model of educational choice under asymmetric information and outline the major difficulties in testing the signaling hypothesis. I then model the endogenous market choice and analyze the optimal occupational choices allowing for differences in the size of signaling and the productivity of human capital. Section 3 examines whether the model with fully separating equilibrium presented here can generate the observed empirical patterns in the employer learning literature. This requires somewhat restrictive assumptions on the information sets of workers. In particular, the model requires that the test scores in the data are either not available to the workers or not fully signaled by the workers. Section 4 presents the numerical analysis. Section 5 concludes.

2 A Model of Employer Learning

The objective of this section is to model the educational decisions under asymmetric information combined with a tractable employer learning process. It rests on the same two key assumptions that the employer learning literature, mentioned above, is predicated upon. First the learning process is symmetric across different employers in the market, i.e., a worker’s output is observed by his potential employers as well as his current employer. This allows for a competitive labor market at each experience level where employers have the same information set. Second, the skill characteristics of the workers are assumed to be time invariant. This assumption is binding especially if the acquisition of additional human capital on the job is dependent on educational attainment or predetermined worker characteristics such as ability. Identifying the employer learning effects separately from the training effects remains an open question to be explored.

The next section presents a single sector economy and demonstrates the (in)distinction between the sorting generated by the signaling and the human capital models. If costs of or marginal returns to education vary systematically with ability, both models generate a sorting of the workers with high ability into higher education groups. If ability is not controlled for, this sorting is reflected in the estimated return to education as ability bias. The literature has focused mainly on purging the estimates of the ability bias and identifying the human capital role of education. Identification of the private return to education, that is the social return plus the return to signaling, requires sepa-

---

ration of the ability bias into its signaling and non-signaling components. This problem has attracted much less attention because of the inherent difficulties associated with a particular condition for a separating equilibrium: the market’s predictions about a worker’s true productivity are accurate on the average.

In the benchmark model the only screening device is education. Under this assumption, it is shown that the return to signaling is decreasing in the speed of learning. The weaker incentive to signal in turn generates a lower educational achievement at each ability level and a smaller education-ability gradient. In a special case where workers learn their productivity along with their employers, return to signaling is zero, and educational choices are independent of the speed of learning. Extending the model to allow for additional screening blurs the link from the speed of employer learning to the size of signaling. In an extreme case where all the uncertainty is uncovered prior to employment, education choices would not depend on the speed of learning at all. This is discussed in the subsection 2.2.

Section 2.3 extends the model to a two-sector economy where sectors differ in the signaling role of education and potentially in the productivity of human capital. The sorting behavior implied by endogenous sectoral choice is analyzed. Workers with higher ability choose to work in the sectors with higher return to human capital and lower role for signaling.

2.1 A Single Sector model of Asymmetric Information with Employer Learning

The model economy consists of a continuum of infinitely-lived workers that are heterogeneous in their skills. The marginal product, $y_i$, of a worker with educational attainment $s_i$ and ability $x_i$ is given by

$$\ln y_{it} = q(s_i, x_i) + h(t) + \varepsilon_i \quad \varepsilon_i \sim i.i.d. N(0, \sigma^2_\varepsilon).$$

where $h(t)$ denotes the experience profile of wages. Following the literature, the experience profile is assumed to be a deterministic process independent of educational attainment and ability. The marginal productivity is increasing and concave in both education and ability, $(q_s, q_x \geq 0, q_{ss}, q_{xx} \leq 0)$. In addition, marginal return to education is strictly increasing in ability, $q_{sx} > 0$. This is the basis for the sorting in the model\(^\text{10}\). The random component $\varepsilon_i$ is independent of $x_i$ and $s_i$, and is unknown to the worker and the employer. Ability $x_i$ is not directly observable by the employer, however is known by the worker.

After the worker starts his career, employers observe a noisy measure of each worker’s output every period.

\(^{10}\)Alternatively one could also consider a model where the marginal cost of education is decreasing in ability. For the theoretical analysis presented here the choice is not important. I discuss the empirical implications of this assumption in the results section.
\[ \eta_{it} = \ln(y_i + u_{it}) \quad u_{it} \sim i.i.d. N(0, \sigma_u^2) \] (1)

Let \( I_{it} = \{\eta_{i1}, \eta_{i2}, ..., \eta_{it}\} \) denote the history of worker \( i \)'s observed output up to experience level \( t \). Assume also that \( I_{i0} = \emptyset \). These output signals are assumed to be observed by the entire market and the employers are competitive in their hiring decisions.

Workers maximize the present discounted value of lifetime income. They choose their schooling levels based on the expected wage stream, conditional on the level of education they choose and their ability. The only cost of education is the foregone earnings during school years\(^{11}\). Let \( W_{it}(s_i) \) be the wage offer to a worker with \( s \) years of education and \( t \) years of experience. The worker’s optimization problem is given below.

\[
\max_s U(s, x) = e^{-rs} \int_0^\infty e^{-rt} E[W_{it}(s)|s_i, x_i]dt
\] (2)

A signaling equilibrium in this simple setting can be defined as follows.

**Definition 1** (Riley 1979a) A signaling equilibrium is a wage function \( W_{it}(s) \) and a policy function \( S(x) \) such that,

1. For all \( x_i \), \( S(x_i) \) solves (2) given \( W_{it}(s) \)
2. Individual choices are consistent with the wage function,

\[ W_{it}(s) = E[y_i|s_i, I_{it}] \]

The second condition follows from the assumption that the labor market is competitive. Existence of an equilibrium wage function with informational consistency is discussed in detail in Riley (1979a). There are essentially two requirements. First, in the economy without asymmetric information, where productivity can be observed directly, there is a unique education level that solves (2) for each ability level \( x \). Second, the opportunity cost of attaining higher educational levels is lower for workers with higher ability. For the model described above, the concavity of \( q(\cdot, \cdot) \) in \( s \) ensures a unique optimal education level when productivity is observed directly. The second requirement is met by the assumption that the marginal return to education is increasing in ability, \( q_{sx} > 0 \). These two conditions ensure the existence of an equilibrium where workers with higher \( x \) choose higher educational credentials.

I now turn to the characterization of the equilibrium. Denote the ability level inferred by the market given an education level \( s \) by the function \( \chi(s) \). Upon observing an output signal \( \eta_{it} \), the employer can extract noisy information on \( \varepsilon_i \) by

\[ \eta_{it} - q(s_i, \chi(s_i)) = q(s, x_i) - q(s, \chi(s_i)) + \varepsilon_i + u_i. \]

Note that these are unbiased observations around the true \( \varepsilon_i \) if and only if the inferred ability \( \chi(s_i) \) is equal to the true

\(^{11}\)Monetary cost of education is immaterial to the theoretical analysis and is ignored for tractability purposes.
ability $x_i$, for instance at the equilibrium. Then the wage offered to a worker with education level $s_i$ and history of signals $I_{it}$ is

\[
\ln W_{it}(s_i) = \ln E[y_{it}|s_i, I_{it}]
\]

\[
= q(s_i, \chi(s_i)) + \ln E[\exp(\varepsilon_i)|I_{it}] + h(t)
\]

\[
= q(s_i, \chi(s_i)) + \lambda(t) (\bar{\eta}_{it} - q(s, \chi(s))) + 0.5(1 - \lambda(t))\sigma^2_{\varepsilon} + h(t)
\]

\[
= (1 - \lambda(t)) q(s, \chi(s)) + \lambda(t)\bar{\eta}_{it} + 0.5(1 - \lambda(t))\sigma^2_{\varepsilon} + h(t).
\]  

(3)

where

\[
\lambda(t) = \frac{\sigma^2_{\varepsilon}}{\sigma^2_{\varepsilon} + \sigma^2_{u}/t}
\]  

and  

\[
\bar{\eta}_{it} = \sum_{k=1}^{t} \eta_{ik}/t
\]

The third line follows from the normality assumption on the distribution of $\varepsilon_i$ and $u_i$. Wages are a weighted average of the productivity as perceived by the market conditional on education alone, and the average of all observed outputs realized up to experience level $t$. Note that $\lambda(t)$ is increasing in $t$, $\lambda(0) = 0$, and $\lambda(\infty) = 1$. As firms learn about the true productivity of a worker, wages rely less on the productivity implied by education alone, and more on the observed output.

Since workers know their ability, they maximize the expected lifetime income, conditional on their ability as well as the level of education they attain. Using equation (3), expected wage stream at the beginning of a worker’s career is

\[
\ln E_0[W_t(s)|s, x] = q(s, \chi(s))(1 - \lambda(t)) + q(s, x)\lambda(t) + h(t) + \frac{\sigma^2_{\varepsilon}}{2}.
\]  

(4)

Expected wage at experience $t$ is a geometric average of the true productivity of a worker and his inferred productivity prior to entry. Plugging equation (4) into the maximization problem, and differentiating with respect to $s$, we have

\[
\Omega (q_s(s, \chi(s)) + q_x(s, \chi(s))\chi'(s)) + (1 - \Omega)q_s(s, x) = r,
\]

where

\[
\Omega = \frac{\int_0^\infty e^{-rt+h(t)}(1 - \lambda(t))dt}{\int_0^\infty e^{-rt+h(t)}dt}.
\]  

(5)

At a separating equilibrium the market extracts the correct information about the workers’ ability, implying $\chi(S(x)) = x$ for all $x$. Using this equality, the first order condition simplifies to

\[
q_s(s, x) + q_x(s, x)\chi'(s)\Omega = r.
\]  

Workers equate the marginal return to attaining one more year of education to the discount rate. The first term on the left of equation (6) is the rise in the productivity due to the higher human capital. The second term is the return to signaling. A marginal increase in the educational attainment of a worker increases the ability inferred by the market by $\chi'(s)$. The contribution of this increment to the wage growth is priced at
$q_x(s, x)$ by the market. Since the market learns the true productivity over time, the return to signaling is limited by $\Omega \in (0, 1)$. To see this, first note that the speed of learning in the model is determined by the noise in the observed output. If $\sigma_u^2$ increases, learning slows down and $\lambda(t)$ is lower for all $t$. This generates a higher $\Omega$, and increases the marginal return to education, and hence the optimal schooling choice for each ability level.

It is important to note that even though at the equilibrium employer learning occurs over $\varepsilon$, a variable that is not related to education or ability, the educational choices are affected by how fast the employers learn. This is because the employers can extract information about $\varepsilon$ only to the extent that their predictions about ability are accurate. This gives the worker the incentive to delude the employer by attaining higher education, but of course, the employers’ predictions are correct at the equilibrium.

$\Omega$ combines the two theories in an intuitive way. At the extreme situation where output is not observed, (or when $\sigma_u^2 \to \infty$), $\Omega$ equals 1, and the return to signaling is maximized. On the other extreme, if the output is observed without any noise, then $\Omega$ would equal zero, and the return to signaling would be zero. In this case, the optimal education choice solves, $q_x(s, x) = r$. This is a purely human capital model and the educational choices are independent of the learning process. Educational choices when $\Omega$ is 0 are socially efficient since each worker’s return to education is the same as the social return. Denote the efficient choice function by $S^*(x)$.

Equation (6) defines a differential equation in $s$, the solution to which describes the equilibrium strategy for the employers. Without an initial condition, the solution to this problem is a continuum of functions that differ in their intercepts. The following lemma establishes that the worker with the minimum ability attains the efficient level of education, pinning down a unique equilibrium strategy.

**Lemma 1** At a separating equilibrium, $S(0) = S^*(0)$

**Proof.** $S(0) < S^*(0)$ clearly can not be the case since the agent can still signal his ability, and improve his lifetime income by receiving $S^*(0)$. Suppose for a contradiction that $S(0) > S^*(0)$. At a separating equilibrium we must have $\chi(S(0)) = 0$. Consider a deviation to $S^*(0)$, for which $\chi(S^*(0)) = 0$ at the equilibrium since it is still the minimum of all education choices. But then, $V(S(0), 0) = e^{-rS(0)} \int_0^\infty e^{-rt} \exp q(S(0), 0)dt < e^{-rS^*(0)} \int_0^\infty e^{-rt} \exp q(S^*(0), 0)dt = V(S^*(0), 0)$ by definition of $S^*(0)$, which contradicts with optimality of $S(0)$ in the signaling environment. ■

We can now characterize the signaling equilibrium with employer learning.

**Proposition 1** The fully separating equilibrium is characterized by functions, $S(x)$ and $\chi(s)$, such that

1. $\chi(s)$ solves the differential equation in (6) with the initial condition $\chi(S^*(0)) = 0$.
2. Schooling choices are consistent with the market’s predictions.

$$S^{-1}(x) = \chi(s)$$
3. Given random components \( \varepsilon_i \) and \( \{ u_i \}_{i=1}^{\infty} \), worker \( i \)'s wage process is given by,

\[
\ln W_{it} = q(s_i, x_i) + \lambda(t)(\varepsilon_i + \bar{u}_it) + h(t) + (1 - \lambda(t))\sigma^2 \varepsilon
\]

where \( \bar{u}_t = \sum_{k=1}^{t} u_{ik}/t \)

The last part of the proposition can be obtained by setting \( \chi(s_i) = x_i \) and using equation (1) in the wage equation.

The total return to education estimated by an econometrician who does not observe \( x_i \) captures the total derivative of wages with respect to education. This is given by,

\[
\frac{d\ln w}{ds} = \underbrace{\frac{\partial \ln q(s, x)}{\partial s}}_{\text{True Return}} + \Omega \underbrace{\frac{\partial \ln q(s, x)}{\partial x} \chi'(s)}_{\text{Ability Bias}} + (1 - \Omega) \frac{\partial \ln q(s, x)}{\partial x} \frac{dx}{ds}. \tag{7}
\]

The first term in the expression above measures the percentage increase in human capital due to education. This is the social return, or the true return to education. Uncovering this component has been an important objective in the literature on the return to schooling. The second term is the return to signaling. The first two components constitute the worker’s private return to education, which is equated to the discount rate at the optimum. The last term is a pure selection bias\(^{12}\) and is not a part of any form of return. The last two terms are known as ability bias in general.

The major difficulty in gauging the return to signaling using wage regressions is outlined by the equation above. If we have access to instruments that are orthogonal to ability and predict the educational choices, then the estimate can be purged of the ability bias and the social return to education can be estimated. But, this method would not separate the signaling bias from the selection bias that would occur in a human capital model. This is because the employers predictions about a worker’s ability are correct at the equilibrium, implying \( \chi'(s) = dx/ds \).

This does not imply however that the total ability bias is independent of \( \Omega \). The education-ability gradient, \( ds/dx \), is a direct function of the size of signaling represented by \( \Omega \) in the model. The ability bias is actually decreasing in the extent of signaling in a wide class of functional forms of \( q(s, x) \) that are used in the literature\(^{13}\). The following examples demonstrate the results discussed above with an analytical form for \( q(s, x) \) that is later used in the numerical analysis.

**Example.** (A Human Capital Model with Symmetric Uncertainty)
Assume that the human capital production function is given by \( \ln q(s, x) = As^\beta x^\alpha \), where \( \beta < 1 \). In addition, assume that the firms observe \( x \) directly prior to employment. In this case the learning process is symmetric between the workers and employers, since

\(^{12}\)Technically, there’s self-selection in both models. To distinguish the two I’ll refer to the selection bias in the signaling model as signaling bias.

\(^{13}\)See Lang (1984)
\( \varepsilon_i \) is not observed by either. The first order condition to the worker’s maximization problem in this case is

\[
\beta A s^{\beta - 1} x^\alpha = r.
\]

An important aspect of the condition above is that it does not rely on the learning parameters. If the worker and employer are learning simultaneously, educational choices are not affected by the speed of learning. Solving for \( s \) yields the following relationship between ability and education at the optimum.

\[
S(x_i) = \left( \frac{r}{A \beta} \right)^{-\frac{1}{1-\beta}} x_i^{\frac{\alpha}{1-\beta}}.
\]

This is a standard human capital model with selection. Since workers with higher ability enjoy higher marginal return to education, schooling is increasing in ability.

Total estimated return to education unconditional on ability can be decomposed as follows.

\[
\frac{d \ln W}{ds} = r + \frac{1 - \beta}{\beta} r = \frac{r}{\beta}
\]

Since there is no asymmetric information, the private return and the social return to education are the same and are equal to the discount rate. However, since ability is positively related to education, the estimated return to education exaggerates this return by a factor of \( \frac{1 - \beta}{\beta} \). Note that even though the underlying production function is highly nonlinear and displays diminishing marginal returns to education, total return to education is linear in schooling when ability is not controlled for. This is because diminishing marginal return is exactly offset by the sorting of higher ability workers into higher educational levels. This has been pointed out in the literature as a justification for Mincerian type regressions, which assume log-wages are linear in education (See Card (2001) for instance).

**Example.** (*A Signaling Model with Employer Learning*)

Assume that the production function is as in the previous example. If firms do not directly observe ability, then the first order condition is

\[
\beta A s^{\beta - 1} x^\alpha + \Omega \alpha A s^{\beta} x^\alpha x^{-1} \chi'(s) = r.
\]

Solving the differential equation above with the initial condition \( \chi(0) = 0 \), optimal education choice is

\[
S(x) = \left[ \frac{r/A}{\beta + \Omega (1 - \beta)} \right]^{-\frac{1}{1-\beta}} x^{\frac{\alpha}{(1-\beta)}}.
\]

Note that as \( \Omega \) approaches 0, employers observe the true productivity directly, and the model reduces to a selection model without asymmetric information. As \( \Omega \) approaches 1, the model is a pure signaling model where employers never learn the true productivity of a worker. Compared to the previous example, \( S(x) \) is higher and steeper at all ability levels provided that \( \Omega > 0 \).
The estimated return to education is

\[
\frac{d\ln W}{ds} = \frac{r\beta}{\beta + \Omega(1 - \beta)} + \frac{r\Omega(1 - \beta)}{\beta + \Omega(1 - \beta)} + \frac{r(1 - \Omega)(1 - \beta)}{\beta + \Omega(1 - \beta)} = \frac{r}{\beta + \Omega(1 - \beta)}.
\]

The first term is the rise in the human capital. This is less than in the previous example, because with asymmetric information, workers attain higher educational levels and the production function exhibits diminishing marginal returns. The second term is the return to signaling. The first two terms add up to \( r \), which is the worker’s private return to education. The last term is the selection bias. Note that if \( \Omega = 1 \), the model is a signaling model without learning, and the total estimated return is \( r \). Although this is the private return to education, it is still higher than the social return. Thus, provided that \( \Omega \in (0, 1) \), total estimated return to education is higher than both the social return and the private return.

Slower employer learning, or equivalently, a higher \( \Omega \), reduces the total estimated return to education for two reasons. First, since workers are choosing higher educational levels, the marginal contribution of education to human capital diminishes. Second, since workers get more education conditional on their ability, the level of ability inferred by employers for a given level of education is lower with a higher \( \Omega \). This also results in a decline in the ability bias term. Note that even though the return to signaling is increasing in \( \Omega \), the sum of signaling and selection biases are decreasing.

### 2.2 Initial Knowledge and Learning Content

The model so far assumes that at the beginning of a worker’s career, firms do not have any information on his ability but that signaled by educational attainment. However, firms might extract information on the true productivity of a worker via cognitive tests, interviews or reference letters prior to hiring. Recent models in the employer learning literature allow firms to have partial information about a worker’s productivity before his entry into the market, and they assume that the econometrician does not have access to this information. The model above can be extended to allow for such information in the following way. Suppose that before workers start their career, employers can observe exactly the true productivity of the worker with probability \( \rho \). This probability is known to the workers. Hence, with probability \((1 - \rho)\), wages will only be learned over time. Expected wages conditional on the worker’s information set is a weighted average of the true productivity of a worker and the productivity that is inferred by education and the observed output signals.

\[
E_0[W_t(s)|s, x] = \left(\rho e^{q(s, x)} + (1 - \rho) e^{q(s, \chi(s))(1 - \lambda(t)) + q(s, x) \lambda(t)}\right) e^{\sigma^2/2}
\]

The weight \( \rho \) represents the fraction of variation in wages that can be extracted through the information that is available to the firms before the worker is hired. The first order condition in this environment is,

\[
q_s(s, x) + q_x(s, x) \chi'(s) \Omega(1 - \rho) = r
\]
The introduction of the possibility of initial knowledge reduces the marginal return to signaling. Optimal choices for education in the example above can be modified for the introduction of initial knowledge by simply replacing $\Omega$ with $\Omega(1 - \rho)$. $(1 - \rho)$ measures the fraction of the uncertainty that can not be ascertained by additional screening devices. This is referred to as the (average) content of learning. If $\rho$ is equal to 1, then all of the uncertainty is cleared before entry, and there’s nothing to learn. In this case, the speed of learning does not matter for educational choices at all. This brings us back to the human capital model.

It is clear that the size of signaling crucially depends on the content of learning as well as the speed of learning. If the content of learning is similar across markets with different speeds of learning, then we should see less room for signaling in the markets with faster employer learning, and hence lower educational attainment controlling for ability. However, an endogenous link from the speed of learning the content of learning is plausible. In markets where it is difficult to observe the marginal value product of a worker, for instance due to teamwork, employers may find it profitable to invest in screening technologies other than education. This generates an offsetting effect to the signaling contribution of the speed of employer learning.

2.3 Endogenous Occupational Choice

As argued above, in order to analyze the implications discussed above, we need more than one market. The literature has focused on self employed workers, for which signaling presumably does not play a significant role in education decision compared to salaried workers (Wolpin 1977). This categorization assumes that agents decide to be self-employed before they complete schooling. It also assumes that the potential uncertainty about the value of services provided by self-employed workers is less than that of salaried workers.

Using a signaling model, Riley (1979b) categorizes the market into screened and non-screened sectors. He assumes that these sectors are not separated with respect to ability, and the sector where signaling is important also has a higher true return to education. These two assumptions imply that the workers who choose the screened sector attain higher education. Since they are assumed to be indifferent across sectors, they must also be paid less controlling for education. The screened group is then defined as the set of occupational categories with high average education and low occupational intercept in a wage regression. Both assumptions in this study are quite restrictive as will be explained in this section.

Allowing for differences in productivity and the extent of signaling does indeed separate workers with respect to their ability in a systematic way. It will be shown that workers with higher ability sort into sectors where human capital is more productive and the signaling role of education is limited.

The tests based on the separation of markets necessarily rely on the assumption that workers are not mobile across these markets. This is a restrictive assumption, yet
one that has to be maintained. In its essence, if educational choices are affected by the extent of signaling in different markets, then these choices must come after making a commitment to a certain market. This can be partially justified for workers with more than high school education on the grounds that the additional years of schooling usually target a subset of the market. Major choice in college and choice of the type of vocational school are common examples of such commitment.

In what follows I describe the aggregate economy. Without loss of generality, I assume that the only screening device is education, i.e. \( \rho = 0 \). This implies that the size of signaling is determined by the speed of employer learning alone. I first analyze the sorting behavior assuming that human capital is equally productive in all markets. Then I relax this assumption and analyze the general sorting behavior.

*The Aggregate Economy*

Consider an economy with two sectors denoted by \( j \in 1,2 \). Sectoral production is linear in the total stock of human capital in that sector, denoted by \( h_j \). The total output is determined by a Cobb - Douglas form aggregator.

\[
Q = Zh_1^\phi h_2^{1-\phi}
\]

The price of human capital per efficiency unit is given as follows.

\[
\omega_j = \frac{\partial Q}{\partial h_j}
\]

A worker’s earnings are determined by the total expected human capital he governs, multiplied by the price per efficiency unit,

\[
W_{ijt} = \omega_j E[y_i|s_i, I_t]
\]

* Differences in the Speed of Employer Learning *

Suppose that marginal product is given by \( \ln q(s, x) = As^\beta x^\alpha \). Human capital is equally productive in both sectors, \( A_1 = A_2 \), but sectors may differ in their learning speed. In particular, assume that in sector 2, the observed output, \( \eta_{it} \), is more precise relative to sector 1. This implies that \( \Omega_2 < \Omega_1 \). Hence, the return to signaling is lower in sector 2. The worker’s problem is to maximize the expected lifetime income given a wage function as before. Since \( \omega_j \) multiplies his total stock of human capital, the only parameter that will generate differences in educational achievements is \( \Omega \). Agents first choose an occupation and then decide how much education to attain. Given a sectoral choice \( j \), the value function for a person with ability \( x_i \) can be written as follows.

\[
\ln V_j(x_i) = \ln \omega_j + q(S_j(x_i), x_i) - rS_j(x_i) + \sigma_\varepsilon^2/2.
\] (8)

Using the assumed functional form for \( q(s, x) \), this can be expressed as

\[
\ln V_j(x_i) = \ln \omega_j + \kappa_j x_i^{\alpha - \beta} + \sigma_\varepsilon^2/2,
\]
where

$$\kappa_j = \left( \frac{r}{\beta + \Omega_j(1 - \beta)} - r \right) \left( \frac{r/A}{\beta + \Omega_j(1 - \beta)} \right)^{-1/(1-\beta)}.$$

The value function is decreasing in $\Omega_j$. A higher $\Omega$ causes workers to attain higher education. The marginal return to education is decreasing at higher education levels but the marginal cost, $r$, is constant. This causes the value function to decrease.

Let $\tau_i$ be the worker’s personal taste for occupation 1. $\tau_i$ is independent of $x_i$, and is distributed according to c.d.f. $G(\tau_i)$ on $(-\infty, +\infty)$. Without loss of generality, assume that $\Omega_1 > \Omega_2$. Under this setting, worker $i$, chooses sector 1 over 2 if and only if

$$\ln V_1(x_i) + \tau_i > \ln V_2(x_i)$$

$$\Leftrightarrow$$

$$\ln (\omega_1/\omega_2) + \tau_i + (\kappa_1 - \kappa_2)x^{1/(1-\beta)} > 0.$$

For a given factor price ratio, $\omega_1/\omega_2$, and taste parameter $\tau_i$, the difference between the two value functions is strictly decreasing in $x$. This is because $\kappa$ is strictly decreasing in $\Omega$. This suggests that given factor prices, for each $\tau_i$, there exists a critical ability threshold $\tilde{x}(\tau_i) \in [0, \infty)$, such that for all $x < \tilde{x}(\tau_i)$, sector 1 is optimal. The critical level of ability given $\tau$ can be obtained from the equation above.

$$\tilde{x}(\tau_i) = \left( -\frac{\ln (\omega_1/\omega_2) - \tau_i}{\kappa_1 - \kappa_2} \right)^{1/(1-\beta)}.$$

Note that $\tilde{x}(\tau_i)$ is non-decreasing in both $\tau_i$ and $\ln (\omega_1/\omega_2)$, and strictly increasing if $\tilde{x}(\tau_i)$ is in the interior. The total measure of human capital allocated to each sector is

$$h_1 = \int_{-\infty}^{\tilde{x}(\tau_i)} \int_{0}^{\tilde{x}(\tau_i)} \exp \left( q(S_1(x), x) + \sigma^2_\varepsilon/2 \right) dF(x) dG(\tau)$$

and

$$h_2 = \int_{-\infty}^{\tilde{x}(\tau_i)} \int_{\tilde{x}(\tau_i)}^{\infty} \exp \left( q(S_2(x), x) + \sigma^2_\varepsilon/2 \right) dF(x) dG(\tau).$$

The factor allocations must be consistent with the factor prices at the equilibrium. This requirement implies

$$\ln (\omega_1/\omega_2) = \ln (\phi/(1 - \phi)) + \ln h_2/h_1.$$

The left hand side of this equation is increasing in the factor price ratio, $\omega_1/\omega_2$, and the right hand side is decreasing. Moreover, as the factor price ratio approaches 0, more workers sort into sector 2, and the RHS approaches $\infty$. Similarly, as $\omega_1/\omega_2 \to \infty$, the RHS approaches $-\infty$. Therefore, by continuity, there exists a unique $\omega_1/\omega_2$ that provides a factor allocation consistent with the factor prices.

---

14See appendix for the derivation.
At the equilibrium, workers with higher ability are more likely to sort into the sector with faster employer learning. Figure 1 displays the equilibrium sorting behavior. Low ability workers choose the occupation with more room for signaling, unless they have a strong taste for the occupation where signaling is limited by employer learning.

Educational choices conditional on ability are depicted in Figure 2. Since slower learning implies larger signaling, workers in this sector attain higher education at each ability level. Furthermore, the education - ability gradient is larger. The educational choices depicted here crucially rely on the assumption that the distribution of occupational taste has an unbounded domain. If the magnitude of $\tau$ is restricted to lie in a bounded interval, then this might disrupt the optimal educational choices. In particular, if $\tau$ is bounded below by $\tau_b$, then there is a critical ability level $\bar{x}(\tau_b)$ associated with this bound such that all workers below this threshold choose sector 1. If this threshold is known by both the market and workers, then a worker with ability $\bar{x}(\tau_b)$ chooses to attain the efficient level of education since he would be the worker with minimum ability who has chosen sector 2. This is a credible deviation if employers know the value of $\bar{x}(\tau_b)$. Symmetrically, an upper bound on $\tau$ implies the existence of $\bar{x}(\tau_b)$ above which workers do not choose sector 1. This situation is depicted in Figure 3. Although the unboundedness condition is crucial, the standard deviation of $\tau$ is not constrained. One may consider $\tau$ as a small perturbation, not in its usual sense, but in the sense that it can have a small (but non-zero) variance.

**Differences in the Productivity of Human Capital**

Suppose now that sectors have the same speed of employer learning, but they differ in the productivity of human capital, $A_j$. Note that $A_j$ and $\Omega_j$ enter the value function in the same way, through $\kappa_j$ only. Since $\kappa_j$ is increasing in $A_j$, with equal speeds of learning, an analogous derivation shows that workers with higher ability choose to sort into the sectors where human capital is more productive.

In an environment where both the speed of learning and the productivity of human capital vary across sectors, sectoral allocation of human capital is more complicated. Clearly, if one of the sectors rewards human capital more and also offers less room for signaling, then both effects would sort higher ability workers into this sector. However if a sector rewards human capital more, but the signaling role of education is important, then the nature of the sorting depends on the relative gains from higher productivity in comparison to the efficiency losses generated by larger signaling. This suggests that if the relative productivity of human capital is considerably higher in one sector, then at the equilibrium high ability workers could sort into this sector despite the larger size of signaling.
3 Empirical Implications and the Employer Learning Literature

The employer learning model as presented by Farber and Gibbons (1996) predicts that a measure of ability that is not directly observable to employers at the beginning of a worker’s career becomes increasingly correlated with wages if the market learns the true productivity type of a worker over his experience. Using a similar model, Altonji and Pierret (2001) show that if the measure of ability is also correlated with education, then the relation between wages and education would weaken as employers rely less on education and more on the observed output to determine wages. This argument can be tested by regressing wages on interactions of education and the measure of ability with experience. The presence of employer learning indicates that the coefficient on education declines as experience increases, and the coefficient on the measure of ability rises. The model presented in the previous section does not generate these patterns in the data, unless one is willing to impose restrictive assumptions on the worker’s information set. To see this, suppose that productivity is linear in education and ability,

$$\ln y_{it} = q(s_i, x_i) + \varepsilon_i = \beta s_i + \alpha x_i + \varepsilon_i$$

Assume also, for simplicity, that firms do not have additional information besides education before the worker is employed, i.e. $\rho = 0$. The model predicts that at the equilibrium wages are given by,

$$\ln w_{it} = \beta s_i + \alpha x_i + E[\varepsilon_i | I_t] + h(t)$$

The ability of the model to generate a decreasing coefficient on education and an increasing coefficient in ability depends on how one interprets the observed measure of ability. If the econometrician has a proxy for $x$, then the perfectly separating equilibrium feature of the model prevents these coefficients from changing over time. In the end, $x$ is perfectly known when the worker starts his career. To produce changing coefficients over experience, econometrician’s information set must contain some information on the variable that is learned over time, $\varepsilon_i$. In order to capture both possibilities, suppose that the econometrician has access to a measure of ability, $z_i$, that is potentially correlated with both the unobserved component, $\varepsilon_i$, and $x_i$.

$$z_i = ax_i + b\varepsilon_i + \nu_i$$

$z_i$ is not directly observable to the employers. Suppose, perhaps heroically, that this measure is either not observed or observed but not utilized by workers to the full extent in making educational choices.

In order to bring the model closer to the literature, assume that the workers’ equilibrium strategy is linear in ability.

$$s_i = \phi_0 + \phi_1 x_i$$
Consider the following projection of wages on education and ability with time varying coefficients

\[ E[\ln w_{it}|s_i, z_i] = \kappa_{st}s_i + \kappa_{zt}z_i \]

Using the multivariate regression formula, it is easy to show,

\[
\hat{\kappa}_{st} = \beta + \alpha/\phi_1 - a/\phi_1 \frac{b \sigma^2_{\nu}}{b^2 \sigma_x^2 + \sigma^2_{\nu}} \text{cov}(\varepsilon, E[\varepsilon|I_t])
\]

\[
\hat{\kappa}_{zt} = \frac{b \sigma^2_{\nu}}{b^2 \sigma_x^2 + \sigma^2_{\nu}} \text{cov}(\varepsilon, E[\varepsilon|I_t])
\]

The set of equations in (9) summarize the basic results in Altonji and Pierret (2001). As employers have more precise information about a worker’s true productivity, \(\text{cov}(\varepsilon, E[\varepsilon|I_t])\) will increase leading to a decrease in \(\hat{\kappa}_{st}\) and an increase in \(\hat{\kappa}_{zt}\). The revelation of information about true productivity causes wages to reflect the variation in true productivities, and \(z\) captures this variation, raising \(\hat{\kappa}_{zt}\) in our regressions. If \(z\) is also related to educational choices, via its relation to \(x\) in this case, then the coefficient on education will decline.

Note that if \(z\) is a measure that is uncorrelated with \(x\), \(a = 0\), then the coefficient on education will be constant. This is the case analyzed in Farber and Gibbons (1996). They extract such a variable by regressing the measure of ability on education first and collecting the residuals from this regression. By construction, the residuals are orthogonal to education, and hence \(x\), but are correlated with the \(\varepsilon_i\). The results in this study confirms a constant estimated return to education over experience and increasing coefficient on the constructed ability residuals. If \(b = 0\), then the ability measure \(z\) is not related to \(\varepsilon\), and both coefficients would be constant over a worker’s experience. If the measure of ability is exact, i.e. \(z_i = \alpha x_i + \epsilon_i\), then the coefficient on education would converge to the productivity augmenting effect of education in this simple version of the model\textsuperscript{15}.

It is crucial that this result is independent of the presence of asymmetric information in the model. Even when the market observes ability \(x_i\) directly, wage regressions will display the same characteristics as long as there is some component of productivity that is not observable to the employer at the time a worker is hired, and employers obtain information about this component later. This is why Altonji and Pierret (1997) and more recently Lange (2005) interpret their results as an upper bound on the contribution of the signaling motive to wages.

**Measuring the Speed of Learning**

The empirical patterns discussed above can be used to calculate a measure of learning in our model. By imposing the normality assumptions on \(\varepsilon_i\) and observed output

\textsuperscript{15}We have abstracted from the unobservable components of productivity that may be correlated with education even after controlling for \(z_i\). The existence of such variates would prevent measuring the true return to education by analyzing workers with high experience levels.
η_{it}, one can estimate the speed at which employers learn about the true productivity of their workers. Lange (2005) proposes an estimate based on the set of equations in (9). I will first go over the estimation procedure in his paper, then propose an alternative procedure based on the empirical patterns in Farber and Gibbons (1996).

Assuming that ε_{i} and observed output η_{it} are normally distributed, the estimated coefficients in equation (9) can be expressed as

\begin{align*}
\hat{\kappa}_{st} &= \kappa_{s0}(1 - \lambda(t)) + \kappa_{s\infty}\lambda(t) \\
\hat{\kappa}_{zt} &= \kappa_{z0}(1 - \lambda(t)) + \kappa_{z\infty}\lambda(t)
\end{align*}

where

\begin{align*}
\kappa_{s0} &= (\beta + \alpha/\phi_1) + \rho(\beta + \alpha/\phi_1 - a/\phi_1) \frac{b\sigma^2_{\varepsilon}}{b^2\sigma^2_{\varepsilon} + \sigma^2_{\nu}} \\
\kappa_{s\infty} &= (\beta + \alpha/\phi_1 - a/\phi_1) \frac{b\sigma^2_{\varepsilon}}{b^2\sigma^2_{\varepsilon} + \sigma^2_{\nu}} \\
\kappa_{z0} &= \frac{b\sigma^2_{\varepsilon}}{b^2\sigma^2_{\varepsilon} + \sigma^2_{\nu}} \rho \\
\kappa_{z\infty} &= \frac{b\sigma^2_{\varepsilon}}{b^2\sigma^2_{\varepsilon} + \sigma^2_{\nu}}
\end{align*}

(10)

λ_{t} can be expressed as a single parameter process defined as,

\[ \lambda_{t} = \frac{tK}{1 + (t - 1)K} \]

with \( K = \frac{\sigma^2_{\varepsilon}}{\sigma^2_{\varepsilon} + \sigma^2_{\nu}} \). If we have access to variables \( z_{i} \) and \( s_{i} \) as well as wages up to an experience level T, then parameters \( \{\kappa_{s0}, \kappa_{s\infty}, \kappa_{z0}, \kappa_{z\infty}, K\} \) can be estimated using coefficients on education and the measure of ability over experience, \( \{\hat{\kappa}_{st}, \hat{\kappa}_{zt}\}_{t=0}^{T} \).

Note that equation (10) suggests that an estimate of ρ can be obtained by \( \kappa_{z0}/\kappa_{z\infty} \). This is because the learning content is modeled in a particular way that implicitly assumes that the measure of ability \( z \) captures all the initial information that is available to the employer at the beginning of a worker’s career. If \( z \) is only partially correlated with this information, then the actual ρ is greater than the one predicted by \( \kappa_{z0}/\kappa_{z\infty} \).

An alternative estimate of the speed of learning is possible under the perfectly separating equilibrium. I begin by relaxing the assumption that education is linear in ability. Since the relation between ability and education is perfect in our model, it is possible to identify the part of ability measure \( z \) that is not signaled (or not perfectly observed by the employer). Then at the equilibrium \( z_{i} = a\chi(s_{i}) + b\varepsilon_{i} + \nu_{i} \) where \( \chi(s) = S^{-1}(x) \) is the equilibrium market strategy. The residuals, \( \tilde{z} \), obtained from a regression of \( z \) on a general function of education that approximates \( \chi(s) \), would decompose \( z \) into two: a component that is transmitted to the employer and a residual component that is unraveled over experience.
Consider the following regression of wages on education and $\tilde{z}$ interacted with experience.

$$E[\ln w_{it} | s_i, z_i] = \kappa_{st}s_i + \kappa_{\tilde{z}t}\tilde{z}_i$$

Using the multivariate regression formula once again, $\kappa_{\tilde{z}t}$ are

$$\hat{\kappa}_{\tilde{z}t} = \frac{\text{cov}(E[\tilde{z}_i | I_{it}], \varepsilon)}{b^2 \sigma^2_{\varepsilon} + \sigma^2_{\nu}}$$

$$= \kappa_{\tilde{z}0}(1 - \lambda(t)) + \lambda(t)\kappa_{\tilde{z}\infty}$$ (11)

where

$$\kappa_{\tilde{z}0} = \frac{b}{b^2 \sigma^2_{\varepsilon} + \sigma^2_{\nu}}$$

$$\kappa_{\tilde{z}\infty} = \frac{b}{b^2 \sigma^2_{\varepsilon} + \sigma^2_{\nu}}$$ (12)

The coefficient on education is time invariant and equals the OLS return to education when ability is not controlled for\textsuperscript{16}. Then using the estimated coefficients $\hat{\kappa}_{\tilde{z}t}$, the speed of learning can equivalently be estimated as above. In particular, if we have access to variables $z_i$ and $s_i$ as well as wages up to an experience level $T$, then parameters $\{\kappa_{\tilde{z}0}, \kappa_{\tilde{z}\infty}, K\}$ can be estimated by a non-linear regression of the estimated coefficients $\{\hat{\kappa}_{\tilde{z}t}\}_{t=T}^T$ on experience using equation (11). Unlike the previous method, this method is robust to nonlinear forms of $q(.,.)$. This is the methodology employed in the numerical section to measure the speed of learning.

*An upper bound on signaling*

Measuring the speed of learning enables a quantitative bound to be imposed on the signaling bias. Once we have an estimate of the speed of learning, $K$, we can calculate $\Omega$ using equation (5). Rewrite the first order condition to the worker’s problem given in equation (6) as follows,

$$[q_s(s, x) + q_x(s, x)\chi'(s)] + (\Omega(1 - \rho) - 1)q_x(s, x) = r$$

The first term inside the brackets is the total return to education that contains the true productivity augmenting effect and the ability bias. This return can be estimated by regressing wages on educational achievement without controlling for ability. Let $\hat{b}_s$ denote the coefficient on education in this regression. Rearranging terms and solving for the return to signaling we get,

$$\Omega(1 - \rho)q_x(s, x)\chi'(s) = \frac{\Omega(1 - \rho)}{1 - \Omega(1 - \rho)}(\hat{b}_s - r) \leq \frac{\Omega}{1 - \Omega}(\hat{b}_s - r)$$ (13)

If the initial knowledge is zero, then the actual return to signaling is equal to the bound. To the extent that firms have extra information that is not available to the

\textsuperscript{16}One can show that $\text{plim} \ k_{st} = \beta + \alpha E_i[\chi'(s_i)]$, which does not depend on $t$. 

20
econometrician, the return to signaling is lower than the bound provided here. In the extreme case when \( \rho = 1 \), the actual return to signaling is zero. Another crucial aspect of this upper bound is that it depends on the marginal cost of education, which is the discount rate here.

Lange (2005) estimates the speed of learning using the first estimation strategy discussed above and derives a bound on the return to signaling. His bound builds on the observation that \( \hat{b}_s - q_s(s, x) \), the ability bias, is greater than the signaling bias. This is identical to the bound given in equation (13)\(^{17}\). He reports his results based on an estimate of tuition costs and various discount rates. The absence of monetary cost of education enables this bound to be represented in a simple form without the calculation of lifetime income, which inevitably necessitates assumptions on the wage levels for workers with higher experience levels than observed in his dataset\(^{18}\).

Table 1 calculates the bound presented above using the speed of learning and the estimated return to education reported in Lange (2005). I estimate the experience profile of wages using a wage regression with a quartic polynomial in experience. The last two columns of the table are taken from Table 3 of his paper. The upper bound on signaling implied by the speed of learning varied from zero to at most 15.8% of the estimated total return to education depending on the discount rate. Expressed as a percentage of the private return, this value can be up to 46% when the discount rate is 3%. These numbers are slightly higher than those presented in Lange (2005). The reason is that the marginal cost measure he uses to calculate the upper bound includes an estimate of the tuition costs. Here, the marginal cost of education is equal to the discount rate only. This results in a smaller denominator when the return to signaling is expressed as a percentage of the private return.

4 Numerical Implementation

A mapping from the potential return to signaling to educational achievement necessitates two markets that are different in their employer learning characteristics. The model presented above predicts that, accounting for differences in productivity of human capital, workers with higher ability are more likely to sort into markets where the return to signaling is lower. Building on this implication of the model, the labor market is categorized into two main groups based on the average AFQT score of workers in each occupation category.

The parameters of the model can be classified into three main groups. The production function parameters \( (\alpha_j, \beta_j, A_j) \), parameters related to employer learning \( (\sigma_{uj}, \sigma_s, \rho_j) \), and the environment parameters \( (r, \sigma_x, \mu_x, T, \phi) \). The numerical exercise presented here attempts to control for the productivity by estimating the production function

\(^{17}\)See Appendix for the derivation

\(^{18}\)Lange (2005) uses data from NLSY, which consists of fairly young and middle aged workers who are in the mid stages of their working lives.
controlling for ability using only the workers that are later in their career. The underlying assumption is that the employers have likely learned the true productivity so that estimated coefficients are close to their true values. The speed of employer learning is estimated as outlined above. For the educational decision the variable of interest is $K = \frac{\sigma^2}{\sigma^2 + \sigma^2_t}$. The environment variables except for the discount rate are calibrated to match certain moments in the data as discussed below. The remaining four variables ($r, \rho_1, \rho_2$ and $\sigma_\tau$) are estimated by the method of simulated moments. The moments that the simulations are asked to meet are, average educational attainment, the education-ability gradient and the average ability in each sector.

The procedure is carried out in two steps. First, the differences between the simulated moments and the data moments were minimized for several random draws of ability, $x$, and occupational taste $\tau$ for a population of 10,000. Each moment was weighted equally. Using the simulated values for the parameters, the variance-covariance matrix is constructed. Then, the first step is repeated weighting the moments by the inverse of the variance-covariance matrix.

Data are taken from the NLSY in order to take advantage of the AFQT score. The particular sample used in the estimations is the nationally representative cross-sectional sample. Female subjects are dropped. Additional sampling restrictions and the definition of variables used in the regressions are summarized in the appendix. The reported estimation statistics are unweighted statistics.

Table 2 shows the average AFQT score in the 12 single digit occupational classifications. Occupations with less than average AFQT score are referred to as the low AFQT group. Average educational attainment is 14.64 years in the high AFQT occupations, and 12.02 years in the low AFQT occupations. The model provides three main reasons that might explain this difference. First, the association between ability and education directly implies that the high AFQT occupations should attain higher education on the average. Second, workers with high AFQT are more likely to sort into occupations with higher return to human capital. Therefore, workers with higher AFQT may achieve higher education also because the return to education is higher in the occupations where they work. Third, the disparities in the signaling role of education across sectors might contribute to this difference. In particular, if signaling is limited in one group, then workers in this group should choose to attain lower education levels conditional on their ability.

I start with the estimation of the speed of learning. Using the perfectly separating feature of the model, the AFQT is regressed on education dummies. The residuals from this regression represent the component of the AFQT that can not be inferred by education alone. Then wages are regressed on a full set of experience dummies, education and the AFQT residual interacted with these dummies and racial controls. If the AFQT residuals contain information about productivity that is not available to the employer at the beginning of a worker’s career, the coefficient on the AFQT residual increases over experience. Figure 4 displays the estimated coefficients. Since these residuals (by construction) are not related to education, the coefficient on education...
does not decrease.

Using the estimated coefficients on the AFQT residual, the speed of learning is estimated by non-linear least squares applied to equation 11. Table 3 displays the estimation results for the speed of learning in the two occupational categories. Occupations with higher AFQT are subject to slower employer learning. Assuming that the learning content in the two occupations is similar, estimated speeds suggest that the high AFQT occupations are potentially subject to higher returns to signaling. However, the estimated initial values for the coefficient on the AFQT residual suggest that the assumption of similar learning contents is a strong one. The initial value for the AFQT residual in the low ability group is close to zero. This suggests that prior to employment, the market does not have any information on the productivity of these workers besides what can be inferred from their education levels. For occupations with high average AFQT, the initial value of the coefficient is estimated as 3.9%, which implies that, relative to the low AFQT occupations, employers have more information on a worker’s productivity before the hiring decision is made\(^\text{19}\).

In order to estimate the difference in the productivity of human capital in the two groups, the production function is assumed to be given by \(q(s, x) = A s^\beta x^\alpha\). This functional form is estimated using data on wages, education and the AFQT for workers with more than 15 years of experience. This requires the learning process to be mostly completed in 15 years. The estimated speeds of learning in the previous section suggest that even when education is the only screening device, 54% and 70% of the variation in the unobserved ability is uncovered within the first 5 years of employment in high and low AFQT occupations respectively. These numbers are even higher if the employers are using additional credentials to separate workers. After 15 years, 11% and 21% of this variation is yet to be uncovered. The assumptions on the functional structure of learning imply that after the first few years of experience, the marginal contribution of an additional output signal is small. This generates a small but persistent part of the variation in the unobserved component that remains to be uncovered even after in late in one’s career\(^\text{20}\). Table 5 displays the estimations results. \(A\) is normalized to 1, and \(\alpha\) and \(\beta\) are allowed to vary across the two groups. Return to education, estimated by \(\hat{\beta}\) is higher in the group with higher AFQT, as predicted by the model.

In the model, agents’ economic life begins when they start schooling, which corresponds to 6 years of age. The lifetime is set to 59 years in the simulations assuming that working life ends when agents are 65 years old. Mean and variance of log-ability,\(\ln x\), are set to 0 and 1 respectively to match the overall distribution of the AFQT score. The aggregate production parameter \(\phi\) is set to 0.56 to match the share of total income that goes to each occupation group. Occupational taste parameter, \(\tau_i\), is assumed to

\[\text{\textsuperscript{19}Both of these statements are true only to the extent that the variation in the initial information available to the firms can be captured by the AFQT residuals in the data, and equally so in the two occupational categories.}\]

\[\text{\textsuperscript{20}For workers with 45 years of experience, 5% and 9% of the initial uncertainty remains unknown in high and low AFQT sectors respectively.}\]
have a zero mean. The remaining parameters of the model are the discount rate, $r$, the learning content in the two sectors, $\rho_i$, and the standard deviation of the occupational taste parameter, $\sigma^2_t$. We simulate these parameters to match the education profiles in the two occupations and the differences in the AFQT. The particular moments we match are the average AFQT in the two sectors, average education, and the gradient of the education-AFQT profile. The model predicts that conditional on AFQT, the size of signaling raises educational achievement, and also makes the education-AFQT profile steeper.

Table 4 displays the regression results from the projection of education on AFQT, controlling for background variables. For the high AFQT group, a standard deviation increase in the AFQT score leads to an increase of 1.25 years in educational attainment, whereas this figure is 0.81 years for the low AFQT group. Table 6 summarizes the parameters of the model and their values used in the simulation.

The simulation results are summarized in Table 7. The first two rows display the true moments obtained from the data and their standard errors, respectively. Given these moments, four parameters $\{r, \rho_1, \rho_2, \sigma^2_{\text{tau}}\}$ are chosen to minimize the distance between the simulated moments and the true moments. The last two rows show the simulated moments and their standard errors based on several replications of the minimization process. The model closely generates the differences in educational attainment in the two occupational groups. The model undershoots the slope of the education-AFQT profile in general, but produces a steeper profile for the high AFQT group, as observed in the data.

Table 8 shows the parameter values associated with the simulated moments. The estimated discount rate is 9.8% which is considerably high compared to the average interest rate in the US economy. This basically is due to the fact that the only cost of education in the model is foregone earnings. Without any monetary costs, more reasonable interest rates generate substantially higher educational attainment. In order to bring the educational attainment closer to that observed in the data, the model has to compensate for tuition costs by choosing a higher discount rate.

The model suggests that 24% of the uncertainty about productivity is resolved prior to employment for low skilled workers, and the remaining is learned with experience. This figure is 95% for the high AFQT occupations. These numbers, combined with the estimated speeds of learning enables the calculation of the return to signaling. The average return to signaling for low skilled workers is 22% of the total OLS return, where it is only 1% for the high skilled group. Using the standard errors on initial knowledge, one can calculate the return to signaling within two standard deviations from the average value of $\rho$. The lower and the upper bound of return to signaling for low skilled workers are 19% and 25% of the OLS return. For high skilled workers, the return to signaling is 0 to 2.7% of the OLS return. These bounds are tight but they should be taken with caution. Some of the parameters used in the simulations, such as the speed of employer learning, also have standard errors that would result in wider bounds for signaling.
The identification of the signaling levels in both sectors rely on the endogeneity of occupation choice. To see this suppose the simulated model without any role for signaling, $\rho_1 = \rho_2 = 0$, matches perfectly the average ability across groups but generates a substantial difference in educational achievement. Also suppose that the interest rate is adjusted such that the educational level of the high AFQT sector is matched exactly to the data. The only correction needed in this scenario is to raise the educational achievement of low AFQT group. Since the only free parameters at hand are the levels of signaling in each sector, this must be achieved by a higher signaling role of education in the low AFQT sector, $\rho_1 > 0$. Had the markets been disconnected, this would work perfectly and the signaling role of education for the high AFQT group would have been unidentified. However with endogenous occupation choice, presence of signaling in the low AFQT group sorts more workers into the high AFQT sector. The general equilibrium adjustments through efficiency wages less than compensates for this effect. The net effect is a higher proportion of workers in the high AFQT sector. This decreases the average AFQT in both sectors. The remaining free parameter is the signaling role of education in the high AFQT sector. Increasing this as well brings the average AFQT back to its initial point but at a cost of higher education differences. Then the determination of the signaling role of education in both sectors boils down to the relative weights assigned to the differences in education profiles and the average AFQT in each sector.

The model assumes that the cost of education is similar for workers of different ability. The sorting behavior can also be generated by modeling the heterogeneity in schooling decisions on the cost side. Allowing the interest rate to decrease (increase) by ability will increase (decrease) the differences in the level of signaling. If the cost of education is non-increasing in ability, then the differences in the educational achievements implied by the human capital would be even wider. This would necessitate an even wider gap in the size of the signaling across sectors to explain the observed differences in education profiles.

**Hypothetical Experiments**

In order to get a better sense of the model I carry out two numerical experiments that are of particular interest. The focus is kept on the differences in the educational choices and occupational sorting behavior. The first experiment asks whether the human capital theory can explain the differences in educational profiles across the high and low skill sectors. Keeping all the parameters at their estimated values, I shut down the asymmetric information in the model by setting $\rho_1 = \rho_2 = 1$. This implies that the uncertainty about workers’ true productivity is completely resolved prior to entry by additional screening devices, therefore leaving no room for signaling via education. Third row in Table 9 shows the results. The difference in educational achievement implied by the differences in AFQT and productivity of human capital is 6.8 years, much higher than the observed difference of 2.6 years.

The educational achievement decreases in both occupations. Since the signaling is
particularly important for the low ability occupations, eliminating it results in lower educational attainment for this group. However, this can not be the only reason to why the educational achievement decreases in the occupational category with an already small role for signaling. The additional reason is that the average AFQT decreases in the high AFQT sector once workers update their occupational choices under the new set of parameters. Shutting down the signaling curbs the incentives for occupational sorting, because the only incentive to sort comes from the higher productivity of human capital in the high AFQT sector. The standard deviation of the personal taste for occupational choice is now relatively larger compared to the strength of sorting. This generates an increase in the noise of the sorting behavior, and results in a lower difference in AFQT scores between the two groups. This contributes to the decline in the average education in the high AFQT sector.

The Second experiment emphasizes the role of the learning content in the model’s ability to match the moments. For this experiment $\rho_1$ and $\rho_2$ are both set to zero. In this case, the size of signaling is determined by the speed of learning. Since the group with higher AFQT is subject to slower employer learning, the return to signaling is relatively higher. This results in a substantial increase in the education gap, because the high AFQT sector also has a higher true return to education. This is despite the distorted sorting behavior. The sector with higher productivity is now also subject to larger role for signaling. These two characteristics offset each other, and the differences in the AFQT scores simulated from the model completely vanish.

5 Conclusion

The analysis presented above emphasizes two crucial results that contributes to the understanding of the educational attainment and occupational choice under asymmetric information and the presence of employer learning. Testing the signaling hypothesis has traditionally relied on the comparison of markets that are potentially different in the signaling role they assign to education. The workers in these markets systematically differ in their ability characteristics if the market choice is endogenous. In particular, markets that rely less on education as a signaling device, draw in the workers with higher ability.

Using the data from the NLSY, the economy is divided into two broad occupational groups based on the the average AFQT score. Accounting for differences in the productivity of human capital the differences in educational profiles and the average ability implied by the model are compared to those observed in the data. The two sets of moments can only be reconciled if the signaling role of education is more important in the lower ability group.

The effect of signaling on these moments in each sector is further disciplined by the speed of employer learning estimated from the data. The estimated speeds signify the role of additional screening devices other than education. In particular, the extent of the signaling function of education crucially depends on how well the employers can
gauge the productivity of the workers using alternative screening methods. Results suggest that these methods are used intensively in occupations where it is harder to observe workers’ productivity on the job.

Estimation results show that the signaling role of education is equal to 22% of the return to education estimated by OLS for workers with low ability. For workers of higher ability most of the uncertainty is resolved prior to employment, leaving a very small role for signaling, 1% of the OLS return, despite the slower employer learning on the job.

The results reported here require that the learning process is symmetric across employer in the market, that workers commit to a market before they complete their schooling, and that post-school training are do not depend on education and ability. Relaxing these assumptions is a relevant target for future work.

References


6 Appendix

Derivations

**Claim 1** $\kappa_j$ is decreasing in $\Omega$

**Proof.** Rearrange $\kappa_j$ as

$$
\kappa_j = (1 - \beta) r^{- \frac{\beta}{1-\beta}} \left[ \frac{\beta + \Omega (1 - \beta)}{1-\Omega} \right]^{\frac{\beta}{1-\beta}}
$$
The derivative of $\Xi$ in squared brackets with respect to $\Omega$ is
\[
\frac{d\Xi}{d\Omega} = (1 - \beta)(1 - \Omega)^{\frac{1 - \beta}{\beta} - 1} \left[ (1 - \Omega) - (1 + \frac{\Omega(1 - \beta)}{\beta}) \right] < 0
\]

by chain rule
\[
\frac{d\kappa_j}{\Omega} = \frac{d\kappa_j}{d\Xi} \frac{d\Xi}{d\Omega} < 0
\]

**Lange’s Bound Revisited**
Denote the estimated return by $\hat{b}_{ls}$. By definition
\[
\hat{b}_{ls} = q_s(s, x) + q_x(s, x)\chi'(s) \geq q_s(s, x) + (1 - \rho)q_x(s, x)\chi'(s)
\]

implying $\hat{b}_{ls} - q_s(s, x) \geq (1 - \rho)q_x(s, x)\chi'$. Using this in the first order condition we have
\[
q_s(s, x) + \Omega(\hat{b}_{ls} - q_s(s, x)) \geq r
\]

Rearranging terms gives
\[
q_s(s, x) \geq \frac{(r - \Omega\hat{b}_{ls})}{1 - \Omega}
\]

This is his lower bound on the productivity effects of schooling. Using the first order condition one more time, we get an upper bound on signaling,
\[
\frac{(r - \Omega\hat{b}_{ls})}{1 - \Omega} + \Omega(1 - \rho)q_x(s, x)\chi'(s) \leq r
\]
or equivalently
\[
\Omega(1 - \rho)q_x(s, x)\chi'(s) \leq \frac{(\hat{b}_{ls} - r)\Omega}{1 - \Omega}
\]

which is equal to the one given in equation (13).

**Data**
The data is taken from the 1979 - 2004 waves of National Longitudinal Survey of Youth (NLSY). The NLSY consists of three samples. The main or cross-sectional sample is a random, nationally representative sample of 6,111 subjects between the ages of 14 and 21 at the time of the first interview in 1979. The supplemental sample of 5,295 subjects oversamples the hispanic, black or disadvantaged white population. The military sample consists of subjects aged 17-21 in September 1978, and were enlisted in the military. The analysis here is restricted to 3003 men of all races from the cross-sectional sample. 192 subjects who do not have AFQT scores in the data were dropped. Remaining AFQT scores are standardized within age groups.
The wage variable used in the regressions is the real average hourly rate of pay for the subject’s current or most recent job. Consumer price index was used to express wages in 2002 prices. Hourly wage observations less than a $1 and more than $100 are dropped. This resulted in a loss of 16 subjects. The analysis is restricted to jobs after the subject leaves the school for the first time. This is determined by the first interview when the respondent is not enrolled in school. Invalid observations for educational attainment and observations with less than 8 years of education are dropped. These restrictions results in a loss of 206 subjects that were either not finished with their education during the sample or do not have valid education higher than 8 years. The final data available for regressions consist of 2588 respondents 37136 observations. Further restrictions imposed by the availability of the variables in particular regressions, such as parents’ education, are mentioned in the text and the corresponding tables.
7 Tables and Figures

Figure 1: Occupational Choice under Differences in the Speed of Employer Learning
Figure 2: Education Choice under Differences in the Speed of Employer Learning

Figure 3: Education Choice with Bounded Taste for Occupations
Figure 4: The Speed of Learning across Occupational Categories

High AFQT Occupations

Low AFQT Occupations
Table 1: An Upper Bound on the Return to Signaling

<table>
<thead>
<tr>
<th>Discount Rate</th>
<th>Ω</th>
<th>Contribution to log-wages</th>
<th>% of OLS Return</th>
<th>% of Private Productivity Return</th>
<th>Productivity Effects</th>
<th>% of Private Productivity Return</th>
<th>Productivity Effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>3%</td>
<td>0.19</td>
<td>1.4%</td>
<td>15.8%</td>
<td>46.0%</td>
<td>1.6%</td>
<td>25.7%</td>
<td>3.4%</td>
</tr>
<tr>
<td>4%</td>
<td>0.21</td>
<td>1.3%</td>
<td>14.4%</td>
<td>31.3%</td>
<td>2.8%</td>
<td>19.6%</td>
<td>4.3%</td>
</tr>
<tr>
<td>5%</td>
<td>0.23</td>
<td>1.1%</td>
<td>12.5%</td>
<td>21.7%</td>
<td>3.9%</td>
<td>13.8%</td>
<td>5.3%</td>
</tr>
<tr>
<td>6%</td>
<td>0.24</td>
<td>0.9%</td>
<td>10.0%</td>
<td>14.5%</td>
<td>5.1%</td>
<td>8.4%</td>
<td>6.5%</td>
</tr>
<tr>
<td>7%</td>
<td>0.26</td>
<td>0.6%</td>
<td>6.9%</td>
<td>8.5%</td>
<td>6.4%</td>
<td>3.1%</td>
<td>7.9%</td>
</tr>
<tr>
<td>8%</td>
<td>0.28</td>
<td>0.3%</td>
<td>3.1%</td>
<td>3.3%</td>
<td>7.7%</td>
<td>&lt; 0</td>
<td>8.1%</td>
</tr>
<tr>
<td>9%</td>
<td>0.29</td>
<td>&lt; 0</td>
<td>&lt; 0</td>
<td>&lt; 0</td>
<td>9.1%</td>
<td>&lt; 0</td>
<td>9.3%</td>
</tr>
</tbody>
</table>

Reported values are based on a total return of 8.7% and speed of learning, $K = 0.2592$. Results are for workers with 12 years of education. Workers are assumed to work for 45 years after graduating from high school. These figures are taken from Lange (2005).
Table 2: Average AFQT Score Across Occupations

<table>
<thead>
<tr>
<th>Occupation Title</th>
<th>Percent Frequency</th>
<th>Average AFQT Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Professional and Technical Workers</td>
<td>0.15</td>
<td>0.79</td>
</tr>
<tr>
<td>Sales Workers</td>
<td>0.05</td>
<td>0.43</td>
</tr>
<tr>
<td>Managers and Administrators</td>
<td>0.13</td>
<td>0.42</td>
</tr>
<tr>
<td>Farmers and Farm Managers</td>
<td>0.01</td>
<td>0.21</td>
</tr>
<tr>
<td>Clerical and Unskilled Workers</td>
<td>0.07</td>
<td>0.16</td>
</tr>
<tr>
<td>Craftsmen and Kindred Workers</td>
<td>0.21</td>
<td>-0.22</td>
</tr>
<tr>
<td>Service Workers</td>
<td>0.10</td>
<td>-0.29</td>
</tr>
<tr>
<td>Operatives</td>
<td>0.11</td>
<td>-0.37</td>
</tr>
<tr>
<td>Transport Equipment Operatives</td>
<td>0.06</td>
<td>-0.43</td>
</tr>
<tr>
<td>Private Household Workers</td>
<td>0.00</td>
<td>-0.44</td>
</tr>
<tr>
<td>Farm Laborers and Farm Foremen</td>
<td>0.01</td>
<td>-0.51</td>
</tr>
<tr>
<td>Laborers</td>
<td>0.10</td>
<td>-0.53</td>
</tr>
</tbody>
</table>

Table 3: Speed of Employer Learning

<table>
<thead>
<tr>
<th>Gradient: High AFQT</th>
<th>Low AFQT</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \hat{\kappa}_{20} )</td>
<td>0.039</td>
</tr>
<tr>
<td>( \hat{\kappa}_{\infty} )</td>
<td>0.146</td>
</tr>
<tr>
<td>( \hat{\kappa}_{\infty} )</td>
<td>0.023</td>
</tr>
</tbody>
</table>
Table 4: Education and AFQT Score

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>High AFQT</th>
<th>Low AFQT</th>
</tr>
</thead>
<tbody>
<tr>
<td>AFQT</td>
<td>1.26</td>
<td>0.81</td>
</tr>
<tr>
<td>Library Card</td>
<td>0.44</td>
<td>0.18</td>
</tr>
<tr>
<td>Mother’s Education</td>
<td>0.11</td>
<td>0.07</td>
</tr>
<tr>
<td>Father’s Education</td>
<td>0.06</td>
<td>0.06</td>
</tr>
<tr>
<td>Black</td>
<td>1.15</td>
<td>0.77</td>
</tr>
<tr>
<td>Hispanic</td>
<td>0.55</td>
<td>0.37</td>
</tr>
<tr>
<td>Constant</td>
<td>12.04</td>
<td>10.37</td>
</tr>
</tbody>
</table>

N 11182 19516
R² 0.37 0.32
Average AFQT 0.59 -0.33
Average Education 14.64 12.02
Table 5: Educational Achievement and Wages

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>High</th>
<th>Low</th>
</tr>
</thead>
<tbody>
<tr>
<td>ln w</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AFQT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>constant</td>
<td>3.933</td>
<td>4.515</td>
</tr>
<tr>
<td></td>
<td>0.487</td>
<td>0.254</td>
</tr>
<tr>
<td>α</td>
<td>0.034</td>
<td>0.045</td>
</tr>
<tr>
<td></td>
<td>0.008</td>
<td>0.008</td>
</tr>
<tr>
<td>β</td>
<td>0.437</td>
<td>0.325</td>
</tr>
<tr>
<td></td>
<td>0.018</td>
<td>0.022</td>
</tr>
<tr>
<td>Experience</td>
<td>0.041</td>
<td>0.039</td>
</tr>
<tr>
<td></td>
<td>0.044</td>
<td>0.022</td>
</tr>
<tr>
<td>Experience²</td>
<td>-0.001</td>
<td>-0.001</td>
</tr>
<tr>
<td></td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td>Black</td>
<td>-0.113</td>
<td>-0.181</td>
</tr>
<tr>
<td></td>
<td>0.072</td>
<td>0.037</td>
</tr>
<tr>
<td>Hispanic</td>
<td>-0.064</td>
<td>-0.004</td>
</tr>
<tr>
<td></td>
<td>0.072</td>
<td>0.051</td>
</tr>
<tr>
<td>N</td>
<td>2568</td>
<td>5616</td>
</tr>
<tr>
<td>R²</td>
<td>0.22</td>
<td>0.15</td>
</tr>
</tbody>
</table>

Table 6: Parameters of the Model

<table>
<thead>
<tr>
<th>Parameter</th>
<th>High AFQT</th>
<th>Low AFQT</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>α</td>
<td>0.04</td>
<td>0.05</td>
</tr>
<tr>
<td>β</td>
<td>0.44</td>
<td>0.33</td>
</tr>
<tr>
<td>( \frac{\sigma_x^2}{\sigma_x^2 + \sigma_z^2} )</td>
<td>0.19</td>
<td>0.32</td>
</tr>
<tr>
<td>( \mu_{\ln x}, \sigma_{\ln x}^2 )</td>
<td>(0.1)</td>
<td>(0.1)</td>
</tr>
<tr>
<td>T</td>
<td>59</td>
<td>59</td>
</tr>
<tr>
<td>φ</td>
<td>0.56</td>
<td>0.56</td>
</tr>
</tbody>
</table>
Table 7: Simulation Results: Moments

<table>
<thead>
<tr>
<th></th>
<th>High Ability</th>
<th>Low ability</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Average AFQT</td>
<td>Average Education</td>
</tr>
<tr>
<td>Data</td>
<td>0.59</td>
<td>14.64</td>
</tr>
<tr>
<td>Moments</td>
<td>0.01</td>
<td>0.06</td>
</tr>
<tr>
<td>Simulation</td>
<td>0.44</td>
<td>15.06</td>
</tr>
<tr>
<td>Moments</td>
<td>0.02</td>
<td>0.05</td>
</tr>
</tbody>
</table>

Table 8: Simulation Results: The size of signaling

<table>
<thead>
<tr>
<th>Parameter</th>
<th>$r$</th>
<th>$\rho_1$</th>
<th>$\rho_2$</th>
<th>$\sigma_{\tau u}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimates</td>
<td>0.003</td>
<td>0.052</td>
<td>0.034</td>
<td>0.007</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Low Ability</th>
<th>High Ability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average Return to Signaling</td>
<td>22.0%</td>
<td>1.1%</td>
</tr>
<tr>
<td>Bounds on Return to Signaling</td>
<td>19.1%</td>
<td>24.7%</td>
</tr>
</tbody>
</table>

Table 9: Numerical Experiments

<table>
<thead>
<tr>
<th></th>
<th>High Ability</th>
<th>Low ability</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Average AFQT</td>
<td>Average Education</td>
</tr>
<tr>
<td>Data Moments</td>
<td>0.59</td>
<td>14.64</td>
</tr>
<tr>
<td>Standard Error</td>
<td>0.01</td>
<td>0.06</td>
</tr>
<tr>
<td>No Signaling</td>
<td>0.31</td>
<td>11.69</td>
</tr>
<tr>
<td>Full Signaling</td>
<td>-0.10</td>
<td>34.79</td>
</tr>
</tbody>
</table>