

Disentangling the College Wage Premium: Estimating a Model with Endogenous Education Choices*

Hanming Fang[†]

First Version: September 1998

This Version: October 1999

Abstract

College graduates on average earn more than high school graduates. This paper provides an equilibrium model of education choices which endogenously determine both the college wage premium and college enrollment rate. The model is then empirically implemented using 1990 U.S. Census 5% PUMS data to *quantitatively* evaluate the contributions of productivity enhancement and ability signaling to the college wage premium. We find that college education enhances attendees' productivity by about 40%. If college education did not enhance productivity at all, then the college wage premium would be *at most* 20.93% in contrast to about 58.64% in the data. From this we conclude that productivity enhancement accounts for *at least* 64.31% of the college premium observed in 1989. Other findings regarding the welfare consequences of college education are also obtained.

Key Words: College Wage Premium, Education Choices

JEL Classifications: I20, J24, J30

1 Introduction

A large empirical literature has documented the college wage premium and its fluctuations. Mincer (1974) is an early study on the “return” to schooling. Using the 1960 Census,

*I am deeply indebted to Professors Andrew Postlewaite and Ken Wolpin for their invaluable advice and encouragements. I also received help and useful comments from Jim Albrecht, Peter Arcidiacono, Jere Behrman, William Brock, Zvi Eckstein, Edward Green, George Mailath, Andrea Moro, Stephen Morris, Peter Norman, Rafael Rob, Dan Silverman, Petra Todd, Ruilin Zhou, participants at the Interuniversity Student Conference at Columbia University, Empir Micro Seminar at University of Pennsylvania and audience at the North American Econometric Society Summer Meeting in Madison. Ken Wolpin also kindly provided data sets of cost-of-living indices; and Andrea Moro generously helped in Fortran programming. I am responsible for all remaining errors. Financial support from Alfred P. Sloan Foundation in the form of a Doctoral Dissertation Fellowship is gratefully acknowledged.

[†]Department of Economics, University of Pennsylvania, Philadelphia, PA 19104. Email: hfang@econ.sas.upenn.edu

Mincer estimates that, on average, an additional year of schooling is associated with 7% to 10% higher earnings; or, equivalently, that college graduates earn about 50% more than high school graduates. More recent studies have shown that the college premium fluctuates over time. For example, Murphy and Welch (1989), using the Current Population Survey, documented that the college wage premium (in weekly wages) for young workers with 1-5 years of experience was about 46% in 1970, declined to about 31% in 1979, and then increased from 1979 to almost 70% in 1986. While Murphy and Welch's major concern is to explain the recent rise in the college wage premium, the goal of this paper is to understand why college graduates can command higher wages in labor markets.

There are two common theoretical explanations for the correlation between higher wages and college education. The first, which we call the *productivity enhancement* explanation, is associated with human capital theory pioneered by T.W. Schultz (1961) and further developed by Becker (1964) and Mincer (1974). It postulates that college education increases wages by directly increasing the worker's productivity. The second, which we call the *ability signaling* explanation, is first presented by Spence (1973). It postulates that college education is associated with higher wages because by acquiring a college education a worker signals to firms that he has higher innate ability than a high school graduate. That is, college wage premium can be generated even if college education is not productivity enhancing. A central difference between these two theories is that in human capital theory, different years of schooling is *responsible* for part of the differences of productivity among workers, while in signaling models, schooling is correlated with differences among workers that were present *before* the schooling choices were made.

Disentangling the contributions of productivity enhancement and ability signaling to the college wage premium is important since these two complementary (and sometimes competing) explanations have dramatically different policy implications. First, if productivity enhancement explains the college premium, college education contributes to economic growth directly by increasing the productivity of workers; while if ability signaling explains the college premium, then college education contributes to economic growth only to the extent that it facilitates a better match between workers and jobs (Stiglitz 1975). Second, the two theories have different implications for the congruence between private and social returns to education. For example, if productivity enhancement explains the college premium, and if the labor market is competitive, then the private return to education will coincide with the social return if there is no human capital externality, and will be lower than social return if there are positive externalities. However, if the ability signaling explanation is important, the marginal private returns to schooling could be higher or lower than the marginal social returns depending on the importance of the match between workers of different abilities and jobs of different ability requirements. Because individuals choose their levels of schooling based on the private rather than the social rate of returns, knowing the importance of productivity enhancement versus ability signaling explanations for the college wage premium can help us determine whether the market is likely to provide an efficient level of schooling.

A fundamental problem is then to empirically estimate the extent to which the college wage premium is explained by productivity enhancement and/or ability signaling factors. This question, however, has proven a challenging one. Weiss (1995) provides a critical survey of various attempts in the literature and his conclusion is that all the studies are rather inconclusive: Many findings are consistent with both productivity enhancement and signaling explanations.

One strand of this literature either tries to control unobservable worker characteristics using identical twins data (e.g. Ashenfelter and Krueger 1994)¹ or exploit some natural experiments resulting from, for example, changes in education laws, to estimate the private returns to education (e.g. Angrist and Krueger 1991).

Another strand of this literature tries to distinguish directly these two explanations by testing *in reduced form* different implications of the theories. Wolpin (1977) explains that a self-employed individual is more certain about her own ability than a firm is about the ability of its salaried workers. Therefore a self-employed worker is less likely to purchase schooling for motivations other than human capital investment. Wolpin then divides the NBER-Thorndike sample of World War II veterans into the self-employed and salaried workers and finds that the difference in years of schooling between these two groups is small, which suggests that ability signaling function of education is minor. Riley (1979) divides occupations into an “unscreened sector” for which a worker’s productivity may be easily ascertained and thus education signaling is unimportant, and a “screened sector” for which a worker’s productivity is difficult to observe and thus education signaling plays a more important role. The implication of Riley’s model is that for any given number of years in school, productivity levels and hence lifetime earnings are on average higher in the unscreened sector than in the screened sector. Riley confirms the above hypothesis using the Current Population Survey cross section data and thus concludes that ability signaling is a significant phenomenon. Albrecht (1981) argues that if education is used as an ability screen one should expect to observe that it is used more for separating out workers from outside for whom direct evidence of ability is weaker. He then tests this hypothesis using the recruitment record of auto workers by Volvo, and concludes that Volvo’s hiring behavior indicates no support for the signalling hypothesis.

The above studies provide an impression of the *qualitative* importance of ability signaling to the college wage premium. The goals of this paper are two-fold: First, I present a general equilibrium model of education choices which not only allows us to endogenously determine both college wage premium and college enrollment rate, but also is flexible enough so that both productivity enhancement and ability signaling components of the college wage premium are well defined; Second, I confront the model with 1990 U.S. Census data and take a structural approach to estimate the fundamental parameters of the model. Simulations using the estimated parameters show that the structural model does an excellent job in matching the wage distributions for both the high school and college graduates in 1989. I then use the estimates of the structural parameters of the model to conduct a counter-

¹The approach of using identical twin data to control for unobserved ability is pioneered by Behrman and Taubman (1976).

factual experiment, which can provide a *quantitative* assessment of the importance of ability signaling in the U.S. education. Specifically, I consider a hypothetical economy which differs from the actual economy only in that its college education did not enhance productivity at all, what would be the maximum equilibrium college wage premium? Since productivity enhancement component is ruled out in the hypothetical economy, its college wage premium is due only to ability signaling. We can then ascribe the difference in college wage premium between the actual and the hypothetical economy to productivity enhancement of the college education. An alternative counter-factual experiment would be to consider an economy in which worker’s ability was perfectly observable and investigate the equilibrium college wage premium in such a hypothetical economy.

In contrast to Spence (1973), we assume that ability is only one of *many* factors that determines whether an individual attends college.² For example, if the credit market is imperfect, two individuals from families with different financial resources will make different educational choices even if their innate abilities are identical. The ability signaling value of college education is determined by the average ability differences between a college and a high school student, and thus is limited by equilibrating forces from endogenous education choices.³ To the extent that agents, when making education choices, trade off the expected college wage premium with the utility cost of college education, the proportion of the population that finds attending college an optimal choice – the college enrollment rate – is informative of the distributions of ability and college attendance utility cost in the population.⁴

We propose a model in which each individual is characterized not only by his intrinsic ability a , but also by a variable v which is interpreted as his utility cost of college education. An individual’s pair of attributes (a, v) is privately known. An outsider knows only that in the population, a and v are distributed with cumulative distribution functions $\Phi(\cdot)$ and $H(\cdot)$ respectively.

Each individual, at the moment he graduates from junior high school, makes two sets of decisions: First, should he attend college or immediately enter the labor market after high school? Second, should he invest in skills that prepare him for more complex tasks (e.g.

²A natural question is that why we do not use an obvious extension of Spence original education signaling model. Suppose that agents differ only in their innate ability a , which is distributed in the population according to an absolutely continuous c.d.f, and as in Spence the college education cost is strictly decreasing in a . Then the unique equilibrium will be characterized by a cut-off level of ability: agents with ability above the cutoff will all attend the college and below the cut-off will not. Then one implication of such a model will be that the lowest wage earned by a college educated worker must be larger than the highest wage earned by a high school educated worker. Such a prediction is hard to reconcile with the data. In our sample, the lowest wage earned by college graduates is at about the *tenth* percentile of the high school wage distribution.

³Heuristically, a higher benefit of attending college will induce more people to go to college, but a changing population of college attendees will affect the average ability of both college and high school attendees, which in turn affects the ability signaling function of college education. We need to analyze the equilibrium of this whole process in order to understand the data.

⁴Though the variability of college enrollment rates has been documented by Murphy and Welch (1989), and U.S. Department of Education (1997), it has not been utilized in previous studies of returns to education. As we will see in Subsection 4.1.2, it plays a crucial role in our approach.

managerial jobs)? We assume that qualification for complex tasks only depends on whether the costly skill investment (e.g. mastering technical material) is undertaken, and not on college attendance. We make the assumption that skill investment cost is inversely related to the worker’s ability. Though college education does not affect a worker’s qualification, it does magnify a worker’s ability (in efficiency units) by a factor ρ .

Employers observe a worker’s educational attainment (college or high school), as well as some noisy signal of the worker’s qualification. Each firm has two tasks to be performed: a simple task that does not need qualification, and a complex task at which only qualified workers can succeed.⁵ Firms compete for the workers by offering wage schedules. It can be shown that in equilibrium a worker is paid his expected productivity in his more productive task. Thus in this paper, within education-group wage inequality arises because of the differences among workers in their qualifications and their luck, while between education-group wage inequality arises because of both the productivity enhancement of college education and the difference in average abilities between college and high school graduates.

It is shown in Fang (1998) that this model might generate multiple equilibria. Econometric implementation of models with multiple equilibria is often complicated by the lack of an equilibrium selection mechanism. Fortunately, the equilibrium of the model exhibits a particular structure that allows us to adopt an estimation strategy similar to that in Moro (1997). More specifically, given the education and qualification choices made by the agents, the subsequent wage distributions of the college and high school attendees are completely determined by some aggregates derived from the equilibrium education and qualification choices (Proposition 1). Hence we can use the wage distributions for each education group to estimate those aggregates. Subsequently we use our characterization of equilibrium education and qualification choices (Proposition 2) to determine the fundamentals of an economy that could have simultaneously generated those aggregates estimated in the first step, and the observed college enrollment rate.

We implement our model using U.S. 1990 Census 5% PUMS data. In our sample, the college wage premium is about 58.64% in 1989, which is consistent with evidence from other sources (e.g. U.S. Department of Education 1997).⁶ We estimate that college education magnifies an individual’s ability by about 40% (i.e. $\rho = 1.40$). We then conduct a counter-factual experiment and find that college wage premium would *at most* be 20.93% in a hypothetical economy in which ρ were equal to 1 (with other parameters held the same as the actual economy). Therefore, productivity enhancement accounts for *at least* $(58.64 - 20.93) / 58.64 = 64.31\%$ of college wage premium.

The remainder of the paper is structured as follows: Section 2 presents details of the model; Section 3 presents the theoretical analysis of the model, in particular, two useful propositions that allow us to empirically implement the model; Section 4 presents the

⁵The distinction of “complex task” is not crucial for the analysis, rather it is introduced to have a more natural interpretation of costly investment in school as being socially productive.

⁶College wage premium is defined as the ratio of the differences in average wages between college and high school graduates over the high school average wage.

estimation strategy and briefly discusses the identification problem; Section 5 describes the data set and presents the estimation results; Section 6 presents the result from a counter-factual experiment. Section 7 discusses the robustness of the estimation methodology and results; Section 8 concludes. Appendix A contains all the proofs, and Appendix B documents results from an alternative counter-factual experiment.

2 The Model

The model is a slight modification of Fang (1998). In a unit mass population, each agent is endowed with two characteristics: first, his innate ability in efficiency units, denoted by $a \in \mathbb{A} \subset \mathbb{R}$;⁷ second, his utility cost of attending college, denoted by $v \in \mathbb{V} \subset \mathbb{R}$.⁸ An agent's pair (a, v) is his private information. In the population a and v are distributed according to continuous and strictly increasing cumulative distribution functions Φ and H respectively, with ϕ and h the corresponding probability densities. For tractability, we assume that a and v are independent.⁹

Consider the population just after junior high school graduation, when all agents are contemplating whether they should enter the labor market after high school or after attending college.¹⁰ It is assumed that attending high school is compulsory, while attending college is not. We will write these two education choices as $\{\mathbf{h}, \mathbf{c}\}$ where \mathbf{h} stands for “high school” and \mathbf{c} for “college”.

There are two firms in the labor market, each with two technologies: one is called the *simple task*, and the other the *complex task*. We assume that all agents can perform the simple task, and an agent with y efficiency units produces y units of output at the simple task. However, an agent must undertake *costly skill investment* in order to succeed at the complex task. A worker with y efficiency units produces yx_q units of output at the complex task if he is qualified, and 0 if he is not. It is assumed that $x_q > 1$. Importantly, we assume that qualification for the complex task is achieved if and only if the agent invests in skills. College education is neither necessary nor sufficient for qualification. We assume that *regardless of his education choice*, the skill investment cost c for an agent with *innate* ability a is $k(a)$ where $k'(a) < 0$.¹¹

⁷Some authors (e.g. Wolpin 1977) view an individual's ability to be multi-dimensional. We are assuming that there exists a mapping to transform a multi-dimensional skill vector into a real number.

⁸The utility cost v can be due to college tuition fees, room and board and forgone income, or simply psychological cost. Arguably these costs do not depend on the individual's ability.

⁹The independence of a and v does not play any role in the theoretical analysis. In the econometric implementation, any joint distribution of a and v with four parameters can in principle be implemented.

¹⁰We assume here that an individual's education choice is not made sequentially. Disentangling college wage premium with sequential educational choices is left for future research.

¹¹Having two tasks are not essential for neither the characterization of equilibrium nor the estimation strategy. A single task model is nested in the two task model when $x_q = 1$. However, in later estimation, the null hypothesis $x_q = 1$ is rejected (see Table 2). But the estimates of the ability signaling component is only slightly changed if we restricted x_q to be 1.

Attending college augments a worker's efficiency units by a factor $\rho > 1$.^{12,13} For example, if a worker has innate ability a , then his efficiency units become ρa if he attends college. Hence his productivity in the simple task will be ρa and if he also invests in skills in college, his productivity on the complex task will be $\rho a x_q$.

After an agent has chosen education level $j \in \{\mathbf{h}, \mathbf{c}\}$, and made qualification decision, he receives a signal $\theta \in [0, 1]$ according to p.d.f. $f_q(\cdot)$ if he is qualified and $f_u(\cdot)$ if unqualified.¹⁴ It is assumed that f_q and f_u are continuously differentiable and satisfy:

Monotone Likelihood Ratio Property (MLRP): f_q/f_u is strictly increasing in θ .

The MLRP implies that qualified workers are more likely to receive higher test signals than unqualified workers. We denote F_q and F_u as their associated cumulative distributions.

Firms observe a worker's educational choice and his test signal, but do not observe his qualification. They simultaneously offer wage schedules to compete for workers. Observing the wage schedules, workers decide for which firm to work. Finally each firm allocate the workers who accept its offers between the tasks.

We now describe the model and strategies in more detail.¹⁵

Stage 1 (Education Choice). An agent with characteristics (a, v) chooses $j \in \{\mathbf{h}, \mathbf{c}\}$. He incurs a cost v if he attends college and zero cost otherwise. Write the education choice profile in the population as $J : \mathbb{A} \times \mathbb{V} \rightarrow \{\mathbf{c}, \mathbf{h}\}$. We assume that an agent's education choice is observable to firms.

Stage 2 (Qualification Decision).¹⁶ After the education choice is made, an agent of type (a, v) chooses an action from $\{e_q, e_u\}$, where e_q means that he invests in skills (and qualifies for the complex task) and e_u that he does not invest (and thus remains unqualified). He pays a utility cost of $c = k(a)$ if he chooses e_q , and no cost if he chooses e_u . Write the qualification decision profile in the population as $e : \mathbb{A} \times \mathbb{V} \rightarrow \{e_q, e_u\}$. The firms do not observe the workers' qualification decisions.

Stage 3 (Test Signal). Each agent receives a signal $\theta \in [0, 1]$. The worker's qualification decision e affects the distribution of his signal: The signal θ is distributed according to probability density function f_q for workers who chose e_q in Stage 2 and f_u for those who chose e_u . It is assumed that f_q and f_u are continuously differentiable, bounded away from zero and satisfy strict MLRP. A worker's signal is observable to both firms.

¹²The interpretation of ρ is that college education provides general human capital that makes the individual more productive at every job in his future career.

¹³More generally we should allow ρ to depend on a . It is not clear, though, whether ρ should increase or decrease in a . We consider the special case that $\rho(a)$ is a constant. See Section 7 for further discussions.

¹⁴In the theoretical analysis, we assume for simplicity that education choice j does not affect the test technology. In the econometric implementation, we allow the testing technology to differ for college and high school attendees.

¹⁵We adopt the following notational convention: firms are indexed by a subscript $i = 1, 2$, and education groups are indexed by a superscript $j = \mathbf{h}, \mathbf{c}$.

¹⁶The sequential order of stage 1 and 2 does not play a role.

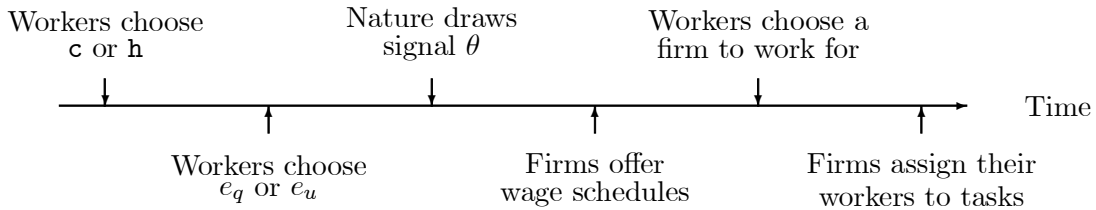


Figure 1: The Time Line

Stage 4 (Wage Offers). The firms compete in the labor market for workers by simultaneously announcing wage schedules. The wage schedules may depend on the worker's educational attainment, as well as on the test signal θ . A pure strategy of firm i at this stage is a measurable function $\mathbf{w}_i : \{\mathbf{h}, \mathbf{c}\} \times [0, 1] \rightarrow \mathfrak{R}_+$. We write $\mathbf{w}_i \equiv (w_i^{\mathbf{h}}, w_i^{\mathbf{c}})$ where $w_i^j : [0, 1] \rightarrow \mathfrak{R}_+$. We write \mathbb{W} as the set of all feasible wage offers, i.e., $\mathbb{W} \equiv \{\mathbf{w} | \mathbf{w} : \{\mathbf{h}, \mathbf{c}\} \times [0, 1] \rightarrow \mathfrak{R}_+\}$.¹⁷

Stage 5 (Offer Acceptance). The workers observe the wage schedules \mathbf{w}_1 and \mathbf{w}_2 announced in Stage 4, and decide for which firm to work. A worker's offer acceptance rule is a pair of maps $I_i : \mathbb{W}^2 \times [0, 1] \times \{\mathbf{h}, \mathbf{c}\} \rightarrow [0, 1], i = 1, 2$. For notational simplicity, we write $I_{i < \mathbf{w}_1, \mathbf{w}_2 >}^j(\theta)$ to represent $I_i(\mathbf{w}_1, \mathbf{w}_2, \theta, j)$. For example, $I_{i < \mathbf{w}_1, \mathbf{w}_2 >}^{\mathbf{c}}(\theta) = p$ means that given \mathbf{w}_1 and \mathbf{w}_2 , a college attendee with test signal θ accepts firm i 's offer with probability p . We require that $\sum_{i=1}^2 I_{i < \mathbf{w}_1, \mathbf{w}_2 >}^j(\theta) = 1$ for all $\theta \in [0, 1], j \in \{\mathbf{h}, \mathbf{c}\}$ and all $(\mathbf{w}_1, \mathbf{w}_2) \in \mathbb{W}^2$.

Stage 6 (Task Assignment). In the final stage each firm allocates its available workers between the complex and simple tasks using an assignment rule which is a measurable function $t_i : [0, 1] \times \{\mathbf{c}, \mathbf{h}\} \rightarrow \{0, 1\}$. The interpretation is that if $t_i(\theta, j) = 1$ (respectively, 0) means that firm i assigns all workers from education group j with signal θ to the complex (respectively, simple) task.

We summarize the timing of the game in Figure 1. Finally we assume that workers are risk neutral and their payoff functions are additively separable in their wage income, their skill investment costs and their utility costs of attending college. We assume that workers do not assign an intrinsic value to working at either task. If a worker is of type (a, v) and is paid wage w , his payoff will be: $w - v(j) - c(e, a)$ where $v(\mathbf{c}) = v, v(\mathbf{h}) = 0, c(e_q, a) = k(a)$ and $c(e_u, a) = 0$. We assume that the firms are risk neutral and maximize expected profits.

¹⁷To simplify the description of the firms' strategy set, we assumed that the firms can observe neither the aggregate distribution of test signals nor the college enrollment rate when announcing wage schedules. This assumption does not affect any of the equilibrium outcomes.

3 Analysis of the Model

We begin our analysis by deriving the distribution of skill investment costs c in the population from the distribution of abilities Φ . Denote the c.d.f. of c as G and its corresponding p.d.f. g . Since $c(e_q, a) = k(a)$, we have by change-of-variable:

$$g(c) = \phi(k^{-1}(c)) \left| \frac{dk^{-1}(c)}{dc} \right|, \text{ for all } c \in \mathbb{C} \equiv \{k(a) : a \in \mathbb{A}\}, \quad (1)$$

where \mathbb{C} is the support of g . Because there is a one-to-one relationship between an agent's ability a and his qualification cost c , it is sometimes useful to index an agent of type $(a, v) \in \mathbb{A} \times \mathbb{V}$ by $(c, v) = (k(a), v) \in \mathbb{C} \times \mathbb{V}$.

We use *Nash Equilibrium* as our equilibrium concept.¹⁸ A Nash equilibrium of the economy consists of a list of decision profiles of agents: $\{J, e, I_1, I_2\}$, and a pair of decision rules by each firm $\{\mathbf{w}_i, t_i\}$ for $i = 1, 2$, such that: (1) every agent's decision rules are optimal given other agents' and the firms' decision rules; (2) each firm's decision rules are optimal given the other firm's decision rules and agents' decision profiles.

The Nash equilibrium of the economy can be analyzed by backward induction.

Equilibrium Task Assignments: In Stage 6 (task assignment stage), suppose that a non-empty set of the workers has accepted firm i 's wage offer. Denote Θ_i^c (respectively, Θ_i^h) as the set of test signals realized by college (respectively, high school) attendees who accept firm i 's offer. Define $\Theta_i \equiv \Theta_i^c \cup \Theta_i^h$. Since in this stage, the task assignment rule of a firm is no longer strategic, we can simply analyze the optimal task assignment rule for firm i facing Θ_i . It is proved as Lemma 2 in the appendix that the optimal task assignment rule for firm i is of the form:

$$t_i(\theta, j) = \begin{cases} 1 & \text{if } \theta \geq \tilde{\theta}_i^j \\ 0 & \text{if } \theta < \tilde{\theta}_i^j \end{cases}, j = \mathbf{h}, \mathbf{c}. \quad (2)$$

This cut-off property is a consequence of the strict MLRP of the test technology.

Equilibrium Offer Acceptance Decisions: In Stage 5, when a worker from education group j with test signal θ is offered a pair of wage schedules, $\{w_1^j(\theta), w_2^j(\theta)\}$, his optimal offer acceptance decision, assuming without loss of generality that he flips a coin when he is indifferent, is:

$$I_{i < \mathbf{w}_1, \mathbf{w}_2}^j(\theta) = \begin{cases} 1 & \text{if } w_i^j(\theta) > w_{i'}^j(\theta) \\ \frac{1}{2} & \text{if } w_i^j(\theta) = w_{i'}^j(\theta), \text{ where } i \neq i' \\ 0 & \text{if } w_i^j(\theta) < w_{i'}^j(\theta) \end{cases} \quad (3)$$

¹⁸Note that the only possible off equilibrium path in the model arises in the task assignment stage, which is not strategic. It can be shown that any Nash equilibrium of the model can be properly modified off the equilibrium path into a Perfect Bayesian Equilibrium. Hence using Perfect Bayesian Nash instead of Nash will not affect any of the equilibrium outcome.

The set of test signals realized by firm i 's workers with educational attainment j , Θ^j , is the support of the function $I_{i < \mathbf{w}_1, \mathbf{w}_2}^j$.

Equilibrium Wage Schedules: Now we consider the firms' wage offering decisions in Stage 4. We prove as Lemma 1 in the appendix that in equilibrium the wage schedules offered by the two firms to workers from education group j are almost everywhere equal, that is, $w_1^j(\theta) = w_2^j(\theta)$ except possibly for a set of signals with measure zero. The argument is simple: Suppose to the contrary one firm offers a higher wage to a positive measure set of workers from education group j , that firm could profitably deviate by cutting the difference in half, thereby lowering the wage bill while keeping its work force unchanged. Together with the offer acceptance rule (3), this in particular implies that the set of test signals for the two firms are the same, i.e., $\Theta_1^j = \Theta_2^j$. Hence in the equilibrium task assignment rule (2) it must be that $\tilde{\theta}_1^j = \tilde{\theta}_2^j$ because the two firms are solving the same problem. Write $\tilde{\theta}^j$ as the common threshold.

Now we arbitrarily fix the education choice profile J and skill investment decision profile e and ask what wage schedules are compatible with equilibrium. We write $w^j(\theta|J, e)$ as the firms' common sequentially rational wage schedules following (J, e) . Write π^j as the fraction of qualified workers in education group j implied by (J, e) . Denote $\mathbb{E}^j[a|e(a, v) = e_q]$ (respectively, $\mathbb{E}^j[a|e(a, v) = e_u]$) as the expected *innate* ability in efficiency units of a worker from education group j who chooses to be qualified (respectively, unqualified). For notational simplicity we write $f_{\pi^j}(\theta) = \pi^j f_q(\theta) + (1 - \pi^j) f_u(\theta)$ as the probability density of test signals in education group j when its fraction of qualified workers is π^j . It can be shown that in equilibrium, with $\rho^c = \rho$ and $\rho^h = 1$,

$$w^j(\theta|J, e) = \max \left\{ \frac{\pi^j f_q(\theta)}{f_{\pi^j}(\theta)} \rho^j \mathbb{E}^j[a|e(a, v) = e_q] + \frac{(1 - \pi^j) f_u(\theta)}{f_{\pi^j}(\theta)} \rho^j \mathbb{E}^j[a|e(a, v) = e_u], \right. \\ \left. \frac{\pi^j f_q(\theta)}{f_{\pi^j}(\theta)} \rho^j \mathbb{E}^j[a|e(a, v) = e_q] x_q \right\}. \quad (4)$$

Note that π^j serves as the firms' prior belief about a randomly drawn worker from education group j being qualified, and $\pi^j f_q(\theta) / f_{\pi^j}(\theta)$ is their posterior belief that the worker is qualified when a test signal θ is realized. In the event that the worker is qualified, his expected productivity on the simple task is ρ^j times the expected innate ability conditional on his choosing e_q ; similarly in the event that he is unqualified, his expected productivity on the simple task is ρ^j times the expected innate ability conditional on his choosing e_u . His expected productivity on the complex task is the probability that he is qualified times the expected efficiency units conditional on his being qualified times x_q . Thus formula (4) essentially states that each worker is paid his expected productivity in his more productive task. We will henceforth write $\mathbb{E}^j(a|e_q)$ and $\mathbb{E}^j(a|e_u)$ as short-hand for $\mathbb{E}^j[a|e(a, v) = e_q]$ and $\mathbb{E}^j[a|e(a, v) = e_u]$, respectively.

Proposition 1 *Suppose in a Nash equilibrium agents' education choice and skill investment decision profiles are respectively J and e . Then the firms' wage offers must be given by (4)*

where,

$$\pi^j = \frac{\int_{\{(a,v) \in J^{-1}(j) \cap e^{-1}(e_q)\}} d\Phi(a) dH(v)}{\int_{\{(a,v) \in J^{-1}(j)\}} d\Phi(a) dH(v)} \quad (5a)$$

$$\mathbb{E}^j(a|e_q) = \frac{\int_{\{(a,v) \in J^{-1}(j) \cap e^{-1}(e_q)\}} ad\Phi(a) dH(v)}{\int_{\{(a,v) \in J^{-1}(j) \cap e^{-1}(e_q)\}} d\Phi(a) dH(v)} \quad (5b)$$

$$\mathbb{E}^j(a|e_u) = \frac{\int_{\{(a,v) \in J^{-1}(j) \cap e^{-1}(e_u)\}} ad\Phi(a) dH(v)}{\int_{\{(a,v) \in J^{-1}(j) \cap e^{-1}(e_u)\}} d\Phi(a) dH(v)} \quad (5c)$$

We can alternatively define the *threshold signal* for education group j in the task assignment rule $\tilde{\theta}^j$ by:

$$\frac{\pi^j f_q(\tilde{\theta}^j)}{f_{\pi^j}(\tilde{\theta}^j)} \mathbb{E}^j(a|e_q) + \frac{(1 - \pi^j) f_u(\tilde{\theta}^j)}{f_{\pi^j}(\tilde{\theta}^j)} \mathbb{E}^j(a|e_u) = \frac{\pi^j f_q(\tilde{\theta}^j)}{f_{\pi^j}(\tilde{\theta}^j)} \mathbb{E}^j(a|e_q) x_q \quad (6)$$

if a solution to equation (6) exists. Due to strict MLRP, the solution is unique if it exists. Set $\tilde{\theta}^j = 1$ if there is no solution to equation (6). The wage offer schedule (4) can then be rewritten as:

$$w^j(\pi^j|J, e) = \begin{cases} \frac{\pi^j f_q(\theta)}{f_{\pi^j}(\theta)} \rho^j \mathbb{E}^j(a|e_q) x_q & \text{if } \theta \geq \tilde{\theta}^j \\ \frac{\pi^j f_q(\theta)}{f_{\pi^j}(\theta)} \rho^j \mathbb{E}^j(a|e_q) + \frac{(1 - \pi^j) f_u(\theta)}{f_{\pi^j}(\theta)} \rho^j \mathbb{E}^j(a|e_u) & \text{if } \theta < \tilde{\theta}^j. \end{cases} \quad (7)$$

From either expression (4) or (7), one can see that in this model, the within-education-group wage variation among workers only comes from different values of test signal θ , which can be due to luck, and/or to qualification; while between-education-group wage variation can come from differences in ρ^j , π^j , $\mathbb{E}^j(a|e_q)$ and $\mathbb{E}^j(a|e_u)$.

We will now characterize the equilibrium education choice and skill investment decision profiles (J, e) . For this purpose it is more convenient to refer to each worker by his characteristics (c, v) . Define $\hat{J} : \mathbb{C} \times \mathbb{V} \rightarrow \{c, \mathbf{h}\}$ and $\hat{e} : \mathbb{C} \times \mathbb{V} \rightarrow \{e_q, e_u\}$ respectively by:

$$\begin{aligned} \hat{J}(c, v) &= J(k^{-1}(c), v) \\ \hat{e}(c, v) &= e(k^{-1}(c), v). \end{aligned}$$

Definition 1 Fix education choice and skill investment profiles (J, e) . The **membership sets** Ω_q^j and Ω_u^j under (J, e) are defined to be respectively the set of qualified and unqualified workers in education group j :

$$\begin{aligned} \Omega_q^j &= \{(c, v) : \hat{J}(c, v) = j, \hat{e}(c, v) = e_q\} \\ \Omega_u^j &= \{(c, v) : \hat{J}(c, v) = j, \hat{e}(c, v) = e_u\}. \end{aligned}$$

We write $\Omega^j = \Omega_q^j \cup \Omega_u^j$ as the set of workers from education group j .

Characterizing Ω_q^j and Ω_u^j provides the same information as characterizing (J, e) .

Definition 2 Let (J, e) be the education choice and investment decision profiles in a Nash equilibrium. **Expected gross equilibrium continuation payoff** for a qualified and an unqualified worker in education group j , denoted by V_q^j and V_u^j respectively, are:

$$V_q^j = \int_0^1 w^j(\theta|J, e) f_q(\theta) d\theta \quad (8)$$

$$V_u^j = \int_0^1 w^j(\theta|J, e) f_u(\theta) d\theta \quad (9)$$

where $w(\theta|J, e)$ is given by (4). Further, define

$$\tilde{v}_q = V_q^c - V_q^h \quad (10)$$

$$\tilde{v}_u = V_u^c - V_u^h \quad (11)$$

$$\tilde{c}^j = V_q^j - V_u^j. \quad (12)$$

In words, \tilde{v}_q (\tilde{v}_u) is the threshold value of v that a would-be qualified (unqualified) agent is willing to give up to enjoy the wage schedule offered to college attendees; and \tilde{c}^j is the threshold value of skill investment cost c that a group j agent is willing to incur to become a qualified worker in that group. Note that $\tilde{v}_q - \tilde{v}_u = \tilde{c}^c - \tilde{c}^h$.

Finally, if $\tilde{v}_q - \tilde{v}_u > 0$, we define a piecewise linear function $L : \mathbb{C} \rightarrow \mathbb{V}$ by:

$$L(c) = \begin{cases} \tilde{v}_q & \text{if } c \in [\underline{c}, \tilde{c}^h] \\ \tilde{v}_q + \tilde{c}^h - c & \text{if } c \in [\tilde{c}^h, \tilde{c}^c] \\ \tilde{v}_u & \text{if } c \in [\tilde{c}^c, \bar{c}] \end{cases}. \quad (13)$$

The following proposition characterizes the membership sets. We will state the proposition assuming that $\tilde{v}_q > \tilde{v}_u$.

Proposition 2 Suppose (J, e) is the education choice and investment decision profiles in a Nash equilibrium. If $\tilde{v}_q > \tilde{v}_u$, then:

1. $\Omega^c = \{(c, v) : c \in \mathbb{C}, v \leq L(c)\}$;
2. $\Omega_u^c = \{(c, v) : c \in [\tilde{c}^c, \bar{c}], v \in [\underline{v}, \tilde{v}_u]\}$;
3. $\Omega^h = \{(c, v) : c \in \mathbb{C}, v > L(c)\}$;
4. $\Omega_q^h = \{(c, v) : c \in [\underline{c}, \tilde{c}^h], v \in [\tilde{v}_q, \bar{v}]\}$.

Proposition 2 is illustrated in Figure (2). It says that agents with low investment cost c but high college attendance cost v will choose not to attend college but invest in skills when in high school; while agents with low college attendance cost v but with high skill investment cost c will choose to attend college but not to invest in skills. The 45

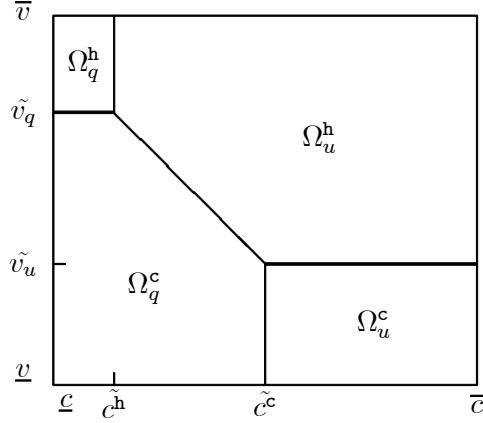


Figure 2: Equilibrium Education and Skill Investment Choices

degree straight-line that separates Ω_q^c and Ω_u^h is a consequence of the additively separable preferences. The proposition is proved using “revealed preference” arguments. It is worth noting that Proposition 2 holds true regardless of whether c and v are independent.

With the configuration of the equilibrium membership sets characterized in Proposition 2, we can compute the values of $\mathbb{E}^j(a|e_q)$ and $\mathbb{E}^j(a|e_u)$ if we know the four numbers $\tilde{v}_q, \tilde{v}_u, \tilde{c}^c, \tilde{c}^h$. To see this, note that knowing Ω^j , we can in principle determine the marginal distribution of c over Ω^j as g^j . But then:

$$\mathbb{E}^j(a|e_q) = \int_{c < \tilde{c}^j} k^{-1}(c) g^j(c) dc \quad (14)$$

$$\mathbb{E}^j(a|e_u) = \int_{c > \tilde{c}^j} k^{-1}(c) g^j(c) dc. \quad (15)$$

We define *college enrollment rate* λ^c as the fraction of agents from a certain age cohort that attends college. It can be calculated in this model economy as:

$$\lambda^c = \int_{\Omega^c} dG(c) dH(v). \quad (16)$$

Note that characterizations of wage offer schedules (Proposition 1) and membership sets (Proposition 2) are premised on a particular Nash equilibrium of the economy. In a closely related model, Fang (1998) shows that multiple equilibria might exist under certain distributions G and H . The existence of equilibrium is not an issue for our present purpose because we interpret the data as an equilibrium of *some* economy.

4 A Parameterization and the Estimation Strategy

The above theoretical framework allows us to find equilibria for very general environments. However since we are interested in empirical implementation, further development

of the model is best done in a more fully parameterized context. In this section, we present in detail the estimation strategy and describe the parametric assumptions in the process. A brief discussion of identification is also included to provide some intuition.

4.1 Estimation Strategy

The model potentially admits multiple equilibria (See Fang 1998). Nevertheless, the structure of the equilibrium of the model permits estimating the structural parameters in three steps. The crucial idea is to realize a particular structure of the model highlighted by Proposition 1: Given education choice and skill investment decision profiles (J, e) , the subsequent wage distributions of college and high school graduates are unique.¹⁹ The first step of the estimation procedure is to separately estimate the relevant variables that characterize the wage distributions of college and high school graduates using maximum likelihood. Those estimated variables of the wage distribution of education group j provide us with information about $\pi^j, \mathbb{E}^j(a|e_q), \mathbb{E}^j(a|e_u)$, which are closely related to (J, e) as expressed by equations (5a)-(5c). The first step estimation also provides us with some idea of how the underlying theory of wage distributions fares with the data. The wage offer schedules estimated in the first step enables us to compute the values of $\tilde{v}_q, \tilde{v}_u, \tilde{c}^c$ and \tilde{c}^h via equations (8) through (12). In the second step of the estimation procedure, we use the characterization of the membership sets provided by Proposition 2 to solve, for a certain class of parameterization of the underlying distributions G, H and the cost-ability relation k , their parameters. In the final step, we utilize the estimates we get in the first step, and the functions G, H and k that are estimated in the second step, to solve the parameter ρ^c , which measures the productivity enhancement of college education.

4.1.1 First Step: Matching The Wage Distributions

Here we will describe in greater detail how we can formulate the likelihood function of a sample of wage observations of group j members.

To begin we impose a parametric form for the distribution of test signals, f_q^j and f_u^j :

$$f_q^j(\theta) = \eta^j \theta^{\eta^j - 1} \quad (17)$$

$$f_u^j(\theta) = \eta^j (1 - \theta)^{\eta^j - 1}. \quad (18)$$

for $j = h, c$. The above parametrization of the test technology satisfies our MLRP assumption if $\eta^j > 1$.²⁰

Using the above parameterization of f_q^j and f_u^j , we can rewrite equation (6) as:

$$\mathbb{E}^j(a|e_q) x_q^j = \mathbb{E}^j(a|e_q) + \frac{1 - \pi^j}{\pi^j} \left(\frac{1 - \tilde{\theta}^j}{\tilde{\theta}^j} \right)^{\eta^j - 1} \mathbb{E}^j(a|e_u), \quad (19)$$

¹⁹The same structure is first exploited in Moro (1997).

²⁰Note that we allow η^j to be different for high school and college. We also allow x_q^j to be different for high school and college. It turns out that their estimated differences are rather small.

INPUT:DATA

OUTPUT:ESTIMATES

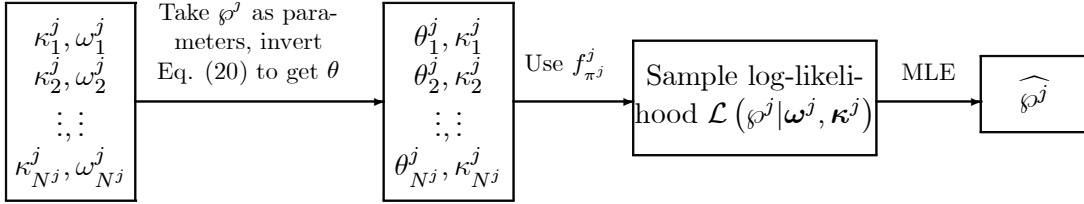


Figure 3: First Step of the Estimation Strategy

and the equilibrium wage schedule (7) can be rewritten as:

$$w^j(\theta) = \begin{cases} \rho^j \mathbb{E}^j(a|e_q) x_q^j \frac{\pi^j \theta^{\eta^j-1}}{\pi^j \theta^{\eta^j-1} + (1-\pi^j)(1-\theta)^{\eta^j-1}} & \text{if } \theta \geq \tilde{\theta}^j \\ \rho^j \mathbb{E}^j(a|e_q) \frac{\pi^j \theta^{\eta^j-1}}{\pi^j \theta^{\eta^j-1} + (1-\pi^j)(1-\theta)^{\eta^j-1}} + \rho^j \mathbb{E}^j(a|e_u) \frac{(1-\pi^j)(1-\theta)^{\eta^j-1}}{\pi^j \theta^{\eta^j-1} + (1-\pi^j)(1-\theta)^{\eta^j-1}} & \text{otherwise.} \end{cases} \quad (20)$$

For notational ease, define $\wp^j \equiv \{\pi^j, \eta^j, \tilde{\theta}^j, \rho^j \mathbb{E}^j(a|e_q) x_q^j, \rho^j \mathbb{E}^j(a|e_q), \rho^j \mathbb{E}^j(a|e_u)\}$. At this stage, we take \wp^j in wage schedule (20) as parameters. Due to the definition of $\tilde{\theta}^j$ and the strict MLRP, wage schedule $w^j(\theta)$ in (20) is strictly increasing in θ . Let us write the inverse of $w^j(\cdot)$ as $\tau^j: \mathbb{R}_+ \rightarrow [0, 1]$. Since we know that in group j , the test signal θ is distributed according to $f_{\pi^j}^j = \pi^j \eta^j \theta^{\eta^j-1} + (1-\pi^j) \eta^j (1-\theta)^{\eta^j-1}$, we can determine the theoretical distribution of $w^j(\theta)$ for fixed values of \wp . More explicitly the theoretical density distribution of wages in group j , ψ^j , is:

$$\psi^j(w) = f_{\pi^j}^j(\tau^j(w)) \frac{d\tau^j(w)}{dw} \quad (21)$$

where $d\tau^j(w)/dw$ can be computed analytically using $(dw^j(\tau^j(w))/d\theta)^{-1}$.

A data set for group j member can be represented by a pair $\{\omega^j, \kappa^j\}$, where $\omega^j = [\omega_n^j]_{n=1}^{N^j}$ is the observation of wages for group j members (without loss of generality $[\omega_n^j]_{n=1}^{N^j}$ can be ordered ascendingly); and $\kappa^j = [\kappa_n^j]_{n=1}^{N^j}$ is the population weights for each observed wage level (for example, κ_n^j denotes population weights associated with wage level ω_n^j).

To ease the computational burden of the estimation, we note first that our model predicts that the lowest wage in education group j equals $w^j(0) = \rho^j \mathbb{E}^j(a|e_u)$. Since a consistent estimate of $w^j(0)$ is simply ω_1^j , ω_1^j is also a consistent estimate of $\rho^j \mathbb{E}^j(a|e_u)$:

$$\rho^j \widehat{\mathbb{E}^j(a|e_u)} = \omega_1^j \quad (22)$$

Second, we shall address top-coding in the data. Suppose the maximum reported wage ω_N^j is the topcode level, then we will first calculate the required test signal, θ_N^j , for a group j member to earn this topcode wage by matching the right tail of the theoretical wage distribution with the fraction of observations at the topcode level. That is,

$$\pi^j \theta_N^j \eta^j + (1 - \pi^j) \left(1 - \left(1 - \theta_N^j \right)^{\eta^j} \right) = 1 - \frac{\kappa_{N^j}^j}{\sum_{n=1}^{N^j} \kappa_n^j}, \quad (23)$$

provided that there is at least one individual in group j who is earning the topcode wage. Then assuming that at least one individual in group j is employed in the complex task, we can use the information of ω_N^j and θ_N^j to get the following relationship:

$$\omega_N^j = \rho^j \mathbb{E}^j (a|e_q) x_q^j \frac{\pi^j \theta_N^j \eta^{j-1}}{\pi^j \theta_N^j \eta^{j-1} + (1 - \pi^j) \left(1 - \theta_N^j \right)^{\eta^j - 1}}. \quad (24)$$

The three relationships (19), (22) and (24) allow us in principle to write $\rho^j \mathbb{E}^j (a|e_q)$, $\rho^j \mathbb{E}^j (a|e_u)$ and $\rho^j \mathbb{E}^j (a|e_q) x_q^j$ as a function of π^j , $\tilde{\theta}^j$, η^j and/or ω_1^j and $\omega_{N^j}^j$. Thus we will be able to express the likelihood of the group j wage distribution as a function of π^j , $\tilde{\theta}^j$ and η^j .

To derive the log-likelihood of the empirical wage observations for group j , we first find the value of θ corresponding to each wage level by inverting the wage function (20). More explicitly, we calculate, for $n = 2, \dots, N^j - 1$,

$$\theta_n^j = \tau^j (\omega_n^j).$$

Second, we need to calculate the Jacobian of τ^j at every wage observations from 2 to $N^j - 1$. To do this, we will compute the derivative of $w^j(\cdot)$ at each θ_n^j as obtained above, and use the identity $d\tau^j(w)/dw = (dw^j(\tau^j(w))/d\theta)^{-1}$ to derive the Jacobian. Let us write $D_n^j \equiv d\tau^j(\omega_n^j)/dw = (dw^j(\theta_n^j)/d\theta)^{-1}$.

Now, the likelihood of group j data set (ω^j, κ^j) using observations from 2 to $N^j - 1$, is:

$$\mathcal{L}^j \left(\pi^j, \tilde{\theta}^j, \eta^j | \omega^j, \kappa^j \right) = \sum_{n=2}^{N^j-1} \left(\log \left(\pi^j \eta^j \left(\theta_n^j \right)^{\eta^j-1} + (1 - \pi^j) \eta^j \left(1 - \theta_n^j \right)^{\eta^j-1} \right) + \log D_n^j \right) \cdot \kappa_n^j \quad (25)$$

Notice that in this expression of the log-likelihood function, the three variables $(\pi^j, \tilde{\theta}^j, \eta^j)$ not only enter directly in the expression, but also appear implicitly in θ_n^j and D_n^j . The first step estimation is completed by maximizing the log-likelihood function (25) by a numerical method. Figure 3 provides a schematic summary of the first step of estimation strategy.

4.1.2 Second Step: Recovering G , H and k

The consistent estimates $\widehat{\varphi}^j = \left\{ \widehat{\pi}^j, \widehat{\eta}^j, \widehat{\theta}^j, \rho^j \widehat{E}^j(a|e_q), \rho^j \widehat{E}^j(a|e_u), \widehat{x}_q^j \right\}^{21}$ obtained in the first step generate consistent estimates of the wage offer equations w^j , and the test technologies f_q^j and f_u^j (see equations 20, 17 and 18). Write the estimated wage functions by $\widehat{w}^j(\cdot)$, and the estimated test technologies by $\widehat{f}_q^j(\cdot)$ and $\widehat{f}_u^j(\cdot)$. Now we can compute estimates of $\tilde{v}_q, \tilde{v}_u, \tilde{c}^c, \tilde{c}^h$ using the estimated wage function and test technologies. We have:

$$\widehat{v}_q = \int_0^1 \widehat{w}^c(\theta) \widehat{f}_q^c(\theta) d\theta - \int_0^1 \widehat{w}^h(\theta) \widehat{f}_q^h(\theta) d\theta \quad (26a)$$

$$\widehat{v}_u = \int_0^1 \widehat{w}^c(\theta) \widehat{f}_u^c(\theta) d\theta - \int_0^1 \widehat{w}^h(\theta) \widehat{f}_u^h(\theta) d\theta \quad (26b)$$

$$\widehat{c}^c = \int_0^1 \widehat{w}^c(\theta) \widehat{f}_q^c(\theta) d\theta - \int_0^1 \widehat{w}^c(\theta) \widehat{f}_u^c(\theta) d\theta \quad (26c)$$

$$\widehat{c}^h = \int_0^1 \widehat{w}^h(\theta) \widehat{f}_q^h(\theta) d\theta - \int_0^1 \widehat{w}^h(\theta) \widehat{f}_u^h(\theta) d\theta. \quad (26d)$$

The four thresholds computed above allow us to calculate, for any joint distributions of (a, v) , values of e.g. the size of Ω^j , $\mathbb{E}^j(a|e_q)$ etc. However, for simplicity we will make the following parametric assumptions:

$$G(c) = \beta_1 + \beta_2 c, \text{ for all } c \in [-\beta_1/\beta_2, (1 - \beta_1)/\beta_2], \quad (27)$$

$$H(v) = \beta_3 + \beta_4 v, \text{ for all } v \in [-\beta_3/\beta_4, (1 - \beta_3)/\beta_4], \quad (28)$$

$$k(a) = \beta_5 + \beta_6 a. \quad (29)$$

There are natural restrictions on the betas. For example, in order for G and H to be cumulative distributions, β_2 and β_4 must be positive, and in order to capture the idea that it is easier for an innately able agent to qualify of the complex task, β_6 must be negative. Thus the linear parameterizations are not adequate if the subsequent calculations yield inadmissible values for the betas.

We will uncover the values of betas by exploiting the restrictions provided by both the data and theory. Note first that from our data set we have a consistent estimate of the size of Ω^j – the observed group size – which we denote by $\widehat{\lambda}^j$. That is,

$$\widehat{\lambda}^j = \frac{\sum_{n=1}^{N^j} \kappa_n^j}{\sum_{n=1}^{N^c} \kappa_n^c + \sum_{n=1}^{N^h} \kappa_n^h}, j = c, h. \quad (30)$$

Yet theory (c.f. Proposition 2) yields alternative consistent estimates of the group sizes of Ω^j and the size of qualified workers π^j (see Figure 2) using the above parameterization of

²¹Carets on the variables are used to denote their estimates.

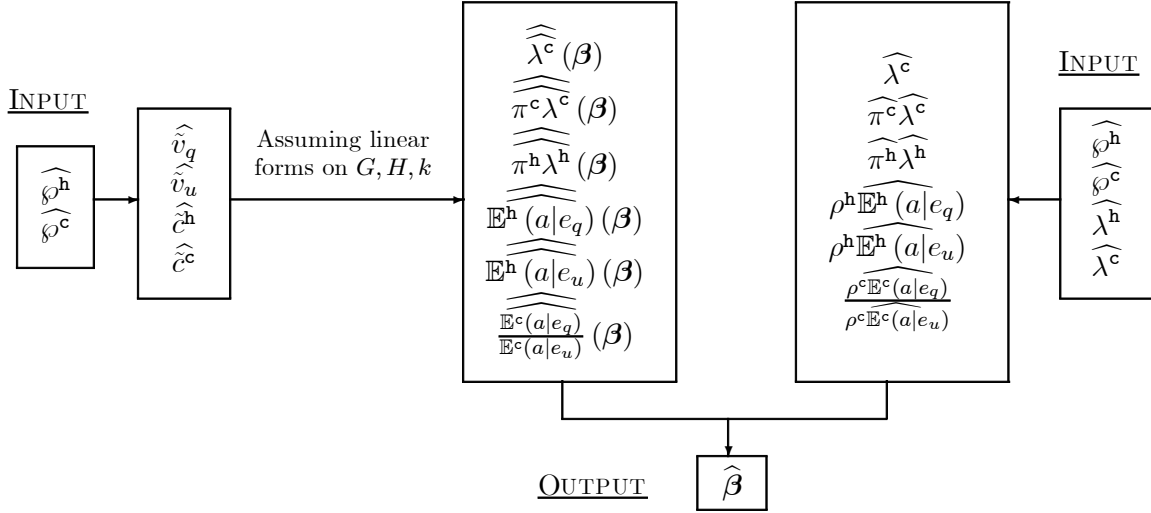


Figure 4: Second Step of the Estimation Strategy

G, H . By equating the estimates provided by theory and by data, we get three equations in the betas:

$$\widehat{\lambda}^c = \beta_3 + \beta_4 \widehat{v}_u + \beta_4 (\beta_1 + \beta_2 \widehat{c}^c) (\widehat{v}_q - \widehat{v}_u) - \frac{1}{2} \beta_2 \beta_4 (\widehat{v}_q - \widehat{v}_u)^2 \quad (31a)$$

$$\widehat{\pi}^h \widehat{\lambda}^h = (1 - \beta_3 - \beta_4 \widehat{v}_q) (\beta_1 + \beta_2 \widehat{c}^h) \quad (31b)$$

$$\widehat{\pi}^c \widehat{\lambda}^c = (\beta_3 + \beta_4 \widehat{v}_q) (\beta_1 + \beta_2 \widehat{c}^c) - \frac{1}{2} \beta_2 \beta_4 (\widehat{v}_q - \widehat{v}_u)^2. \quad (31c)$$

For example, the right hand side of equation (31a) is the size of college attendees dictated by theory, and it is equal to the direct consistent estimates of college group size $\widehat{\lambda}^c$. Similar interpretations hold for equations (31b) and (31c).²²

Three more restricting equations in the betas can be obtained by noting that Proposition 2 also allows us to get alternative consistent estimates of $\mathbb{E}^j(a|e_q)$ and $\mathbb{E}^j(a|e_u)$, exploiting our linear parametrization of the cost-ability relation (29).

Exploiting the uniform distributions assumed in (27) and (28), we obtain consistent estimates of $\mathbb{E}^j(c|e_q)$, $\mathbb{E}^j(c|e_u)$ using the characterization given by Proposition 2. We will

²²Note that $\widehat{\lambda}^h$ is not used since it is redundant with equation (31a).

denote these alternative estimates by double carets.

$$\begin{aligned}
\widehat{\mathbb{E}^h(c|e_q)} &= \frac{1}{2} \left(\widehat{c}^c - \frac{\beta_1}{\beta_2} \right) \\
\widehat{\mathbb{E}^h(c|e_u)} &= \frac{1}{(1 - \widehat{\pi}^h) \widehat{\lambda}^h} \left[\frac{1}{2} \left(\widehat{c}^c + \frac{1 - \beta_1}{\beta_2} \right) (1 - \beta_3 - \beta_4 \widehat{v}_u) (1 - \beta_1 - \beta_2 \widehat{c}^c) \right. \\
&\quad + \frac{1}{2} \beta_2 (\widehat{c}^c + \widehat{c}^h) (\widehat{c}^c - \widehat{c}^h) (1 - \beta_3 - \beta_4 \widehat{v}_q) \\
&\quad \left. + \frac{1}{2} \beta_2 \beta_4 \left(\widehat{c}^c - \frac{\widehat{c}^c - \widehat{c}^h}{3} \right) (\widehat{v}_q - \widehat{v}_u)^2 \right] \\
\widehat{\mathbb{E}^c(c|e_u)} &= \frac{1}{2} \left(\widehat{c}^c + \frac{1 - \beta_1}{\beta_2} \right) \\
\widehat{\mathbb{E}^c(c|e_q)} &= \frac{1}{\widehat{\pi}^c \widehat{\lambda}^c} \left[\frac{\widehat{c}^h - \beta_1/\beta_2}{2} (\beta_1 + \beta_2 \widehat{c}^h) (\beta_3 + \beta_4 \widehat{v}_q) \right. \\
&\quad + \frac{\widehat{c}^c + \widehat{c}^h}{2} \beta_2 (\widehat{c}^c - \widehat{c}^h) (\beta_3 + \beta_4 \widehat{v}_u) \\
&\quad \left. + \frac{1}{2} \beta_2 \beta_4 \left(\widehat{c}^h + \frac{\widehat{c}^c - \widehat{c}^h}{3} \right) (\widehat{v}_q - \widehat{v}_u)^2 \right]
\end{aligned}$$

Matching these consistent estimates for a given set of betas with their corresponding estimates based on the first stage estimates $\{\rho^j \widehat{\mathbb{E}^j}[a|e_q], \rho^j \widehat{\mathbb{E}^j}[a|e_u]\}$ and the assumed linear cost-skill relation k (using the normalization $\rho^h = 1$), we have the following three equations in betas:

$$\beta_5 + \beta_6 \rho^h \widehat{\mathbb{E}^h}(a|e_q) = \widehat{\mathbb{E}^h}(c|e_q) \quad (31d)$$

$$\beta_5 + \beta_6 \rho^h \widehat{\mathbb{E}^h}(a|e_u) = \widehat{\mathbb{E}^h}(c|e_u) \quad (31e)$$

$$\frac{\rho^c \widehat{\mathbb{E}^c}[a|e_q]}{\rho^c \widehat{\mathbb{E}^c}[a|e_u]} = \frac{\widehat{\mathbb{E}^c}(c|e_q) - \beta_5}{\widehat{\mathbb{E}^c}(c|e_u) - \beta_5} \quad (31f)$$

In the second step, we will solve the nonlinear equation system (31a)-(31f). I am unable to show that the existence and uniqueness of the solution. For this reason, we will rely on numerical methods to search for solutions (if they exist) and if we indeed find a solution, we will change the initial guess to see if the algorithm still converges to the same solution. In our estimation, solutions exist and there is a unique admissible solution.²³ The schematic summary of the second step of the estimation strategy is given in Figure 4.

²³We use Broydn's method (which is a variation of Newton's method) to solve the nonlinear system of equations.

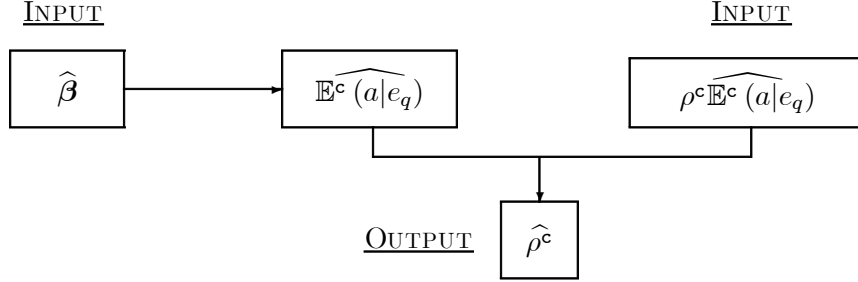


Figure 5: Third Step of the Estimation Strategy

4.1.3 Third Step: Estimating ρ^c

Once we obtain consistent estimates $\{\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3, \hat{\beta}_4, \hat{\beta}_5, \hat{\beta}_6\}$, we can generate consistent estimates of $\mathbb{E}^c(a|e_q)$ and $\mathbb{E}^c(a|e_u)$ using (14) and (15). A consistent estimate of ρ^c can be obtained by dividing the consistent estimate $\rho^c \widehat{\mathbb{E}^c(a|e_u)}$ we obtained in the first stage by the consistent estimate $\widehat{\mathbb{E}^c(a|e_u)}$. Note that equation (29) implies that

$$\widehat{\mathbb{E}^c(a|e_u)} = \frac{\widehat{\mathbb{E}^c(c|e_u)} - \hat{\beta}_5}{\hat{\beta}_6},$$

therefore a consistent estimate of ρ^c is

$$\begin{aligned} \hat{\rho}^c &= \frac{\rho^c \widehat{\mathbb{E}^c(a|e_u)}}{\widehat{\mathbb{E}^c(a|e_u)}} \\ &= \frac{\hat{\beta}_6 \rho^c \widehat{\mathbb{E}^c(a|e_u)}}{\widehat{\mathbb{E}^c(c|e_u)} - \hat{\beta}_5} \end{aligned} \quad (32)$$

A schematic summary of the third step is given in Figure 5.

This concludes the estimation of the model. If we find admissible values for the fundamental parameters of the economy, we know that the wage offer schedule estimated in the first step is part of a Nash equilibrium of the estimated economy.

4.2 Identification

From the detailed description of the estimation strategy, one can see that only the first step of the estimation procedure involves estimation. The second and third steps only involve solving non-linear equations. A formal proof of identification needs to establish, first, the log-likelihood function in (25) has a unique maximizer; second, the system of non-linear

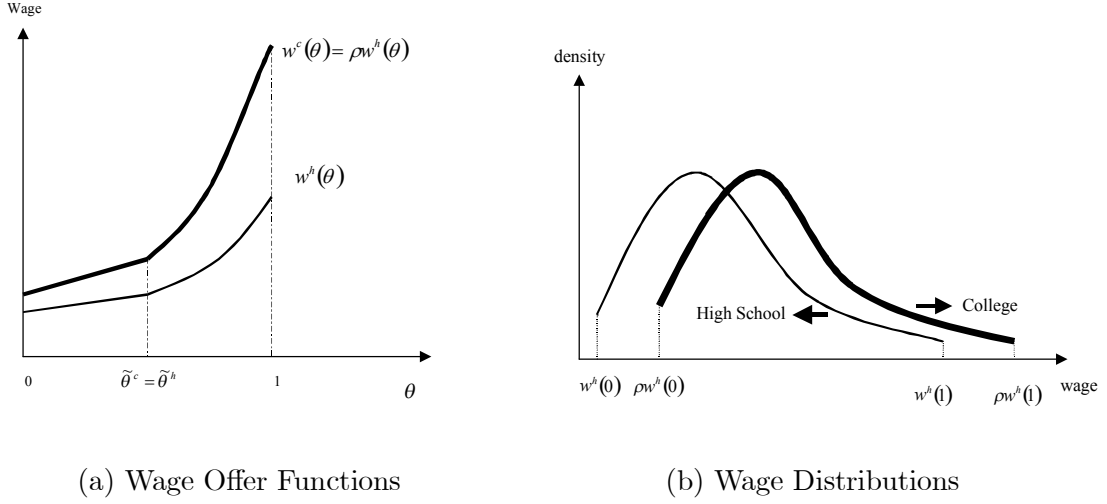


Figure 6: Pure Productivity Enhancement Case

equations (31a)-(31f) has a unique admissible solution in betas. Unfortunately, this is very hard to accomplish due to the complexity of the model. Thus the “identification” in this paper will be done only through careful numerical procedures. In the first step, obviously the parametric forms on the testing technologies (eqs. 17 and 18) play an important role; in the second step, eqs. (31a)-(31c) would not have been possible if we did not have an equilibrium model of education choice that enables us to use the information of the college enrollment rates. The assumed linear forms on G , H , and k are less important than they seem (see Section 7).

Albeit the difficulty in analytically demonstrating identification, the following thought experiments might provide some intuition about the identification. Let us consider two polar cases: the pure ability enhancing case, and the pure signaling case.

In the first thought experiment, suppose that the agents randomly chose to attend college or high school, which would imply that the firms, in offering wage schedules, will not consider the education choices a signal of ability. This hypothetical case is called the pure ability enhancing case since all the wage variations between college and high school graduates would come only from productivity enhancement. Suppose ψ^h is the p.d.f. of the high school wage distribution with support $[\underline{w}^h, \bar{w}^h]$, then the p.d.f. of the college wage distribution will be $\psi^c(w) = \psi^h(w/\rho)$ with support $[\underline{w}^c, \bar{w}^c] = [\rho\underline{w}^h, \rho\bar{w}^h]$. (See Figure 4 for an illustration.)

On the other hand, in the pure ability signaling case in which the college education were not productive enhancing (i.e., $\rho^c = 1$), then the difference in the wage distributions between college and high school graduates will result only from the ability signaling. Depending on the particular equilibrium, the differences in π^j , $\mathbb{E}^j(a|e_q)$ and $\mathbb{E}^j(a|e_u)$ between $j = c$ and $j = h$ might cause the threshold signals to be assigned to the complex task to differ for

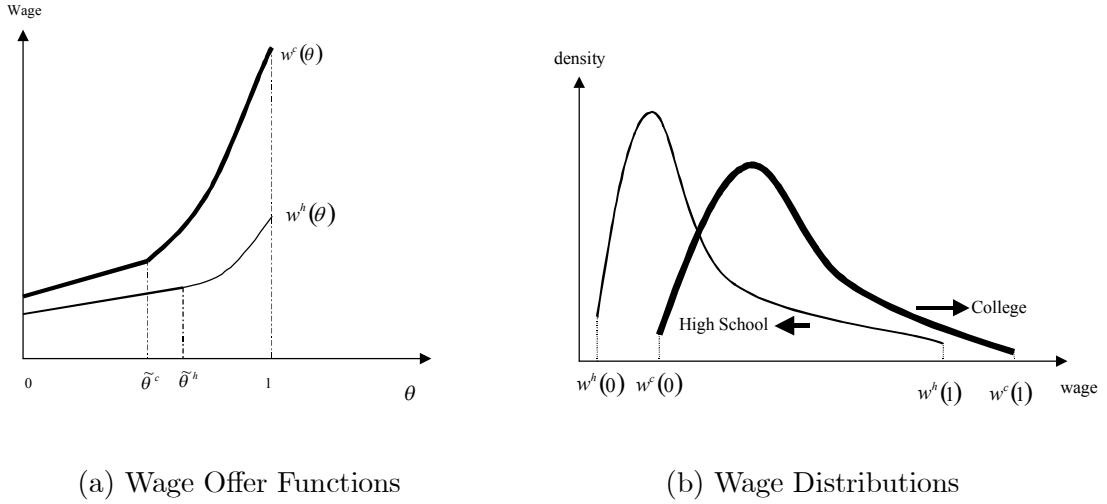


Figure 7: Pure Ability Signaling Case

the two education groups (see Figure 5a), and the difference in the wage distributions will take a different form (see Figure 5b). In our first step estimation, we use the actual wage distributions of college and high school graduates to measure the above two forces underlying the between-education-group wage distribution differences. At the same time, Proposition 2 tells us that π^j , $\mathbb{E}^j(a|e_q)$ and $\mathbb{E}^j(a|e_u)$ are completely determined by the agents' education choice and skill investment decisions, which also determines the size of the college and high school attendees. Therefore college enrollment rate plays a prominent role in our second step estimation.

Another useful piece of intuition is that the scope of ability-signalling in generating wage differentials between high school and college graduates is limited by the equilibrating forces in the model, while the scope of productivity enhancement in generating the college premium is not restricted by the equilibrating process since it is treated as an exogenous parameter.

5 The Data Set and Estimation Results

In this section we describe the data set used in the estimation and estimation results.

5.1 The Data Set

The model is estimated using 1990 U.S. Census 5% PUMS data. It includes 5% of the U.S. households who took the long-form questionnaire.

Sample Selection Criterion. We select only those white males who were within three

years of *leaving* either high school or college and were employed full time (no less than 35 hours per week) for at least 40 weeks in 1989.²⁴ We only select white males so that elements of racial and sexual discriminations do not play a direct role. We only select those who have left school within three years so that experience and tenure effects play a lesser role. We do not include self-employed individuals since they often do not have well defined wages. We only select those who worked full time for at least 40 weeks so that the differences between permanent and temporary jobs do not affect results. None of the individuals selected were in school.

We categorize those who finished high school (with or without diploma or GED) *and* those who had some college but no degree as “high school graduates”; while those with associate degree or Bachelor’s degree in college as “college graduates”. Since the 1990 Census does not have precise information of how many years of college one has, I assume for the college dropouts that they drop out in an average of two years. We assume that all individuals start primary school at the age of 6 and give one cushion year since most people graduate in May or June. Therefore “high school graduates” in our selected sample include individuals in 1989 of ages between 19 to 23, and “college graduates” include individuals of ages between 21 to 25.²⁵

We calculate weekly wages for individuals selected using the above criterion in the following way: First, we divide the wage and salary income by the total weeks worked,²⁶ second, we normalize the above calculated wage by the cost of living indices according to which state the individual works in 1989.

We observe that about 1.55% of college graduates earn weekly wages below the minimum wage in 1989 (which is around \$3.35 per hour). We assume that those are due to measurement error, and these observations are deleted. We assume that the incidences of measurement errors are the same for the two education groups, thus we delete a similar fraction of high school attendees from the bottom wage distribution.

In 1990 census the wage and salary income is topcoded by 140,000 dollars in *annual* wage and salary income. The fraction of observations above the topcode is very small in our selected sample. We decide to topcode the *weekly* wages for college graduates at \$1,000 and the high school graduates at \$718.5 so that about 1% of each education group is topcoded.

The above selection and calculation leave us respectively 18,567 *different* wage observations for college graduates education group and 42,667 for high school graduates. Note that each observation in the 1990 Census 5% PUMS has a person’s weight to adjust for the stratified sampling process to make it representative, the total weights in our selected

²⁴These restrictions reduce the number of observations, which is the reason we did not use the smaller but annually available March Supplement to the CPS data.

²⁵It would be more appropriate if we use wages of high school and college attendees from the same *age* cohort. The problem is that the census is held decennially. We could have compared the 1989 wage of college attendees with the 1989 wages of those who left high school in 1985, but then we would have to deal with the experience issue.

²⁶There are in the data variations in hours worked per week. We neglect this variation, and simply assume that all the selected individuals work the same amount of hours per week.

sample are 411,326 for the college group and 858,419 for the high school group.

5.2 The Estimation Results

I. First Step Estimation Results

The first step estimates using the 1989 wage distributions of high school and college attendees are given in Table 1. We use the outer-product-of-the-gradient (OPG estimator) estimator to calculate the standard deviations for the estimates $\hat{\theta}^j$, $\hat{\pi}^j$ and $\hat{\eta}^j$ (see Davidson and MacKinnon, 1993, for a discussion of OPG estimator). The standard deviations for the other variables are simulated using the estimated variance-covariance matrix of $\hat{\theta}^j$, $\hat{\pi}^j$ and $\hat{\eta}^j$.²⁷

Table 1: Estimates of φ^j : The First Step

Variables	College	High School
π^j	.315 (1.0E-03)	.227 (.167E-02)
$\tilde{\theta}^j$.877 (.1E-06)	.915 (.57E-02)
η^j	1.55 (.18E-02)	1.57 (.11E-02)
$\rho^j \mathbb{E}^j(a e_q)$	1,139.45 (8.94E-01)	920.37 (4.25)
$\rho^j \mathbb{E}^j(a e_q) x_q^j$	1,237.4 (1.21)	1006.4 (2.14)
$\rho^j \mathbb{E}^j(a e_u)$	133.9	98.76

We estimated that in 1989 about 31.5% of college graduates invested in the skills that qualify them for the complex task, while that fraction is about 22.7% among high school graduates. The test technology parameters η^c and η^h are similar (1.55 for college, and 1.57 for high school), although a formal t -test would reject the hypothesis that they are equal. Conditional on being qualified, an average college graduate has efficiency units 1,139.45 (including the productivity enhancement due to college education), while an average high school graduate has efficiency units 920.37. Conditional on being unqualified, an average college attendee has efficiency units 133.9, while a high school attendee has 98.05.

²⁷The simulated standard errors for derived estimates are obtained by drawing a random sample of 1,000 of $\tilde{\theta}^j$, π^j , and η^j according to their asymptotic variance-covariance matrix. For every simulated sample, we numerically re-calculate the relevant estimates, including re-solving the system of non-linear equations (31a)-(31f).

Table 2: Some Other Estimates

Variables	Estimates	Simulated Stand. Errors
λ^c	.324	...
λ^h	.676	...
\tilde{v}_q	182.95	0.85
\tilde{v}_u	146.48	0.63
\tilde{c}^c	143.35	9.8E-02
\tilde{c}^h	106.88	1.87E-01
x_q^c	1.086	2.1E-04
x_q^h	1.095	3.98E-03

In Table 2, we present the estimates of the education group size in our sample, as well as some estimates derived from the first step estimates. These values will be used in the second step estimation. Here we note that the estimate of the group size of the college graduates in our sample at 32.4%. This is in line with the college graduation rate reported elsewhere (e.g. U.S. Department of Education, 1997), which estimates the college graduation rate, including male and female, at about 30% around 1990. This shows that our sample selection criterion does not distort the general education pattern in the U.S. among young men. We discuss other estimates in item III below.

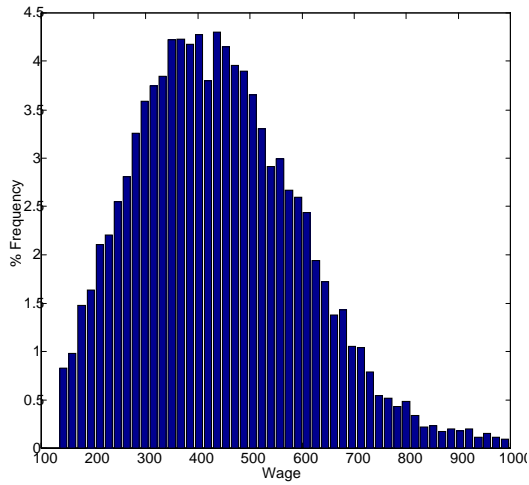
II. Evaluation of the First Step Estimation

It is important to evaluate how the first step estimates replicate the actual wage distributions of the college and high school graduates. For this purpose, I simulate 10,000 wages according to the estimated wage offer schedules and the estimated test technologies for each education group. The histograms of the simulated wages and the actual wages conditional on the wages being lower than the topcoded level are presented in Figure 8.

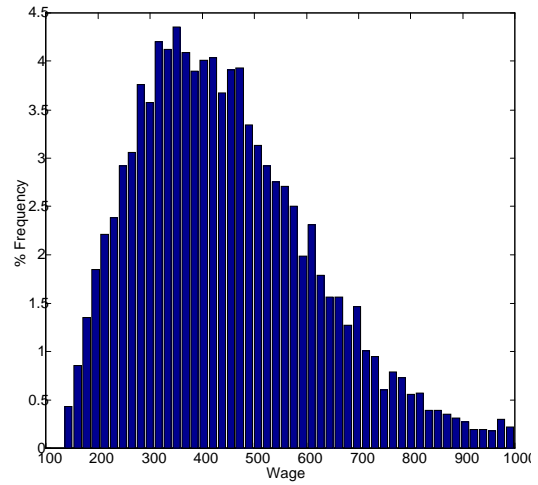
Table 3: Summary Statistics for Simulated and Actual Wage Distributions

		Mean	Std. Dev.	Skewness	Kurtosis
College	Empirical	440.78	158.40	.50	3.08
	Estimated	446.20	169.3	.66	3.11
High School	Empirical	279.07	112.44	1.02	3.98
	Estimated	280.03	114.96	0.99	3.90

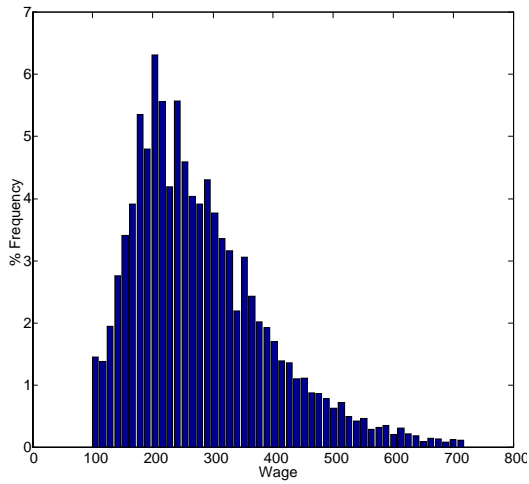
At least on the visual basis, the simulated wage distributions using our estimates seem to capture the shape of the actual wage distributions rather well. The first four centered



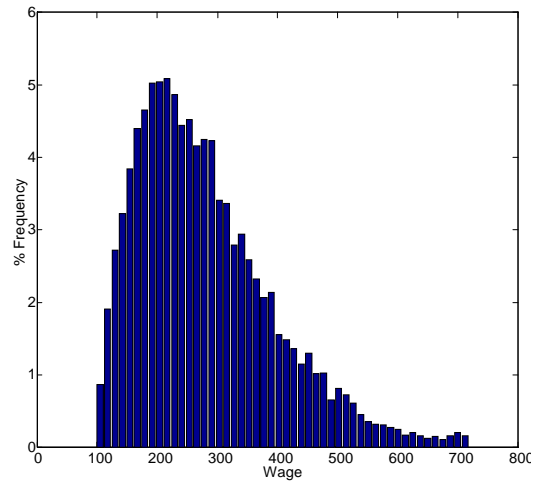
(a) Actual College Wage Distribution



(b) Simulated College Wage Distribution



(c) Actual H.S. Wage Distribution



(d) Simulated H.S. Wage Distribution

Figure 8: Simulated and Actual Wage Distributions for H.S. and College Graduates in 1989

moments for the simulated and actual wage distributions are presented in Table 3. All statistics are conditional on the wage being lower than the topcoded level. These evidences seem to show that the model performs fairly well in representing the wage distributions.

III. Second and Third Step Estimation Results

The first step estimates implies (via eqs. 10 through 12) that the thresholds for the equilibrium membership sets are: $\tilde{v}_q = 182.95$, $\tilde{v}_u = 146.48$, $\tilde{c}^c = 143.35$ and $\tilde{c}^h = 106.88$. (See Table 2 for simulated standard errors of those estimates.) Using these values, the second and third step estimates are presented in Table 4. The estimates of betas imply that the utility cost of attending college v is distributed uniformly on the support $[35.8, 406.3]$. The cost of skill investment is uniformly distributed on the support $[-2924.3, 8994.6]$.²⁸ Also, a one unit increase in ability decreases the skill investment cost by about 7.25 units. We estimate ρ^c to be about 1.40 with a standard error of 4.77E-03. That is, attending college augments an agent's innate ability by about 40%.

Table 4: Estimates: The Second and Third Steps

Parameters	Estimates	Simulated Stand. Errors
β_1	.245353	1.126E-03
β_2	0.839E-04	1.17E-05
β_3	-9.65E-02	7.05E-03
β_4	2.70E-03	4.82E-05
β_5	5263.84	683.21
β_6	-7.25	0.93
ρ^c	1.3991	4.77E-03

IV. Estimated College Wage Premium

We can use the estimated wage offer schedules for college and high school attendees, \widehat{w}^j , together with the estimated test technologies, \widehat{f}_q^j and \widehat{f}_u^j , to estimate the college wage premium as:

$$\widehat{\text{Premium}} = \frac{\int_0^1 \widehat{w}^c(\theta) \left[\widehat{\pi}^c \widehat{f}_q^c(\theta) + (1 - \widehat{\pi}^c) \widehat{f}_u^c(\theta) \right] d\theta - \int_0^1 \widehat{w}^h(\theta) \left[\widehat{\pi}^h \widehat{f}_q^h(\theta) + (1 - \widehat{\pi}^h) \widehat{f}_u^h(\theta) \right] d\theta}{\int_0^1 \widehat{w}^h(\theta) \left[\widehat{\pi}^h \widehat{f}_q^h(\theta) + (1 - \widehat{\pi}^h) \widehat{f}_u^h(\theta) \right] d\theta},$$

²⁸Note that these numbers should be viewed with the understanding that we are measuring all the benefit of attending school in terms of weekly wages. Thus the above numbers represent roughly two or four years' cost or benefit collapsed into a single week.

where the denominator is the average wage of a high school attendee, and the numerator is the difference in average wage between a college and high school attendee. We compute the above value to be about 58.64% with a simulated standard error of 3.71E-03. Computing the college wage premium directly from the data yields approximately the same number. This estimate of college wage premium is consistent with that in Murphy and Welch (1989) and U.S. Department of Education (1997).

6 Disentangling the College Wage Premium: An Experiment

In this section, we consider how much of the estimated 58.64% college wage premium should be attributed to the productivity enhancement of college education. To this end, I conduct the following counter-factual experiment: Suppose that we are in an economy that is the same as the economy we estimated above except that college education is not productive, i.e., $\rho = 1$. What would be the maximum college premium in the hypothetical economy?

The equilibria of the hypothetical economy are numerically solved. The summary statistics of the two equilibria of the hypothetical economy are presented in Table 5.

Table 5: Comparison of Statistics of the Hypothetical and Actual Economies

	Actual Economy	Hypothetical Economy	
	Observed Eq.	Collegeless Eq.	High Premium Eq.
Output	339.7	306.48	308.8
College Utility Cost	31.4	0.	2.33
College Enrollment Rate	32.4%	0.%	5.13%
Qualified HS Attendees	22.68%	24.28%	25.08%
Qualified College Attendees	31.51%	...	32.80%
Average College Wage	452.90	...	369.45
Average High School Wage	285.49	306.48	305.49
College Wage Premium	58.64%	...	20.93%

Since there is a positive utility cost of attending college (recall the support of v is [35.8, 406.3]), the hypothetical economy admits another equilibrium in which no one attends college even though the concept of college exists. This equilibrium is dubbed *collegeless equilibrium* in Table 5. However, even though college education did not enhance productivity, the hypothetical economy does admit an equilibrium in which about 5.13% of the population attend college. This equilibrium is dubbed the *high premium equilibrium* in Table 5. The college wage premium in the high premium equilibrium is about 20.93%. Since a hypothetical economy with $\rho = 1$ could sustain at most a 20.93% college wage premium, we conclude that at least 37.71% ($= 58.64\% - 20.93\%$) wage difference between college and

high school attendees in 1989 should be attributed to the fact that $\rho = 1.40$. That is, *at least* 64.31% ($= 37.71/58.64$) of the college premium is due to productivity enhancement.

In the high premium equilibrium of the hypothetical economy, the *output* is about 9% lower than the actual economy, while *welfare* is only about 0.3% lower. The reason is that in the actual economy although output is higher but there are more people who incur the cost of attending the college.

Sometimes we encounter the argument that high school graduates are earning less because colleges attract a pool of agents with higher average ability. The counter-factual experiment allows us to evaluate the extent to which the low average wage of high school graduates is due to the existence of college. The wage high school graduates would be able to earn in a world without colleges will be the same as the average high school wage in the collegeless equilibrium of the hypothetical economy, which is about 306.48 dollars. This is only about 7.35% higher than the average wage of high school attendees in the actual economy.

An alternative experiment is to consider a hypothetical economy in which a worker's innate ability and his qualification for the complex task are both perfectly observable to firms.²⁹ In this economy, the resulting college wage premium is due only to the productivity enhancement of college education. The results from this counter-factual are described in Appendix B. It is worth noting that the equilibrium structure of this complete information economy is vastly different from the actual economy. In particular, firms will not hire workers whose productivity is negative. Therefore, a non-negligible fraction of high school students would have been unemployed.

7 Robustness

In this section, we discuss the robustness of the estimation methodology and the estimation results.

I. Robustness of Estimation Methodology

A few parametric assumptions can be relaxed in order for the proposed estimation methodology in Section 4 to work. First, we assumed that the worker's preference takes the form of $w - v(j) - c(e, a)$. In fact, an appropriate modification of the estimation procedure can be applied to any utility function of the form $w - \varphi(v(j), c(e, a))$ with known function form of φ . Note that when alternative functional forms of φ is used, the characterization of education and skill investment choices (Proposition 2) will have to be modified. Second, we assumed in the paper that in the population (a, v) are independently distributed. But this is only for simplicity. If instead (a, v) has a joint distribution Γ , which are characterized by no more than four parameters, then our proposed estimation strategy can still be applied. The key idea is to note that in the characterization of the education and skill investment choices (Proposition 2), no independence assumption is needed. Hence, in principle we can for any

²⁹I thank Stephen Morris for suggesting this counter-factual experiment.

To figure out the college enrollment rates and the expected ability for each membership sets as we did in Subsection 4.1.2. The only difference is that with more general forms of Γ , we will be unable to write out analytically the six non-linear equations, thus we will have to use numerical integration methods in deriving the set of non-linear equations. Third, in the paper we assumed that attending college augments a worker's efficiency units from a to ρa where ρ is a constant, namely, we assumed uniformity in ability enhancement. However, our estimation procedure could deal with the scenario that ρ is a particular function of a , say $\rho(a) = zq(a)$ where z is a constant to be estimated, and $q(\cdot)$ is a known function. Fourth, we assumed in the paper that one efficiency unit of a qualified worker on the complex task is x_q^j , a constant independent of the worker's actual ability. The estimation procedure, properly modified, can deal with the case that x_q^j depends on a in the following fashion: $x_q^j(a) = \tilde{z}\tilde{q}(a)$ where \tilde{z} is a constant to be estimated, and $\tilde{q}(\cdot)$ is a known function.

II. Robustness of Estimation Results

In the results presented, we selected individuals who were within three years of leaving high school or college. It is well known that experiences are a crucial determinant of wages, and one might wonder how one can disentangle the college wage premium for workers with longer working experience. For this purpose we select from 1990 Census 5% PUMS data those workers who were at their prime working age, namely those with 21-24 years of working experiences. We find that the average college wage for this old cohort is about 907.6 dollars per week, while the average high school wage is about 673.8 dollars per week. Therefore, the college wage premium for the older cohort is around 34.7% in 1989. This is consistent with Murphy and Welch (1989) who documented that around mid-1980s, the college wage premium for workers with 1-5 years of experiences exceeded that for workers with more experiences. Using the same estimation procedure, we estimated that for the old cohort the college education enhanced that productivity by about 22.6%, i.e. $\rho = 1.226$ for the old cohort. When we conduct the same counter-factual experiment as in Section 6, we find that in a hypothetical economy in which college education did not enhance productivity, the maximum college wage premium sustainable in equilibrium is about 19.5%, i.e., ability signaling can account for about 56.2% of the college wage premium among prime age workers in 1989. The fact that our estimated productivity enhancement factor for this cohort, who graduated from college around 1964 to 1968, is also consistent with the evidence that college wage premium was low in the 1970s (see Murphy and Welch 1989).

8 Conclusion

We *quantitatively* evaluated the importance of productivity enhancement and ability signaling in explaining the college wage premium using an equilibrium model of education choice. We interpret that the extent to which attending college signals ability is determined by the average ability of college attendees. Our characterization of the individuals' education choices and skill investment decisions (Proposition 2) allows us to use the observed college enrollment rate in an informative way.

Our structural estimation yields the following results: (1) We find that college education enhances attendees' productivity by 40%; (2) If college education did not enhance productivity at all, then (i) *at most* 5.13% of the population would attend college in contrast to 32.4% in the data; (ii) the college wage premium would be *at most* 20.93% in contrast to about 58.64% in the data, from which we conclude that productivity enhancement accounts for *at least* 64.31% of the college premium observed in 1989; (iii) output would be about 9% lower while welfare would only be about 0.3% lower than the actual economy.

Finally, some remarks about our model of education are in order. Our model is stylized and some important features of education and labor markets are ignored. For example, we assumed that whether an individual attends college is a one-sided decision problem. In reality colleges also impose admission standards, though they are not solely based on ability. We assumed that the decision of whether to attend college is made upon junior high school graduation. In reality college attendance decisions are more likely made sequentially. We assumed that in the labor market employers only observe a worker's credential and some noisy signal, in reality the employers might observe more information that can change a worker's obtainable wage. Also we took in this paper the narrow view that the benefit of education is its effect on wages in the labor market. In reality, there are more private and social benefits of education: for example, students are taught civics, or art, or music to enrich their lives and make them better citizens. The possibility of the labor market learning a worker's qualification over time is also ignored. To incorporate it, we would need to address the adverse selection problem in the rehiring labor market.

Appendix A: Proofs

Proof of Proposition 1: We first prove two Lemmas:

Lemma 1 *Suppose $\langle \mathbf{w}_i, t_i \rangle_{i=1,2}$ is a pair of best responses, then for each j , $w_1^j(\theta) = w_2^j(\theta)$ for almost all $\theta \in [0, 1]$.*

Proof. Suppose to the contrary that for some j there is a positive measure set $\Theta^j \subseteq [0, 1]$ such that $w_i^j(\theta) > w_{i'}^j(\theta)$ for all $\theta \in \Theta^j$. Then the alternative strategy $\langle \hat{\mathbf{w}}_i, t'_i \rangle : t'_i(\theta) = t_i(\theta)$ for all $\theta \in [0, 1]$, $\hat{w}_i^j(\theta) = w_i^j(\theta)$ for all $\theta \in [0, 1] \setminus \Theta^j$ and $\hat{w}_i^j(\theta) = (w_i^j(\theta) + w_{i'}^j(\theta)) / 2$ for $\theta \in \Theta^j$, and $\hat{w}_i^{j'}(\theta) = w_i^{j'}(\theta)$ is a profitable deviation for firm i , a contradiction. ■

Lemma 2 *If $t_i : [0, 1] \times \{\mathbf{c}, \mathbf{h}\} \rightarrow \{0, 1\}$ is the task assignment rule on the equilibrium path for firm $i = 1, 2$, then for each j there exists some $\tilde{\theta}_i^j \in [0, 1]$ such that $t_i(\theta, j) = 1$ for almost all $\theta > \tilde{\theta}_i^j$ and $t_i(\theta, j) = 0$ for almost all $\theta < \tilde{\theta}_i^j$, and for $i = 1, 2$. Furthermore, $\tilde{\theta}_1^j = \tilde{\theta}_2^j = \tilde{\theta}^j$ where $\tilde{\theta}^j$ is defined in equation (6).*

Proof. First we prove the cut-off property. Suppose that the cut-off property does not hold for some firm i for some education group j . That is, suppose there are positive measure sets $\Theta_{ih}^j, \Theta_{il}^j \subseteq [0, 1]$ such that $\theta_h > \theta_l$ for all $(\theta_h, \theta_l) \in \Theta_{ih}^j \times \Theta_{il}^j$, $t_i(\theta_h) = 0$ for

all $\theta_h \in \Theta_{ih}^j$ and $t_i(\theta_l) = 1$ for all $\theta_l \in \Theta_{il}^j$. Write $f_{\pi^j}(\theta) = \pi^j f_q(\theta) + (1 - \pi^j) f_u(\theta)$, and $\bar{\theta}_l = \sup \Theta_{il}^j$. We can w.l.o.g. assume that $\int_{\Theta_{il}^j} f_{\pi^j}(\theta) d\theta = \int_{\Theta_{ih}^j} f_{\pi^j}(\theta) d\theta$. Write

$$z_q^j(\theta|J, e) = \pi^j \rho^j \mathbb{E}^j(a|e_q) x_q f_q(\theta) / f_{\pi^j}(\theta)$$

as the expected productivity of a worker from education group j with a test signal θ if he is assigned to the complex task; and

$$z_u^j(\theta|J, e) = [\pi^j \rho^j \mathbb{E}^j(a|e_q) f_q(\theta) + (1 - \pi^j) \rho^j \mathbb{E}^j(a|e_u) f_u(\theta)] / f_{\pi^j}(\theta)$$

as his expected productivity if assigned to the simple task. Further, define

$$z^j(\theta|J, e) = z_q^j(\theta|J, e) - z_u^j(\theta|J, e).$$

Due to the MLRP assumption, it is straightforward to verify that since $x_q > 1$, z^j is strictly increasing in θ . Now consider an alternative task assignment rule t'_i , where $t'_i(\theta, j) = 1$ for all $\theta \in \Theta_{ih}^j$, $t'_i(\theta, j) = 0$ for all $\theta \in \Theta_{il}^j$, $t'_i(\theta, j) = t_i(\theta, j)$ for all other $\theta \in [0, 1]$; and $t'_i(\theta, j') = t_i(\theta, j')$ for all $\theta \in [0, 1]$. The difference in profits between t_i and t'_i is

$$\begin{aligned} & \int_{\theta \in \Theta_{ih}^j} z^j(\theta|J, e) f_{\pi^j}(\theta) d\theta - \int_{\theta \in \Theta_{il}^j} z^j(\theta|J, e) f_{\pi^j}(\theta) d\theta \\ & > \int_{\theta \in \Theta_{ih}^j} [z^j(\theta|J, e) - z^j(\bar{\theta}_l|J, e)] f_{\pi^j}(\theta) d\theta > 0, \end{aligned}$$

which is a contradiction to t_i being an equilibrium task assignment rule of firm i .

Moreover, Lemma 1 and workers' equilibrium offer acceptance rule (3) imply that $\Theta_i^j = \Theta_{i'}^j$. Hence the equilibrium task assignment rules for the two firms are the same, i.e., $\tilde{\theta}_i^j = \tilde{\theta}_{i'}^j$. By definition of $\tilde{\theta}^j$, the task assignment rule t with $\tilde{\theta}^j$ as the critical point strictly increases profit over any other thresholds. Hence $\tilde{\theta}_1^j = \tilde{\theta}_2^j = \tilde{\theta}^j$. ■

(*Proof of Proposition 1*) Proposition 1 can be restated as: Given J, e , the equilibrium wage offer schedules are $w^j(\theta|J, e) = \max \{z_q^j(\theta|J, e), z_u^j(\theta|J, e)\}$. Lemma 1 establishes that for each j , $w_1^j(\theta) = w_2^j(\theta) = w^j(\theta)$ almost everywhere. Suppose to the contrary that for some j , $w^j(\theta) < \max \{z_q^j(\theta|J, e), z_u^j(\theta|J, e)\}$ for a positive measure set $\tilde{\Theta}^j \subseteq [0, 1]$. Consider an alternative strategy $\hat{\mathbf{w}}_1$ for firm 1 where $\hat{w}_1^{j'}(\theta) = w^{j'}(\theta)$, and $\hat{w}_1^j(\theta) = w^j(\theta) + \varepsilon$ for $\theta \in \tilde{\Theta}^j$ for some $\varepsilon > 0$ while $\hat{w}_1^j(\theta) = w^j(\theta)$ for $\theta \in [0, 1] \setminus \tilde{\Theta}^j$. All workers from education group j whose test signal $\theta \in \tilde{\Theta}^j$ will then accept firm 1's offer. The difference in profits for firm 1 between $\langle \hat{\mathbf{w}}_1, t_1 \rangle$ and $\langle \mathbf{w}_1, t_1 \rangle$ is:

$$\int_{\theta \in \tilde{\Theta}^j} \left\{ \frac{1}{2} [\max \{z_q^j(\theta|J, e), z_u^j(\theta|J, e)\} - w^j(\theta)] - \varepsilon \right\} f_{\pi^j}(\theta) d\theta,$$

which is strictly positive if ε is sufficiently small. Similar arguments can establish that profitable deviation exists if $w^j(\theta) > \max \{z_q^j(\theta|J, e), z_u^j(\theta|J, e)\}$. *Q.E.D.*

Proof of Proposition 2: We proceed via a series of lemmas:

Lemma 3 *In a Nash equilibrium, if $(\hat{c}, \hat{v}) \in \Omega_q^h$, then $\{(c, v) : c \leq \hat{c}, v \geq \hat{v}\} \subseteq \Omega_q^h$; if $(\hat{c}, \hat{v}) \in \Omega_q^c$, then $\{(c, v) : c \leq \hat{c}, v \leq \hat{v}\} \subseteq \Omega_q^c$; if $(\hat{c}, \hat{v}) \in \Omega_u^h$, then $\{(c, v) : c \geq \hat{c}, v \geq \hat{v}\} \subseteq \Omega_u^h$; if $(\hat{c}, \hat{v}) \in \Omega_u^c$, then $\{(c, v) : c \geq \hat{c}, v \leq \hat{v}\} \subseteq \Omega_u^c$.*

Proof. If $(\hat{c}, \hat{v}) \in \Omega_q^h$, then by revealed preference, the following three inequalities hold: $V_q^h - \hat{c} \geq V_u^h$, $V_q^h - \hat{c} \geq V_q^c - \hat{v} - \hat{c}$, and $V_q^h - \hat{c} \geq V_u^h - \hat{v}$. This implies that for any $c \leq \hat{c}, v \geq \hat{v}$, the following three inequalities hold: $V_q^h - c \geq V_u^h$, $V_q^h - c \geq V_q^c - v - c$, and $V_q^h - c \geq V_u^h - v$. Hence $\{(c, v) : c \leq \hat{c}, v \geq \hat{v}\} \subseteq \Omega_q^h$. Similar proof applies to the other cases. ■

Definition 3 *For $j \in \{c, h\}$, define $c^{j*} = \sup_c \Omega_q^j$, $v_q^* = \sup_v \Omega_q^c$, and $v_u^* = \sup_v \Omega_u^c$.*

Lemma 4 *In any Nash equilibrium, $v_q^* = \tilde{v}_q$, $v_u^* = \tilde{v}_u$ and $c^{j*} = \tilde{c}^j$, for $j \in \{c, h\}$.*

Proof. Consider a worker of type (c, v_q^*) with c sufficiently small (possibly negative). He must be indifferent between being in Ω_q^c and Ω_q^h . Suppose w.l.o.g. that he strictly prefers to be in Ω_q^c , then $V_q^c - v_q^* > V_q^h$. But this implies that we can find $v' > v_q^*$ such that $(c, v') \in \Omega_q^c$, a contradiction to $v_q^* = \sup_v \Omega_q^c$. Hence $v_q^* = V_q^B - V = \tilde{v}_q$. The other claims are similarly proved. ■

Definition 4 *The border of Ω_u^h and Ω_q^c is defined to be $\{(c, v) \in \Omega_u^h \cup \Omega_q^c : V_q^c - c - v = V_u^h\}$. Note that $(\tilde{c}^h, \tilde{v}_q)$ and $(\tilde{c}^c, \tilde{v}_u)$ are on the border of Ω_u^h and Ω_q^c .*

Lemma 5 *In a Nash equilibrium with $\tilde{v}_q > \tilde{v}_u$, the border between Ω_u^h and Ω_q^c is a straight-line connecting $(\tilde{c}^h, \tilde{v}_q)$ and $(\tilde{c}^c, \tilde{v}_u)$.*

Proof. Suppose these two points are not connected by a straight line, then there must be a subset of Ω_u^h below the 45° line or a subset of Ω_q^c above it or both. Suppose there is a subset $\hat{\Omega} \subset \Omega_u^h$ below the straight line. Pick any $(\hat{c}, \hat{v}) \in \hat{\Omega}$. $(\hat{c}, \hat{v}) \in \hat{\Omega}$ implies that $\hat{c} - \tilde{c}^h < \tilde{v}_q - \hat{v}$. But a worker of type $(\tilde{c}^h, \tilde{v}_q)$ is indifferent between being in Ω_u^h and Ω_q^c , i.e., $V_u^h = V_q^c - \tilde{c}^h - \tilde{v}_q$. This implies that $V_u^h < V_q^c - \hat{c} - \hat{v}$, a contradiction to $(\hat{c}, \hat{v}) \in \hat{\Omega} \subset \Omega_u^h$. The same argument can rule out the case that there is a subset of Ω_q^c above the straight line. ■

Proposition 2 is a direct consequence of Lemmas 3-5. *Q.E.D.*

Appendix B: An Alternative Counter-Factual Experiment

In this Appendix, we consider a hypothetical economy in which a worker's innate ability and his qualification for the complex task are both perfectly observable to firms. To calculate the college wage premium in such a hypothetical economy, we need to characterize the agents' education and skill investment choices. A worker of type (a, v) will make education and skill investment choices to solve:

$$\max \{a\rho^c x_q^c - (\beta_5 + \beta_6 a) - v, a\rho^c - v, ax_q^h - (\beta_5 + \beta_6 a), a\},$$

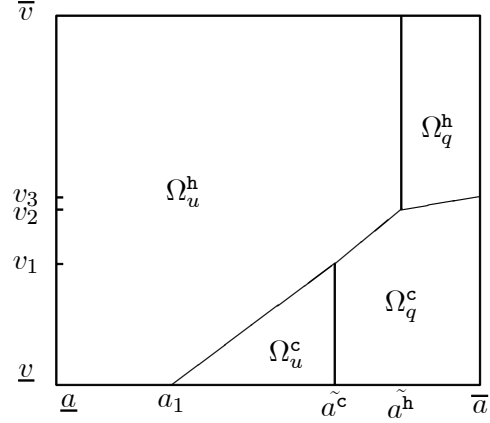


Figure 9: Education and Skill Investment Choices in the Complete Information Economy

where the first term is his benefit from being in group Ω_q^c , the second term his benefit from group Ω_u^c , the third term his benefit from Ω_q^h , and the last term his benefit from Ω_u^h . It is very easy to verify that the equilibrium education and skill investment choices in this hypothetical economy can be characterized as:

$$(a, v) \in \begin{cases} \Omega_q^h & \text{if } a > \tilde{a}^h, v > a(\rho^c x_q^c - x_q^h) \\ \Omega_q^c & \text{if } a \in (\tilde{a}^c, \tilde{a}^h), v < a(\rho^c x_q^c - 1 - \beta_6) - \beta_5 \\ & \text{or } a > \tilde{a}^h, v < a(\rho^c x_q^c - x_q^h) \\ \Omega_u^c & \text{if } a < \tilde{a}^c \text{ and } v < a(\rho^c - 1) \\ \Omega_u^h & \text{otherwise} \end{cases}, \quad (33)$$

where $\tilde{a}^h \equiv \beta_5 / (x_q^h - 1 - \beta_6)$, and $\tilde{a}^c \equiv \beta_5 / (\rho^c (x_q^c - 1) - \beta_6)$. From our estimates presented in Table 4, we know that $\tilde{a}^h > \tilde{a}^c$. The membership sets for the hypothetical economy are described as in Figure 9. According to our estimates presented in Table 1 and 4, the parameter values in Figure 9 are given in Table 6:

Table 6: Parameter Values in Figure 9

Parameter	Value	Parameter	Value
\underline{a}	-514.58	\underline{v}	35.8
\bar{a}	1129.4	\bar{v}	406.3
a_1	89.5	v_1	285.62
\tilde{a}^c	714.19	v_2	304.37
\tilde{a}^h	715.78	v_3	306.76

Table 7: Some Statistics in the Hypothetical Economy with Complete Information

Variable	Value
Average College Wage	996.11
Average High School Wage:	
Unconditional	271.60
Conditional on Employment	503.54
College Enrollment Rate	31.12%
Qualified College Attendees	58.83%
Qualified High School Attendees	10.19%
High School Unemployment Rate	46.06%

Note, however, firms will only make non-negative wage offers to workers with non-negative productivity. Hence high school workers with negative productivity will be unemployed. One needs to take this into account when calculating the average wages for high school and college graduates in this hypothetical economy. Some statistics in the equilibrium of this complete information hypothetical economy are presented in Table 7. When calculating the unconditional average wage of the high school attendees, we take the wages of the unemployed high school attendees to be zero. There are a few notable features: First, The average wage of college graduates in this complete information economy is much higher than the actual economy. This is so since each worker is paid according to his own productivity, and the free riding of the unqualified college graduates on the qualified ones are no present any more. Second, for the same reason, the average wage of high school graduates conditional on being employed is also much higher than that in the actual economy. However, the unconditional high school average wage remain about the same as in the actual economy. Third, the college enrollment rate is about the same as in the actual economy. There are two things that can change the college enrollment rate in this complete information economy: on the one hand, no low ability workers can free ride on high ability ones, which will tend to decrease the enrollment rate; on the other hand, some high ability but with high college attendance cost agents will now find it worth while to attend college, which will tend to increase the enrollment rate. It turns out that two forces almost cancel out. If we use the average high school wage conditional on employment, the college wage premium in this economy is about 97.82% ($= (996.13 - 503.54)/503.54$); while if we use the unconditional average high school wage, then the college wage premium is about 226.8% ($= (996.13 - 271.6)/271.6$).

Does the above results mean that ability signaling plays no role in the observed college wage premium in 1989? My answer is no since the complete information hypothetical economy is drastically different from the actual economy. The counter-factual we conducted in Section 6 is much closer in spirit to the actual economy.

References

- [1] Albrecht, James, “A Procedure for Testing the Signalling Hypothesis”, *Journal of Public Economics*, 1981, Vol 15, 123-132.
- [2] Angrist, Joshua and Alan Krueger, “Does Compulsory School Attendance Affect Schooling and Earnings?”, *Quarterly Journal of Economics*, November 1991, **106**, 979-1014.
- [3] Ashenfelter, Orley and Alan Krueger, “Estimates of the Economic Returns to Schooling from a New Sample of Twins”, *American Economic Review*, **No. 5, Vol 84**, December 1994.
- [4] Becker, Gary, *Human Capital*, New York: Columbia University Press for the National Bureau of Economic Research, 1964.
- [5] Behrman, Jere R. and Paul Taubman, “Intergenerational Transmission of Income and Wealth”, *American Economic Review*, **No 3, Vol 66**, 436-440, May 1976.
- [6] Davidson, Russell and James MacKinnon (1993), *Estimation and Inference in Econometrics*, New York, Oxford University Press.
- [7] Fang, Hanming, “Discrimination with Endogenous Group Choices”, mimeo, June 1998, University of Pennsylvania.
- [8] Mincer, Jacob, *Schooling, Experience and Earnings*, New York: Columbia University Press, 1974.
- [9] Moro, Andrea, “The Effect of Statistical Discrimination on Black-White Wage Differentials: Estimating a Model with Multiple Equilibria”, mimeo, 1997, University of Pennsylvania.
- [10] Murphy, Kevin and Finis Welch, “Wage Premium for College Graduates: Recent Growth and Possible Explanations”, *Educational Researcher*, May 1989, **18**, 17-26.
- [11] Riley, John, “Testing the Educational Screening Hypothesis”, 1979, *Journal of Political Economy*, Vol 87, No 2, pt. 2.
- [12] Schultz, T.W., “Investment in Human Capital”, *American Economic Review*, March 1961, Vol **51**, No. 1, 1-17.
- [13] Spence, A. Michael, “Job Market Signaling”, *Quarterly Journal of Economics*, August 1973, **87**, 355-74.
- [14] Stiglitz, Joseph, “The Theory of ‘Screening’, Education, and the Distribution of Income, June 1975, *American Economic Review*, Vol 65, No. 3, 283-300.

- [15] U.S. Department of Education, *Education and the Economy: An Indicators Report*, National Center for Education Statistics, NCES 97-269, by Paul T. Decker, Jennifer King Rice, Mary R. Rollefson, Project Officer. Washington D.C. 1997.
- [16] Weiss, Andrew, “Human Capital vs. Signalling Explanations of Wages”, *Journal of Economic Perspectives*, **Vol 9**, Number 4, Fall 1995, 133-154.
- [17] Wolpin, Kenneth I., “Education and Screening”, *American Economic Review*, December 1977, Vol 67, No. 5, 949-958.