INTERNATIONAL ECONOMIC REVIEW
Vol. 47, No. 4, November 2006

DISENTANGLING THE COLLEGE WAGE PREMIUM: ESTIMATING A MODEL WITH ENDOGENOUS EDUCATION CHOICES*

By Hanming Fang[1]

Yale University, U.S.A.

This article proposes and empirically implements a structural model of education choices and wage determination to quantitatively evaluate the contributions of productivity enhancement and ability signaling in the college wage premium. The model is estimated under various distributional parameterizations using 1990 U.S. Census 5% Public Use Micro Sample. Under these parameterizations, I find that college education enhances attendees' productivity by about 40%, and productivity enhancement accounts for close to two-thirds of the college wage premium.

1. Introduction

College graduates on average earn substantially more than high school graduates. For example, the college wage premium (in weekly wages) for young workers with 1–3 years of experience was about 60% in the Public Use Micro Sample of the 1990 U.S. Census.

There are two theories for the source of the college wage premium. The first is the productivity enhancement explanation pioneered by 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 is the ability signaling explanation of Spence (1973). It postulates that college wage premium arises because, by acquiring a college education, a worker signals to firms that he has higher innate ability than a high school graduate. The central difference between the two, put simply, is that different years of schooling is responsible for part of the differences of productivity among workers in the human capital theory, whereas in signaling models schooling is correlated with differences among workers that were present before the schooling choices were made. These two explanations of college wage premium have very different policy implications. First, in human capital theory, education contributes to economic growth directly by increasing the productivity of workers; whereas in signaling

* Manuscript received May 2005; revised September 2005.

[1] This article is a substantive revision of chapter two of my Ph.D. dissertation at the University of Pennsylvania. I am most grateful to Kenneth I. Wolpin and Andrew Postlewaite for their advice and many seminar and conference participants for their helpful comments and suggestions. Andrea Moro provided me with some Fortran codes used in the estimation. All remaining errors are my own. Financial support from Sloan Foundation is gratefully acknowledged. Please address correspondence to: H. Fang, Department of Economics, Yale University, P. O. Box 208264, New Haven, CT 06520-8264. E-mail: *hanming.fang@yale.edu*.

theory, education contributes to economic growth only to the extent that it facilitates a better match between workers and jobs (Stiglitz, 1975). Second, they have different implications for the congruence between private and social returns to education, which determines whether the market is likely to provide an efficient level of schooling. In human capital theory, barring human capital externality, the private and social returns to education coincide in a competitive labor market; whereas in the signaling theory, the private returns to education are likely higher than the social returns.[2]

Although most of the literature aims at testing the existence of signaling effects in the college wage premium (for more details, see the literature review in Section 2), this article attempts to empirically evaluate the *quantitative* importance of the above two sources of college wage premium. This question, however, has proven to be a challenging one.[3] The main problem is that the two theories, without imposing further restrictions, are observationally equivalent. However, if one is willing to impose parametric assumptions on the distribution of abilities in the population, then the two theories may be empirically distinguishable *within the structure of the model*. As a heuristic illustration, imagine a population of agents with heterogeneous abilities. In an ability signaling model, the college wage premium is the result of the difference in the average abilities between college and high school graduates. To the extent that education choices are endogenous, the scope of ability signaling in generating wage differentials is restricted by the equilibrating forces of education choices. On the other hand, the scope of productivity enhancement in generating college premium is not restricted by the equilibrating process since it is an exogenous technology parameter in the education production function.[4] This suggests that in general a model that admits the forces of both productivity enhancement and ability signaling can *better rationalize* the observed college wage premium than a pure signaling model, given a family of parametric models. In the extreme case that the observed college wage premium lies outside the level that can be rationalized in a pure ability signaling model, the data will actually call for some role of productivity enhancement. Note that my arguments here apply only within a parametric family of structural models.[5]

This article proposes and estimates a structural general equilibrium model of endogenous education choices and wage determination to assess the quantitative importance of ability signaling and productivity enhancement in the college wage premium. This model nests both ability signaling and human capital enhancement explanations of the college wage premium. It is worth emphasizing that the endogenous modeling of education choices is important for my empirical

[2] See Lang (1994) for a thoughtful analysis on the impact of the human capital and ability signaling debate on educational policy.

[3] See Weiss (1995) for a comprehensive critical survey of various early attempts in the literature.

[4] This argument implicitly assumes that workers of different skill levels are not complements in the production function. Otherwise, the scope of productivity enhancement induced college wage premium would also be restricted by equilibrating forces.

[5] It is useful to point out that the testable hypotheses about the existence of ability signaling in the literature reviewed in Section 2 are often derived from models with a lot of implicit structures on the ability signaling models.

analysis because it allows me to relate the college enrollment rate to the underlying distribution of relevant characteristics in the population. I then confront the equilibrium prediction of the model with 1990 U.S. Census data, and describe a structural empirical procedure to uncover 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 the data.

In order to provide a quantitative assessment of the importance of ability signaling in the U.S. education, I conduct a counterfactual experiment based on estimates of the structural parameters of the model: I consider a hypothetical economy which differs from the actual economy only in that its college education did not enhance productivity at all, and numerically solve for the equilibrium college wage premium in such a hypothetical world. Since productivity enhancement component is ruled out in the hypothetical economy, its college wage premium is due only to ability signaling. I can then ascribe the difference in college wage premium between the actual and the hypothetical economies to productivity enhancement of the college education.

Under different distributional parameterizations of the model, I find, despite differences in the estimates of other parameters, that college education magnifies an individual's ability by about 40%. In a hypothetical world in which college education is not productivity enhancing, I find that the college wage premium is at most about one-third of the actual college wage premium, from which I conclude that productivity enhancement accounts for at least two-thirds of the observed college wage premium in the data.

It is useful to make it clear at the onset that this article is a *first step* toward a long-standing empirical question via a structural approach by presenting and implementing a stylized model of education choices and wage determination. The main departure of the current article from existing literature is the explicit modeling of incomplete information regarding workers' abilities, and its main shortcoming is that the parameters of the model are identified only under parametric assumptions; hence my conclusions have to be qualified under such assumptions as well. In spite of this, I believe that a structural approach to address the human capital versus ability signaling question is a valuable complement to the existing literature.

The remainder of the article is structured as follows. Section 2 reviews the related literature; Section 3 outlines the model; Section 4 characterizes the equilibrium of the model; Section 5 describes the estimation procedure; Section 6 describes the data; Section 7 details the estimation results; Section 8 describes the counterfactual experiment we conduct to disentangle the college wage premium; Section 9 provides an informal discussion about the interpretations of our estimates; and finally Section 10 concludes.

2. RELATED LITERATURE

This article is closely related to two strands of literature. The first is the literature on "returns to schooling." Card (2001) provides a comprehensive survey on the recent progress in this literature.[6] Recent studies have attempted to use identical twins (Ashenfelter and Krueger, 1994), compulsory schooling laws (Angrist and Krueger, 1991) and other similar features to control for the omitted ability bias in estimating the private returns to schooling. An implicit assumption of this literature is that employers *perfectly observe* the productivities of the workers. This assumption directly rules out ability signaling through education. As an illustration, consider the case of identical twins. Suppose that one is a high school graduate with 12 years, and the other a college graduate with 16 years, of schooling. In the identical twins studies, it is assumed that potential employers perfectly observe their productivities, which are a combination of presumably identical innate abilities and schooling, and the differences in their wages would reflect only their productivity differences due to different years of schooling. But if workers' productivities are in fact not perfectly observed (which is particularly true for newly hired workers), then the firms would have to infer the productivities of the twins from noisy signals such as interviews, tryouts, and reference letters together with their education labels, high school versus college.[7] The main point is, because the assumption of perfect observability of productivities implicit in this literature is not likely to be true, the estimated returns to schooling from this literature consists of both productivity enhancement and ability signaling elements.

The other strand of related literature attempts to directly distinguish ability signaling and productivity enhancement explanations by testing *in reduced form* different implications of the theories. The results in this strand of literature are qualitative yes or no answers. Wolpin (1977) argues that self-employed workers are less likely to acquire education for motivations other than human capital investment. He tests if the self-employed have less schooling than salaried workers using the NBER–Thorndike sample of World War II veterans. He found that the difference in schooling for the two groups are small, from which he concludes that ability signaling function of education is minor.[8] Riley (1979) divides occupations into an "unscreened" and "screened" sectors. In unscreened sectors, a worker's productivity can be easily ascertained, hence educational signaling is unimportant, whereas in screened sectors, productivities are difficult to ascertain and thus ability signaling plays a more important role. The implication of Riley's model is that for any given number of years of schooling, productivity levels and hence lifetime earnings are on average higher in the unscreened sector than in the screened sector. He confirms this hypothesis using Current Population Survey cross section data and 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 signaling hypothesis.

[6] Table II in Card (2001, pp. 1146–47) is an excellent summary of the results from recent studies.

[7] This argument is true unless the same firm hires both twins and are aware that they are identical twins.

[8] However, Wolpin (1977) did not take into account that self-employed professionals such as lawyers and doctors may have incentives to acquire education to signal to their clients.

Lang and Kropp (1986) adopted a somewhat different approach by looking at the comparative statics properties of the signaling and human capital models. In a signaling model, a state compulsory school attendance law will increase the educational attainment of high-ability workers who are not directly affected by the law; in contrast, in a human capital model, such laws affect only those individuals whose behavior is directly constrained. They find that compulsory attendance laws do increase enrollment rates in age groups they do not affect directly, and thus support the sorting hypothesis. In similar spirits, Bedard (2001) showed that in a signaling model, better university access in terms of a local presence of colleges may increase the high school drop-out rate, whereas in a human capital model, the high school drop-out rate should be unaffected by the easier university access. She finds evidence in support of signaling.

To summarize, there is no consensus in this literature as to whether ability signaling is qualitatively important. More importantly, as Wolpin (1977) pointed out, "The real issue concerns not the mere existence of one or the other effect, but the extent to which schooling performs each of these roles." This article provides a methodological innovation to deal precisely with this task.

Recent studies by Altonji and Pierret (1998) and Lange (2004) estimate models in which employers may learn about workers' productivity over time. They estimated the speed of employer learning and use it to quantify the signaling value of education. This approach exploits the intuitive idea that as firms learn about a worker's productivity over time, the signaling component to the college wage premium should decline. The main difference between my article and theirs is as follows. My model is a general equilibrium model and the signaling value of education emerges as a result of the agents' endogenous educational choices. Both Altonji and Pierret (1998) and Lange (2004) take the wage equation as fixed in their counterfactual analysis when they assume that education does not enhance productivity. Nonetheless, it is intriguing that we reach very similar upper bounds on the contribution of signaling to the college wage premium.[9]

This article is related, to a lesser degree, to a growing literature that aims to explain the rising college wage premium in the 1980s as documented by Murphy and Welch (1989) and Juhn et al. (1993). Taber (2001) investigates whether the growing college wage premium is due to an increase in the relative demand for workers with high unobserved ability or an increase in the demand for skills accumulated in college. Heckman et al. (1998) empirically implement a general equilibrium model of human capital investment (including both schooling and on-the-job training) with heterogeneous agents. Both papers have the advantage over this article in that their models are dynamic and allow for richer form of wage determinations, but, rightly so for their purpose, they assume that individual workers' skills, though they may not be observable by researchers, are observed by the firms. This assumption rules out the ability signaling potential of college attendance, which is the focus of the current study. As will be seen below, the current article simplifies of many fronts to focus on imperfect information and ability signaling.

[9] Lange (2004)'s upper bound on the contribution of signaling to the returns to education is about 25%.

3. THE MODEL

In this section, I present a structural model of education choices and wage determination specifically designed to incorporate, in the simplest manner, both the ability signaling and productivity enhancement components of college wage premium.[10] Discussions of some of the main features of the model are presented in Section 3.F.

A. *Agents' Characteristics.* There is a continuum of agents with unit measure in the population. I assume that each agent has two characteristics (a, v) where $a \in \mathcal{A}$ is his innate ability in efficiency units and $v \in \mathcal{V}$ is his utility cost of attending college in monetary terms. It is assumed that an agent privately observes his characteristics, but it is commonly known that in the population (a, v) is distributed according to a joint c.d.f. G.

B. *Firms.* Firms behave competitively in the labor market. Each firm has two types of jobs that are, respectively, called simple and complex jobs. All agents can perform successfully on the simple job, and a worker's productivity in efficiency units is transformed one-to-one to outputs on the simple job. In contrast, complex jobs are jobs that require some special skills, that is, only workers with those skills can succeed on the complex job. A skilled worker can transform each efficiency unit to $x_s > 1$ units of output on the complex job, whereas an unskilled worker produces 0 units of output regardless of his ability. Table 1 summarizes the firms' production technologies.

C. *Agents' Education Choices and Skill Investment Decisions.* Individuals in my model are assumed to have just completed junior high school and are contemplating whether they should enter the labor market after high school or after attending college. I assume that high school attendance is compulsory whereas college attendance is not. I denote the education choices by $\{h, c\}$ where h stands for "high school" and c for "college."

Agents also decide whether to undertake skill investment that will qualify them for the complex job. The cost of skill investment, denoted by z, depends on the agent's innate ability a and it is assumed to be $k(a)$ where $k'(a) < 0$. Let $\mathcal{Z} \equiv k(\mathcal{A})$ be the domain of z. Hence, it is less costly for an agent with higher innate ability to acquire the requisite skills for the complex job. Note that I assume that an agent's skill investment cost does not depend on his education choice.

In this model, it is important to distinguish education choices and skill investment decisions as separate and distinct. I interpret college education as an investment in *general* human capital, and I assume that college education augments a

[10] Explicit modeling of educational choices is also the focus of Willis and Rosen (1979), but that paper assumes that firms have perfect information about workers' productivity. Hendel et al.'s (2005) model is closest in spirit to mine in that they also assume that college attendance decisions are affected by both innate ability and other costs. Their focus is to analyze how government-financed educational subsidy affects income inequality in a signaling model. Also see Weiss (1983) for an alternative model of education that allows for both learning and signaling.

TABLE 1

OUTPUT BY A WORKER WITH ONE EFFICIENCY UNIT ON THE TWO JOBS

	Simple Job	Complex Job
Skilled worker	1	x_s
Unskilled worker	1	0

worker's efficiency units. In contrast, skill investments are *specific* human capital or know-how to perform the complex job. That is, it qualifies the worker for the complex job without changing the worker's efficiency units. Thus in my model, college education enhances general productivity and skill investment provides specific knowledge requisite for the complex job. This distinction also plays a crucial role in my empirical identification (see Section 5.D for a detailed discussion).

D. *Productivity Enhancement of College Education.* I assume that college attendance augments an agent's productivity in efficiency units by a factor $\rho^c > 1$. As mentioned earlier, college education is assumed to provide general human capital that makes the individual more productive at every job in his future career. To illustrate, if a worker has innate ability a, then his efficiency units become $\rho^c a$ if he attends college. Therefore his productivity on the simple job will be $\rho^c a$. If he also invests in skill while in college, his productivity on the complex job will be $\rho^c a x_s$.

In general, the productivity augmenting factor ρ^c may vary across agents; specifically, it may depend on the agent's innate ability a. However, it is not clear whether more able agents will benefit more or less from college attendance than less able ones. In this article, I make the simplistic assumption that ρ^c is constant across agents, not for its realism but for its tractability. The assumption that ρ^c is independent of a also plays an important role in identification.

E. *Time Line of The Events.* The events in this model can be divided into four stages.

Stage 1: Education choice. Each agent, based on his privately observed (a, v), chooses whether to attend college. He incurs the college attendance utility cost v if he chooses college. The education choice profile in the population is written as $J : \mathcal{A} \times \mathcal{V} \to \{h, c\}$. Agents' education choices are observed by the firms.

Stage 2: Skill investment decision. Each agent also decides whether to invest in the requisite skills for the complex job (s) or not invest (u). It will become clear later that the sequential order of stages 1 and 2 is unimportant. Each agent incurs a cost $z = k(a)$ if he invests in skills and none if not. The skill investment decision profile in the population is written as $I : \mathcal{A} \times \mathcal{V} \to \{s, u\}$. Agents' skill investment decisions are *not* observed by the firms.

Stage 3: Test signal. The firms observe a noisy but informative signal, $\theta \in [0, 1]$, about whether or not the agent is skilled. The signals could include, for example, certificates, interviews, transcripts, tryouts, and reference letters. An agent's signal θ is drawn from a continuous p.d.f. f_s if he invests in skills in Stage 2 and from p.d.f. f_u otherwise. To capture the idea that the signal is informative about the agent's skill, I assume the following standard monotone likelihood ratio property (MLRP):

ASSUMPTION MLRP: *f_s/f_u is strictly increasing in θ.*

This assumption implies that skilled agents are more likely to receive higher test signals than unskilled ones and, conversely, the posterior probability that an agent is skilled is larger if he receives a higher signal.

Stage 4: Competitive wage determination, labor market clearing, and job assignments. After observing the test signals, firms compete for workers by simultaneously announcing wage schedules. I allow a firm's offer of wage schedule to depend on educational attainment and test signal. A firm's offer of wage schedule to workers with educational attainment j is denoted by $w^j : [0, 1] \to \mathbb{R}_+$ where $j \in \{h, c\}$. After observing the offers of wage schedules, each agent then chooses for which firm to work, and the labor market clears. The firm then decides the job assignment of each worker based on his educational attainment and test signal. The job assignment rule is denoted by $T : \{h, c\} \times [0, 1] \to \{simple, complex\}$.

The payoff for an agent with characteristics (a, v), education level $j \in \{h, c\}$, and skill investment decision $i \in \{s, u\}$ and wage w, is given by

$$w - v \cdot 1_{\{j=c\}} - k(a) \cdot 1_{\{i=s\}},$$

where $1_{\{\cdot\}}$ is an indicator function that takes value 1 if and only if the condition in the bracket is true. The firms are risk neutral and maximize expected profits. The actual productivity of a worker with innate ability a, education choice $j \in \{h, c\}$, skill investment $i \in \{s, u\}$, job assignment $t \in \{simple, complex\}$ can be succinctly summarized by the following expression:

$$a\big[\rho^h \cdot (1 - 1_{\{j=c\}}) + \rho^c \cdot 1_{\{j=c\}}\big]\big[1_{\{t=simple\}} + \big(1 - 1_{\{t=simple\}}\big)x_s \cdot 1_{\{i=s\}}\big],$$

where $\rho^h \equiv 1$. Note that education, skill investment, and skill/job match all contribute to the worker's productivity. Table 2 summarizes the primitives of the model.

TABLE 2

A SUMMARY OF THE PRIMITIVES OF THE MODEL

G:	Joint distribution of (a, v) in the population
k:	Cost of skill investment $z \equiv k(a)$ with $k' < 0$
ρ^c:	Productivity enhancement factor of attending college
x_s:	Productivity of skilled workers on the complex job
f_s:	p.d.f. of θ for skilled workers
f_u:	p.d.f. of θ for unskilled workers

F. *Discussions of the Model.* The model outlined above is clearly very stylized, and it abstracts from many realistic features of education and labor market. But at a general level, I believe that this is a quite natural model. No one would object to the basic ideas that college attendance decisions are affected by innate ability and financial and psychological costs as well as the expected college wage premium, and that firms use educational attainment as well as more personalized signals in determining wage offers. Now I discuss some special modeling assumptions in the details of the model.

F.1. *Why do I depart from Spence's (1973) signaling model?* Spence (1973) assumes that cost of education is strictly decreasing in one's innate ability. In terms of my model, this is equivalent to reducing the two-dimensional individual heterogeneity (a, v) to one dimension, or equivalently, making the c.d.f. G "degenerate" in the two-dimensional space of (a, v). Such a "degenerate" G distribution will have a counterfactual empirical implication: Suppose that the innate abilities are distributed in the population according to a continuous distribution. Then the unique equilibrium with both education groups will be characterized by a cutoff ability level: Agents with ability above the cutoff will all attend college and below the cutoff will not. The immediate implication of such a model will be that the lowest wage earned by a college graduate must be larger than the highest wage earned by a high school graduate. Note that this implication holds despite the assumption that the test signals could be noisy, because in equilibrium firms will correctly infer that the least able college graduate is more productive than the most able high school graduate regardless of their test signals. It will require too big a measurement error to reconcile such a prediction with the data: In my sample, the lowest wage earned by college graduates is at about the 10th percentile of the high school wage distribution. For this reason, I must extend Spence's original formulation of signaling model and assume that ability and college attendance cost are not perfectly correlated.

F.2. *Why do I introduce skill investment and what is its interpretation?* Although the link between the cost of attending college and innate ability is probably quite weak (in the empirical analysis I allow for varying degrees of correlation between a and v), I do believe that one's innate ability affects his effectiveness in acquiring specific skills. Skill investments in my model are best interpreted as time and effort one spent on learning specific knowledge essential to the complex job, and it is reasonable to assume that it takes less time and effort for more able students to learn the specific skill. A concrete example of the complex job is writing computer codes, a task whose requisite skills can conceivably be acquired by both high school and college attendees with effort. Such skills are inherently difficult for the firm to evaluate before employment. One may argue that workers may provide certificates to prove their skills, but such certificates could be noisy, and as such they are incorporated in the test signal θ.

F.3. *Why do I need to introduce simple and complex jobs?* The introduction of the simple and complex jobs allows me to capture the part of productivity differential resulting from the mismatch between skills and job assignments. It also allows me to empirically implement specifications of G such that a and v are independent (see Section 5.B), in which case the channel through which ability signaling component appears in the college wage premium differs from Spence (1973).[11] In order for ability signaling to play a role despite the independence of a and v, I need to allow ability to affect the investment cost for skills requisite for the complex job, which also seems to be a realistic assumption. Most importantly, however, the simple and complex jobs as well as the skill investment play a key role in my identification strategy. In Section 5.D, I describe the difficulties of estimating a "bare bones" model in which agents only choose whether to attend college.

F.4. *What is the interpretation of utility cost of college attendance?* I make the empirically reasonable assumption that ability is not the only determinant of college attendance. For example, individuals with identical innate abilities may make different educational choices depending on the family financial resources.[12] I should also emphasize that the utility cost of college attendance v incorporates consumption value of college education, through, for example, partying, networking, civics and life-enriching art and music, etc. To summarize, I interpret the utility cost of college attendance, v, as the sum of indirect opportunity cost, direct financial cost, and psychological costs, net of any consumption value of college education. Moreover, I make the important simplifying assumption that the college attendance cost v is unobservable to employers. There are three justifications for this simplification. First, this simplifying is realistic if the cost mainly captures nonpecuniary costs. There is some evidence in support of that. Carneiro et al. (2003) show that nonpecuniary factors associated with psychic costs, motivations, and the like play a major role in explaining why individuals from low-income families do not attend college even though it is financially profitable to do so. Second, the pecuniary part of college attendance cost such as tuition expenses and forgone earnings while in college can be partially accounted for by the researcher, but typically such information is not available to the employers. Third, even if we were to estimate directly the pecuniary component of the college attendance cost, we still have to assume a distribution for the nonpecuniary component. Moreover, the marginal distribution of college attendance cost needs to be combined with the ability distribution. For these reasons I think the simplifying assumption of treating college attendance cost v as an unobservable is a reasonable starting point.

[11] When agents differ in two dimensions, both in ability a and in college attendance cost v, the technical reason that college attendance may emerge *endogenously* as a valid signaling instrument even in situations when the distributions of a and v are independent is quite different from Spence (1973). See Fang (2001) for a thorough theoretical analysis for a related model in which signaling instruments are generated endogenously in equilibrium.

[12] See Hendel et al. (2005) for a review of the evidence for borrowing constraints. Keane and Wolpin (2001) evaluate the role of borrowing constraints on educational attainment, whereas Keane (2002) and Ichimura and Taber (2002) present estimates of the impact of financial aid and tuition subsidy on college attendance in the presence of borrowing constraints.

F.5. *What is the role of test signals?* The assumption that workers' skill investment decisions are imperfectly observed by firms are clearly realistic; and the test signals have natural interpretations as described earlier. The test signals also endogenously generate wage distributions in the simplest fashion.

F.6. *Why not directly assume that college education directly changes the distribution of test signals?* One may propose an alternative model in which the signal about a college attendee's ability is more informative than that of a high school attendee. I in fact allow for such a possibility in my empirical implementation (see Equation (11) in Section 5), but fail to find substantial differences in the testing technologies. The reason is that the coefficient of variation for the college and high school wage distributions is close.

4. ANALYSIS OF THE MODEL

In this section, I characterize the equilibrium of the model. An equilibrium of the economy consists of agents' decisions regarding education choices in Stage 1, skill investment decisions in Stage 2, and firms' wage schedules and job assignment rules in Stage 4, such that every agent's decision rules are optimal given other agents' and the firms' decision rules; and every firm's decision rules are optimal given the other firms' decision rules and agents' decision profiles.

A. *Competitive Wage Determination.* In a competitive labor market in Stage 4, firms make each worker a wage offer equal to his expected productivity. Since firms do not perfectly observe a worker's skill investment decision, they make inferences based on the worker's test signal and educational attainment to decide whether he is more productive on the simple or the complex job.

In order to derive the equilibrium wage schedule in a competitive labor market, I first arbitrarily fix the agents' education choice profile J and skill investment decision profile I. Define from J and I four membership sets as follows:

$$\Omega_i^j = \{(a, v) : J(a, v) = j, I(a, v) = i\},$$

where $j \in \{h, c\}$, $i \in \{s, u\}$. The interpretation of these sets is obvious; it is also clear that membership sets convey exactly the same information as the profiles J and I.

For given profiles J and I, firms infer the fraction of skilled workers in education group j, which I denote by π^j, as follows:

$$\pi^j = \frac{\iint_{\Omega_s^j} \mathrm{d}G(a, v)}{\iint_{\Omega_s^j \cup \Omega_u^j} \mathrm{d}G(a, v)}, \tag{1}$$

where the numerator is the measure of individuals who choose education j and invest in skills, and the denominator is the total measure of individuals who choose education j. Given π^j, which serves as the prior belief about a randomly drawn worker from education group j being skilled, firms apply Bayes' rule to arrive at the posterior belief, denoted by $P^j(\theta)$, that a group j worker with test signal θ is skilled. This posterior belief $P^j(\theta)$ is given by

$$P^j(\theta) = \frac{\pi^j f_s(\theta)}{\pi^j f_s(\theta) + (1 - \pi^j) f_u(\theta)}. \tag{2}$$

Firms also calculate the expected *innate* abilities in efficiency units of the workers in each education group conditional on skill levels, which are denoted by $\mathrm{E}^{j,i}[a]$ where $j \in \{h, c\}$, $i \in \{s, u\}$, and given by

$$\mathrm{E}^{j,i}[a] = \frac{\iint_{\Omega_i^j} a \, \mathrm{d}G(a, v)}{\iint_{\Omega_i^j} \mathrm{d}G(a, v)}. \tag{3}$$

Since the firm will optimally choose to assign a worker to the job in which he is *perceived* by the firm to more be productive in expectation, the expected productivity of a worker from education group j with test signal θ is the maximum of his expected productivity on the simple and complex jobs. That is,

$$w^j(\theta) = \max\left\{P^j(\theta)\rho^j \mathrm{E}^{j,s}[a] + [1 - P^j(\theta)]\rho^j \mathrm{E}^{j,u}[a], P^j(\theta)\rho^j x_s \mathrm{E}^{j,s}[a]\right\}, \tag{4}$$

where, as a notational convention, $\rho^h = 1$, and the terms in the bracket are the worker's expected productivity on the simple and complex jobs, respectively. Note that, due to imperfect observability of worker skill investment decision, some skilled workers may be misallocated in equilibrium to the simple job and some unskilled workers to the complex job. The firm's job assignment rule, however, minimizes such misallocation subject to its available information.

An alternative expression of the competitive equilibrium wage schedule (4) is as follows. Define the threshold signal for education group j, $\tilde{\theta}^j$, as the unique solution (if a solution exists) to the following equation:[13]

$$P^j(\theta)\rho^j \mathrm{E}^{j,s}[a] + [1 - P^j(\theta)]\rho^j \mathrm{E}^{j,u}[a] = P^j(\theta)\rho^j x_s \mathrm{E}^{j,s}[a], \tag{5}$$

and set $\tilde{\theta}^j = 1$ if no solution exists. In words, a worker with a test signal $\tilde{\theta}^j$ has equal expected productivity on the simple and complex jobs. Using Assumption MLRP, I can write the wage schedule (4) as

$$w^j(\theta) = \begin{cases} P^j(\theta)\rho^j x_s \mathrm{E}^{j,s}[a] & \text{if } \theta \geq \tilde{\theta}^j \\ P^j(\theta)\rho^j \mathrm{E}^{j,s}[a] + [1 - P^j(\theta)]\rho^j \mathrm{E}^{j,u}[a] & \text{if } \theta < \tilde{\theta}^j. \end{cases} \tag{6}$$

An important feature of the equilibrium wage schedule (6) that I exploit in later empirical analysis is its monotonicity in θ, which directly follows from Assumption

[13] The uniqueness of the solution follows from Assumption MLRP.

MLRP. Note that the wage expression (4) or (6) also highlights the role of educational signaling in facilitating matches between workers and jobs emphasized by Stiglitz (1975). It is also clear that, in this model, the within-education-group wage variation arises from different values of test signal θ, which is affected by the worker's skill investment as well as chances; the between-education-group wage variation arises from differences in ρ^j, π^j, $\mathrm{E}^{j,s}[a]$ and $\mathrm{E}^{j,u}[a]$, which are affected by the productivity enhancement factor of college education and agents' education choices and skill investment decisions.

B. *Equilibrium Education Choices and Skill Investments.* I now characterize the equilibrium education choice and skill investment decision profiles (J, I), or equivalently, the membership sets Ω_i^j, $j \in \{h, c\}$, $i \in \{s, u\}$. I will occasionally abuse my earlier notation by referring to a worker by the pair (z, v) instead of (a, v) where z is his corresponding skill investment cost.

Suppose that (J, I) are candidate equilibrium education choice and investment decision profiles. The expected gross continuation equilibrium payoff for a skilled and an unskilled worker in education group j, which I, respectively, write as V_i^j where $j \in \{h, c\}, i \in \{s, u\}$, are

$$V_i^j = \int_0^1 w^j(\theta) f_i(\theta)\, d\theta, \tag{7}$$

where $w^j(\theta)$ is the continuation equilibrium wage schedule (4). I will write the pairwise differences between V_i^j, $j \in \{h, c\}, i \in \{s, u\}$ as

$$\tilde{v}_i = V_i^c - V_i^h, \tag{8}$$

$$\tilde{z}^j = V_s^j - V_u^j. \tag{9}$$

In words, $\tilde{v}_i$, $i \in \{s, u\}$, is the threshold value of the college attendance cost that an agent with skill level i is willing to incur in order to enjoy the wage schedule offered to college attendees, and $\tilde{z}^j$, $j \in \{h, c\}$, is the threshold value of the skill investment cost that an agent with educational attainment j is willing to incur in order to be a skilled worker in that group. Note that $\tilde{v}_s - \tilde{v}_u = \tilde{z}^c - \tilde{z}^h$.

With these preliminary notations, the following proposition characterizes the membership sets in any candidate equilibrium of the model assuming that $\tilde{v}_s > \tilde{v}_u$, which can be shown to hold in any equilibrium with a college wage premium. The proof, which uses simple revealed preference arguments, is relegated to Appendix A.

PROPOSITION 1. Suppose (J, I) are the education choice and investment decision profiles in a candidate equilibrium. If $\tilde{v}_s > \tilde{v}_u$, then

- $\Omega_s^c = \{(z, v) : z \le \tilde{z}^h, v \le \tilde{v}_s\} \cup \{(z, v) : z \in [\tilde{z}^h, \tilde{z}^c], v + z \le \tilde{v}_u + \tilde{z}^c\}$,
- $\Omega_u^c = \{(z, v) : z \in [\tilde{z}^c, \bar{z}], v \in [\underline{v}, \tilde{v}_u]\}$,

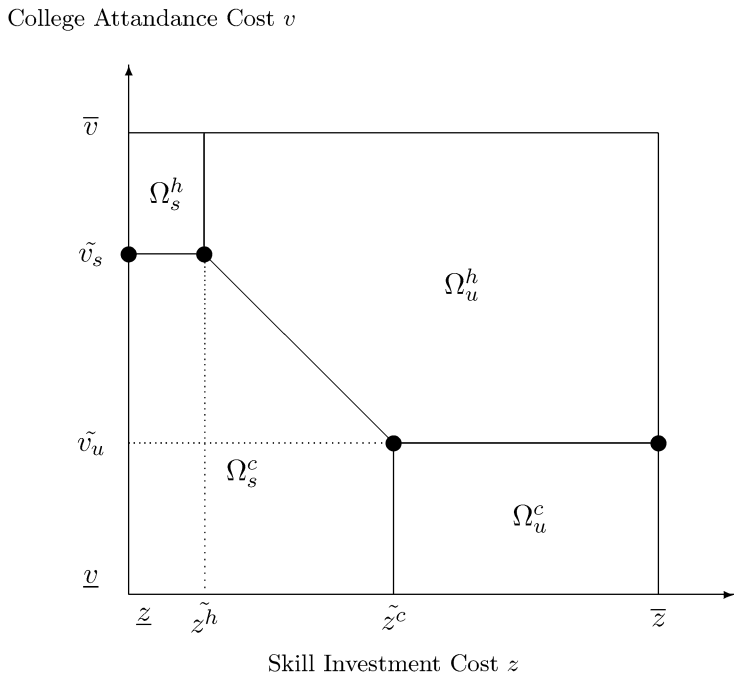


FIGURE 1

EQUILIBRIUM EDUCATION AND SKILL INVESTMENT CHOICES

- $\Omega_u^h = \{(z, v) : z \geq \tilde{z}^c, v \geq \tilde{v}_u\} \cup \{(z, v) : z \in [\tilde{z}^h, \tilde{z}^c], v + z > \tilde{v}_u + \tilde{z}^c\}$,
- $\Omega_s^h = \{(z, v) : z \in [\underline{z}, \tilde{z}^h], v \in [\tilde{v}_s, \bar{v}]\}$.

It is easier to read Proposition 1 through its graphical illustration in Figure 1. Verbally, it states that a piecewise linear function connecting the four dots in Figure 1 essentially delineates individuals' education choices and skill investment decisions. This characterization accords well with intuition: Agents with low skill investment cost and high college attendance cost will acquire skills but not attend college, whereas agents with low college attendance cost but with high skill investment cost will choose to attend college but not to acquire skills. Note that the characterization of equilibrium membership sets in Proposition 1 holds for arbitrary joint distribution of (z, v).[14]

C. *Empirical Content of the Model.* There are two major empirical contents of the above stylized structural model of education choices and wage formation. First, the equilibrium wage schedules (4) or (6) provides the model's prediction of two wage distributions, one for the college graduates and one for high school

[14] One can derive equilibrium existence conditions based on Proposition 1 but I will not focus on such issues. Instead, I will interpret the observed wage distributions and college enrollment rates as equilibrium outcomes of *some* economy.

graduates; second, Proposition 1 provides me with the model's prediction of *college enrollment rate*, λ^c, which is simply given by

$$\lambda^c = \int_{\Omega_s^c \cup \Omega_u^c} \mathrm{d}G(a, v). \tag{10}$$

We write $\lambda^h = 1 - \lambda^c$ to denote the fraction of the population who enter labor market after high school. The major advantage of the explicit modeling of education choices is that college enrollment rate, which is an aggregate statistic, becomes informative about the underlying distribution of unobservable population characteristics. Indeed I will match the model's prediction of the college enrollment rate to the data in step two of the estimation procedure below.

5. ESTIMATION PROCEDURE

I attempt to estimate the primitives of the model summarized in Table 2 from the model's empirical content described above: two wage distributions and college enrollment rate. My estimation strategy follows Moro (2003), who estimated a related structural model of statistical discrimination.[15] Exploiting the fact that conditional on (J, I) the subsequent equilibrium wage schedules are uniquely determined by formula (6), this estimation procedure consists of three steps.

A. *Step One: Matching the Wage Distributions.* It is clear from the equilibrium wage schedules (6) that $w^j(\theta)$ is a monotonic function of θ; moreover, variables π^j, $\tilde{\theta}^j$, x_s, $\rho^j \mathrm{E}^{j,s}[a]$, $\rho^j \mathrm{E}^{j,u}[a]$ and the parameter(s) in the test signal distributions f_s and f_u determine its shape. I proceed by parameterizing the distributions of the test signals. As mentioned in Section 3.F, in the estimation I allow the test signal distributions to differ for college and high school graduates. Specifically, I consider the following family of f_s^j and f_u^j (which is now superscripted by education group j):

$$f_s^j(\theta) = \eta^j \theta^{\eta^j - 1}; \quad f_u^j(\theta) = \eta^j (1-\theta)^{\eta^j - 1}, \tag{11}$$

which are, respectively, distributions of Beta(η^j, 1) and Beta(1, η^j). To satisfy Assumption MLRP, I impose the restriction that $\eta^j > 1$. This particular parameterization of the test technologies, however, is not crucial for identification in step one.[16,17]

[15] In Moro (2003), agents differ only in innate abilities and they make only skill investment choices. Also see Antonovics (2002) for a related model and an alternative step one estimation method.

[16] There is no good justification for choosing such a parameterization except that Beta distributions are versatile and they have been used to fit wage distributions (see Johnson et al., 1995, p. 235 for applications of Beta distributions).

[17] It is not sufficient to only parameterize the likelihood ratio f_s/f_u. The reason is as follows. Although knowing f_s/f_u is sufficient to invert equilibrium wage schedule (6), it is not enough to pin down the likelihood of the observed wage distributions, which involves the distributions f_s and f_u separately (see Equation (13)).

With this parameterization, the "parameters" I estimate in step one can be denoted by

$$\mathcal{P}_1^j = \left\{x_s, \eta^j, \pi^j, \tilde{\theta}^j, \rho^j \mathrm{E}^{j,s}[a], \rho^j \mathrm{E}^{j,u}[a]\right\}.$$

Note that in $\mathcal{P}_1^j$, only x_s and η^j are bona fide primitive parameters, and the others are in fact endogenous variables. How do I estimate $\mathcal{P}_1^j$? Exploiting the monotonicity of $w^j(\cdot)$ in θ, I can, for given values of $\mathcal{P}_1^j$, invert the equilibrium wage schedule (6) and infer the test signal associated with each observed wage. Likelihood of the wage distribution is formulated accordingly. This step of the estimation is analogous to the parametric empirical auction literature: In auctions, the econometricians invert the equilibrium bidding function to infer a bidder's valuation from her bid and then formulate the likelihood of the observed bids.[18]

Specifically, I write $\tau^j : \mathrm{R}_+ \to [0, 1]$ as the inverse of the wage schedule $w^j(\cdot)$, and the Jacobian of τ^j, denoted by D^j, is given by

$$D^j(w) = \frac{\mathrm{d}\tau^j(w)}{\mathrm{d}w} \equiv \left[\frac{\mathrm{d}w^j(\tau^j(w))}{\mathrm{d}\theta}\right]^{-1}. \tag{12}$$

A cross-section wage distribution for education group j is represented by a pair $\{\boldsymbol{\omega}^j, \boldsymbol{\kappa}^j\}$, where $\boldsymbol{\omega}^j = (\omega_1^j, \dots, \omega_{N^j}^j)$ is the vector of wages observed in group j (in ascending order with no loss of generality), and $\boldsymbol{\kappa}^j = (\kappa_1^j, \dots, \kappa_{N^j}^j)$ are the population weights associated with corresponding wage levels, and N^j is the number of distinct wage levels observed in group j sample. Now the likelihood of such a wage distribution, denoted by $\mathcal{L}^j(\mathcal{P}_1^j; \{\boldsymbol{\omega}^j, \boldsymbol{\kappa}^j\})$, is given by

$$\begin{aligned} \mathcal{L}^j\left(\mathcal{P}_1^j; \{\boldsymbol{\omega}^j, \boldsymbol{\kappa}^j\}\right) = \prod_{n=1}^{N^j} \kappa_n^j D^j\left(\omega_n^j\right)\big\{&\pi^j \eta^j \left[\tau^j\left(\omega_n^j\right)\right]^{\eta^j - 1} \\ &+ (1 - \pi^j)\eta^j\left[1 - \tau^j\left(\omega_n^j\right)\right]^{\eta^j - 1}\big\}. \end{aligned} \tag{13}$$

Step one estimates $\mathcal{P}_1^j$ by maximizing $\mathcal{L}^j(\mathcal{P}_1^j \mid \{\boldsymbol{\omega}^j, \boldsymbol{\kappa}^j\})$.[19]

In step one, I also obtain a consistent estimate of the college enrollment rate, $\widehat{\lambda^c}$, given by

$$\widehat{\lambda^c} = \frac{\sum_{n=1}^{N^c} \kappa_n^c}{\sum_{n=1}^{N^c} \kappa_n^c + \sum_{n=1}^{N^h} \kappa_n^h}. \tag{14}$$

The proportion of high school attendees is estimated to be $\widehat{\lambda^h} = 1 - \widehat{\lambda^c}$.

[18] See Paarsch (1992) or Donald and Paarsch (1996) for details.

[19] In order to deal with the income top coding in the census data, I make some modification to this procedure in the actual estimation. The omitted details are in Appendix B.

B. *Step Two: Recovering Parameters in G and k.* Estimates from step one, denoted by $\widehat{\mathcal{P}_1^j}$, provide me estimates of the equilibrium wage schedules, denoted by $\widehat{w^j(\theta)}$, and the corresponding test signal distributions, $\widehat{f_s^j}$ and $\widehat{f_u^j}$. These estimates in turn provide consistent estimates of V_i^j according to their formulas given by (7). That is,

$$\widehat{V_i^j} = \int_0^1 \widehat{w^j}(\theta)\widehat{f_i^j}(\theta)\,d\theta. \tag{15}$$

I then use (15) to calculate estimates of the four thresholds $\tilde{v}_i$ and $\tilde{z}^j$, $i \in \{s, u\}$, $j \in \{h, c\}$, in Figure 1 according to formulas (8) and (9).

Given a set of parameterizations regarding G—the joint distribution of (a, v), and $k(\cdot)$—the skill investment cost as a function of innate ability a, I can use the estimates of the four thresholds $\{\widehat{\tilde{v}_i}, \widehat{\tilde{z}^j}, i \in \{s, u\}, j \in \{h, c\}\}$ to calculate the model's predictions of the sizes of Ω_i^j and $\mathrm{E}^{j,i}[a]$, where $j \in \{h, c\}$, $i \in \{s, u\}$, according to Proposition 1. The idea in step two of the estimation is to find, among the assumed parametric families of G and k, the one that can match, as close as possible in the sense of (simulated) *minimum distance*, the model's predictions of a set of variables (to be detailed below) with those same variables estimated in step one. That the model is identified only under parametric specifications of G and k is a shortcoming of this article. In order to partially alleviate the concerns regarding the robustness of my findings to parametric assumptions, I experiment with different specifications of G and investigate how the main conclusions of the article are affected by different functional form assumptions.

Now I provide more details of this step. Throughout, I assume that the function $k(\cdot)$ takes a linear form:

$$k(a) = \beta_1 + \beta_2 \cdot a, \tag{16}$$

where $\beta_2 < 0$. The linear specification of $k(\cdot)$ is convenient, but is an assumption that I am unable to empirically validate.

Given the linear specification on $k(\cdot)$, I can use the estimates of the four thresholds $\{\widehat{\tilde{v}_i}, \widehat{\tilde{z}^j}\}$ to calculate, possibly numerically, the six objects described in Table 3 for any joint distribution $G(a, v; \boldsymbol{\alpha})$ where vector $\boldsymbol{\alpha}$ denotes to-be-estimated parameters in G. For example, in order to calculate the model's prediction of college enrollment rate (first row in Table 3), we integrate G over Ω^c, which is in turn estimated by $\widehat{\Omega^c}$ according to the estimated thresholds $\{\widehat{\tilde{v}_i}, \widehat{\tilde{z}^j}\}$. Similarly, in order to calculate the model's prediction of the measure of Ω_s^c under a given G, $\widehat{\widehat{\pi^c\lambda^c}}$, I numerically calculate the integral $\int_{\widehat{\Omega_s^c}} dG(a, v; \boldsymbol{\alpha})$, where $\widehat{\Omega_s^c}$ is the estimate of Ω_s^c using the estimated thresholds $\{\widehat{\tilde{v}_i}, \widehat{\tilde{z}^j}\}$.

The model's predictions, all denoted by double carets, are obtained from Monte Carlo simulations if analytical integrals are not available. Clearly, the model's predictions of the six objects are functions of $\mathcal{P}_2 \equiv (\beta_1, \beta_2, \boldsymbol{\alpha})$. Step two of the estimation strategy is to estimate $\mathcal{P}_2$ by minimizing the simulated distance between

TABLE 3

MATCHING THE MODEL'S PREDICTIONS AND THE STEP ONE ESTIMATES IN STEP TWO

Description	Model's Prediction	Step One Estimate
Col. enrollment rate	$\widehat{\widehat{\lambda^c}}$	$\widehat{\lambda^c}$
The measure of Ω_s^c	$\widehat{\widehat{\pi^c\lambda^c}}$	$\widehat{\pi^c}\widehat{\lambda^c}$
The measure of Ω_s^h	$\widehat{\widehat{\pi^h\lambda^h}}$	$\widehat{\pi^h}\widehat{\lambda^h}$
Expected innate ability efficiency units for skilled h.s. attendees	$\widehat{\widehat{\mathrm{E}^{h,s}[a]}}$	$\widehat{\rho^h \mathrm{E}^{h,s}[a]}$
Expected innate ability efficiency units for unskilled h.s. attendees	$\widehat{\widehat{\mathrm{E}^{h,u}[a]}}$	$\widehat{\rho^h \mathrm{E}^{h,u}[a]}$
Ratio of expected innate ability efficiency units of skilled over unskilled col. attendees	$\widehat{\widehat{\frac{\mathrm{E}^{c,s}[a]}{\mathrm{E}^{c,u}[a]}}}$	$\frac{\widehat{\rho^c \mathrm{E}^{c,s}[a]}}{\widehat{\rho^c \mathrm{E}^{c,u}[a]}}$

the double-careted model's predictions with their counterparts from step one, for unknown parameters $\mathcal{P}_2$. This procedure will only work, generically, for parametric families of G with four unknown parameters; moreover, generically for any four-parameter functional specifications of G, there exists a unique solution that minimizes the distance.[20,21]

In Section 7 below, I report estimation results under the following three different parametric specifications of $G(a, v; \boldsymbol{\alpha})$.[22]

1. *Independent uniform distributions specification.* I assume that a and v are independent and uniformly distributed. Specifically, I assume that a is uniformly distributed on $[\underline{\alpha}_a, \bar{\alpha}_a]$, and v is uniformly distributed on $[\underline{\alpha}_v, \bar{\alpha}_v]$. The main advantage of this specification is that I can obtain analytical expressions for the model's predictions (the double-careted variables in Table 3).

2. *Independent symmetric triangular distributions specification.* I assume that a and v have independent symmetric triangular distributions respectively on support $[\underline{\alpha}_a, \bar{\alpha}_a]$ and $[\underline{\alpha}_v, \bar{\alpha}_v]$[23] In both the uniform and triangular specifications, $\boldsymbol{\alpha} = (\underline{\alpha}_a, \bar{\alpha}_a, \underline{\alpha}_v, \bar{\alpha}_v)$

3. *Lognormal distributions specifications.* I assume that a and v have a bivariate lognormal distribution with a *known* correlation coefficient γ. Specifically, the joint probability density function of a and v is given by

[20] Recall we already have two unknowns β_1, β_2 from $k(\cdot)$.

[21] The standard error of the step two and three estimates are obtained through simulations using the estimated variance–covariance matrix of the step one estimates.

[22] Note that when I estimate the model with different specifications of G, step one of the estimation procedure does not have to be repeated.

[23] A random variable X has a symmetric triangular distribution on support $[\underline{x}, \bar{x}]$ if its probability density function P_X is given by

$$P_X(x) = \begin{cases} 4(x-\underline{x})/(\bar{x}-\underline{x})^2 & \text{if } x \in [\underline{x}, (\bar{x}+\underline{x})/2] \\ 4(\bar{x}-x)/(\bar{x}-\underline{x})^2 & \text{if } x \in [(\bar{x}+\underline{x})/2, \bar{x}]. \end{cases}$$

See Johnson et al. (1995, p. 297) for more details.

$$\frac{1}{2\pi\sigma_a\sigma_v a v\sqrt{1-\gamma^2}}$$
$$\times \exp\left\{-\frac{1}{2(1-\gamma^2)}\left[\frac{(\log a-\mu_a)^2}{\sigma_a^2}-\frac{2\gamma(\log a-\mu_a)(\log v-\mu_v)}{\sigma_a\sigma_v}+\frac{(\log v-\mu_v)^2}{\sigma_v^2}\right]\right\},$$

where $\gamma \in (-1, 1)$ is the correlation coefficient of a and v.[24] The main advantage of the lognormal specification over the first two specifications is that correlations between a and v can be allowed. But the correlation coefficient γ has to be specified because, as I discussed earlier, I can generically only deal with distribution function G with four unknown parameters. In Section 7, I present results for $\gamma = 0$ and $\gamma = -0.3$. The choice of $\gamma = 0$ is for the convenience of comparison with the other specifications. The choice of $\gamma = -0.3$ seems to have some qualified support from other researches. For example, Solon (1992) and Zimmerman (1992) have shown that, in the United States, about 20%–40% of the differences in achievement between any two parents persist into the next generation. If I interpret v, the cost of attending college, to be negatively affected by parental earnings, then the evidence from intergenerational income mobility would suggest a negative correlation between a and v in the magnitude between −0.2 and −0.4.[25] In the lognormal specification, $\boldsymbol{\alpha} = (\mu_a, \sigma_a^2, \mu_v, \sigma_v^2)$.

C. *Step Three: Estimating ρ^c.* Finally, I use the estimates of $\mathcal{P}_2$ from step two to simulate values of $\mathrm{E}^{c,s}[a]$ using Figure 1 and the estimated linear function $\hat{k}$. Denote the value of $\mathrm{E}^{c,s}[a]$ evaluated at $\widehat{\mathcal{P}_2}$ obtained in step two as $\widehat{\widehat{\mathrm{E}^{c,s}[a]}}$; then a consistent estimate of ρ^c is given by

$$\widehat{\rho^c} = \frac{\widehat{\rho^c \mathrm{E}^{c,s}[a]}}{\widehat{\widehat{\mathrm{E}^{c,s}[a]}}}, \tag{17}$$

where $\widehat{\rho^c \mathrm{E}^{c,s}[a]}$ is obtained in step one.

This concludes the description of my estimation procedure. If I find admissible values for the fundamental parameters of the economy, my estimation procedure ensures that the observed wage distributions and the college enrollment rate are an equilibrium of the estimated economy.

D. *Identification: Informal Discussions.* My structural model is only parametrically identified. This is revealed by a simple examination of the primitives of the model listed in Table 2 and its empirical content described in Section 4.C. In general, there will not be a one-to-one mapping between observable and/or inferable outcomes from the equilibrium of the model and the fundamental parameters of the model.

[24] Recall that the mean and variance of a lognormal random variable with parameters μ and σ^2 are given by $\exp(\mu + \sigma^2/2)$ and $\exp[2(\mu + \sigma^2)] - \exp(2\mu + \sigma^2)$, respectively (see Casella and Berger, 1990, p. 682).

[25] This argument is not at all tight since our interpretation of v is more inclusive than the tuition costs.

More specifically, identification of $\mathcal{P}_1^j$ in step one requires that the likelihood function (13) has a global maximum in $(x_s, \eta^j, \pi^j, \tilde{\theta}^j, \rho^j \mathrm{E}^{j,s}[a], \rho^j \mathrm{E}^{j,u}[a])$. The complicated nonlinear fashion that these "parameters" enter the likelihood function precludes me from a formal demonstration of identification, but the nonlinearity of the likelihood function in these "parameters" also gives me confidence that they are likely identifiable.[26] I should mention that the parameterizations of the test signal distributions adopted in (11) is not crucial for identifying $\mathcal{P}_1^j$. Any test signal distributions f_s and f_u that satisfies MLRP will suffice for the purpose of identification. Another important question is whether I could nonparametrically identify the testing technologies in step one, given the fact that I am using techniques similar to those used in the empirical auction literature. Unfortunately, the answer is no. The reason is that the nonparametric methods in the empirical auction literature (see Guerre et al., 2000) rely on the fact that, in the first price auction, a bidder's equilibrium bid is related to her belief about the opponents' bid distribution (which can be nonparametrically estimated) and nothing else. In my setting, however, the equilibrium wage schedule (6) depends also on other "parameters" such as $\rho^j \mathrm{E}^{j,s}[a]$, $\rho^j \mathrm{E}^{j,u}[a]$ and x_s. For this reason, the test signal distributions can only be parametrically identified.

The identification of the joint c.d.f. $G(a, v; \boldsymbol{\alpha})$ in the step two relies crucially on some parametric family restrictions on the c.d.f. $G(a, v; \boldsymbol{\alpha})$, even though generically any four-parameter families of bivariate distribution could be allowed. In step two, college enrollment rate plays an important role in identifying the parameters even under a certain parameterization of $G(a, v; \boldsymbol{\alpha})$, since without the college enrollment rate $\widehat{\lambda^c}$, I would not be able to use the the first three equations in Table 3.

At an informal level, the following two thought experiments may help illustrate how ρ^c is identified. In the first thought experiment, suppose that agents randomly choose high school or college; thus the innate ability distributions among high school and college attendees are identical. If college enhances productivity by a factor of ρ^c, then the college wage distribution would simply be a specific transformation of the high school wage distribution. Explicitly, if ψ^j is the p.d.f. of the wage distribution of education group $j \in \{h, c\}$ with support $[\underline{w^j}, \overline{w^j}]$, then in this scenario I have the relationship that $\psi^c(w) = \rho^c \psi^h(w/\rho^c)$, and $\underline{w^c} = \rho^c \underline{w^h}$, $\overline{w^c} = \rho^c \overline{w^h}$. In the second thought experiment, I imagine that college is not productivity enhancing, i.e., $\rho^c = 1$. Thus the difference between college and high school wage distributions would due only to ability signaling. In this scenario, the high school and college wage distributions will not have the specific transformation relationship as in the first thought experiment. My estimation procedure loosely measures the relative importance of the above two thought experiments underlying the between-education-group differences in their wage distributions to identify ρ^c.

[26] Moro (2003) demonstrates through simulation that the "parameters" $(\pi^j, \tilde{\theta}^j, \eta^j)$ affect the wage distributions in different dimensions.

D.1. *Discussion: Role of skill investment and simple/complex tasks in identification.* It is useful to reflect upon the role played by the skill investment in identification of the structural model. Consider a "bare bones" model in which agents are heterogeneous in (a, v), and they choose whether to attend college based on privately observed characteristics (a, v). If they attend college, their productivity in efficiency units will be augmented by a factor $\rho^c > 1$. Firms then observe a noisy signal θ about a worker's productivity from distribution f^c if he is a college graduate and from distribution f^h if he is a high school graduate. There are two possibilities about these distributions f^c and f^h.

The first possibility is that the distributions f^c and f^h do not depend on a, the innate ability of the agent. Then in any equilibrium of the model, the continuation wages of college and high school attendees are independent of agents' innate ability a, thus the college enrollment decision will be strictly determined by v: there will exist a threshold v^* such that agents attend college if and only if $v \leq v^*$. Since the distribution of signal θ does not depend on a, the firms will then in equilibrium offer wages to college and high school graduates independent of $\theta : w^c(\theta) = \mathrm{E}[a \mid v \leq v^*]$ and $w^h(\theta) = \mathrm{E}[a \mid v \geq v^*]$. That is, all the *within-group* wage variation would have to be explained, of course rather unsatisfactorily, by measurement error.

The second possibility is that the distributions f^c and f^h depend on a, i.e., θ is drawn from $f^c(\cdot \mid a)$ and $f^h(\cdot \mid a)$ for a college and high school graduate with innate ability a. As before, write Ω^c and Ω^h, respectively, as the equilibrium set of college and high school attendees in the (a, v)-space; and let $\xi^c(\cdot)$ and $\xi^h(\cdot)$ denote the marginal PDF of innate ability among college and high school attendees. Firms will then offer equilibrium wage to a worker with signal θ from education group j his expected productivity (recall that there is no simple and complex tasks in this discussion), that is,

$$w^j(\theta) = \frac{\rho^j \int a f^j(\theta \mid a) \xi^j(a)\, da}{\int f^j(\theta \mid a) \xi^j(a)\, da}, \quad j \in \{h, c\}. \tag{18}$$

The difference between wage schedules (18) and (6) highlights the role of the simple/complex task and skill investment in my empirical strategy. In (6), we can effectively write the wage schedules as functions of the simple summary statistics of the sets Ω^h and Ω^c, namely, $\rho^j \mathrm{E}^{j,s}[a]$ and $\rho^j \mathrm{E}^{j,u}[a]$, among other "parameters." In contrast, the wage schedules in this "bare bones" model (18) contain the whole marginal PDFs of innate ability $\xi^c(\cdot)$ and $\xi^h(\cdot)$; as a result, this "bare bones" model actually does not allow for a computationally tractable empirical procedure. Furthermore, the characterization of the equilibrium sets Ω^c and Ω^h in this "bare bones" model is more complicated, making it difficult to match the model's prediction of the college enrollment rate and that in the data.

D.2. *Discussion: Skill investment costs.* In this model, I assumed that an agent's skill investment cost $z = k(a)$ does not depend on his educational choice. This allows me to conveniently refer to the types of the agents either in (a, v)-space or in (z, v)-space (see the first paragraph of Section 4.B and Figure 1). An important benefit of this simplification is the tractable characterization of the equilibrium education and skill investment choices in the (z, v)-space depicted in Figure 1. If the skill investment costs depend on education choice, for example, $z^h = k^h(a)$ and $z^c = k^c(a)$, then the thresholds in Figure 1 will not be piecewise linear, and the exact form depends on the functional forms of k^h and k^c.[27] More importantly, because we are only able to identify up to six unknown parameters in step two of our estimation procedure (including parameters for the joint distribution of (a, v) and the ability–skill transformation), allowing skill investment cost to differ by education will cause serious identification problems for other parameters.

6. THE DATA

I empirically implement the model using 1990 U.S. Census 5% Public Use Micro Sample. It includes 5% of the U.S. households who took the long-form questionnaire. I impose the following criterion in selecting my sample: I select only white males who were within 3 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. Now I offer some rationales for the selection criterion. I only select white males to abstract from elements of racial and sexual discriminations. I only select those who have left school within 3 years so that experience/tenure effects and employer learning play a lesser role (see Section 10 for more discussions). I do not include self-employed individuals since they often do not have well-defined wages. I 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.

I 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"; whereas 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 2 years. I 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 my selected sample include individuals in 1989 of ages between 19 to 23, and "college graduates" include individuals of ages between 21 to 25.[28] I calculate weekly wages for individuals selected using the above criterion in the following way: First, I divide the wage and salary income by the total weeks worked and average weekly hours of work to obtain the hourly wage, which I then multiply by 40 hours to calculate the standardized weekly wages; second, I normalize the above calculated wage by the cost of living indices according to which state the individual works in 1989.

[27] Recall that the characterization in Proposition 1 holds for arbitrary functions of $k(\cdot)$.

[28] Since the Census is decennial, we cannot compare the wages of college and high school graduates that are both in the same age cohort and have the same levels of experiences.

I observe that about 1.55% of college graduates earn weekly wages below the minimum wage in 1989 (which is around \$3.35 per hour). I assume that those are due to measurement error, and these observations are deleted. I assume that the incidences of measurement errors are the same for the two education groups, thus I delete an equal fraction of high school attendees from the bottom wage distribution. In the 1990 Census the wage and salary income is top coded by \$140,000 in annual wage and salary income. The fraction of observations above the top code is very small in my selected sample. I decide to top code the weekly wages for college graduates at \$1,000 and the high school graduates at \$718.50 so that about 1% of each education group is top coded.

The above selection and calculation leave me, respectively, 18,567 different wage observations for the college graduates education group and 42,667 for high school graduates. Note that each observation in the 1990 Census has a person's weight to adjust for the stratified sampling process to make it representative; the total weights in my selected sample are 411,326 for the college group and 858,419 for the high school group.

7. ESTIMATION RESULTS

A. *Step One Estimation Results.* Panel A of Table 4 presents parameter estimates of $\mathcal{P}_1^j$ from step one using the 1989 wage distributions of high school and college attendees.[29,30] The asymptotic standard deviations of $\widehat{\tilde{\theta}^j}$, $\widehat{\pi^j}$, and $\widehat{\eta^j}$, and the simulated standard deviations (using the estimated asymptotic variance–covariance matrix of $\widehat{\tilde{\theta}^j}$, $\widehat{\pi^j}$, and $\widehat{\eta^j}$) of the other variables are in parentheses. I estimated that in 1989 about 31.5% of college graduates invested in the skills that qualify them for the complex job, whereas that fraction is about 22.7% among high school graduates. The testing technology parameters η^c and η^h are similar (1.55 for college and 1.57 for high school), although a formal t-test rejects 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), whereas an average high school graduate has efficiency units 920.37. Conditional on being unqualified, an average college attendee has efficiency units 133.9, whereas a high school attendee has 98.05. These estimates provide me estimates of the equilibrium wage schedules, denoted by $\widehat{w^j(\theta)}$, and the corresponding test signal distributions, $\widehat{f_s^j}$ and $\widehat{f_u^j}$.

Panel B of Table 4 provides the estimates and simulated standard errors of other variables that will be used in step two, including the college enrollment rate λ^c,

[29] We use the outer-product-of-the-gradient (OPG) estimator to calculate the standard deviations for the estimates $\widehat{\tilde{\theta}^j}$, $\widehat{\pi^j}$, and $\widehat{\eta^j}$ (see Davidson and MacKinnon, 1993, for a discussion of OPG estimator).

[30] The standard errors of $\rho^j \mathrm{E}^{j,u}[a]$ and λ^c are obtained using bootstrap. Recall from Section 6 that I interpret the the bottom of the wage distributions (below those corresponding to the minimum wage) for college attendees as measurement errors, and an equal fraction of high school wage observations are also deleted. However, in estimating $\rho^j \mathrm{E}^{j,u}[a]$ and its standard errors, I bootstrap on the *whole* samples (i.e., the bottoms of the wage distributions are not discarded in the bootstrap).

TABLE 4
ESTIMATES FROM STEP ONE

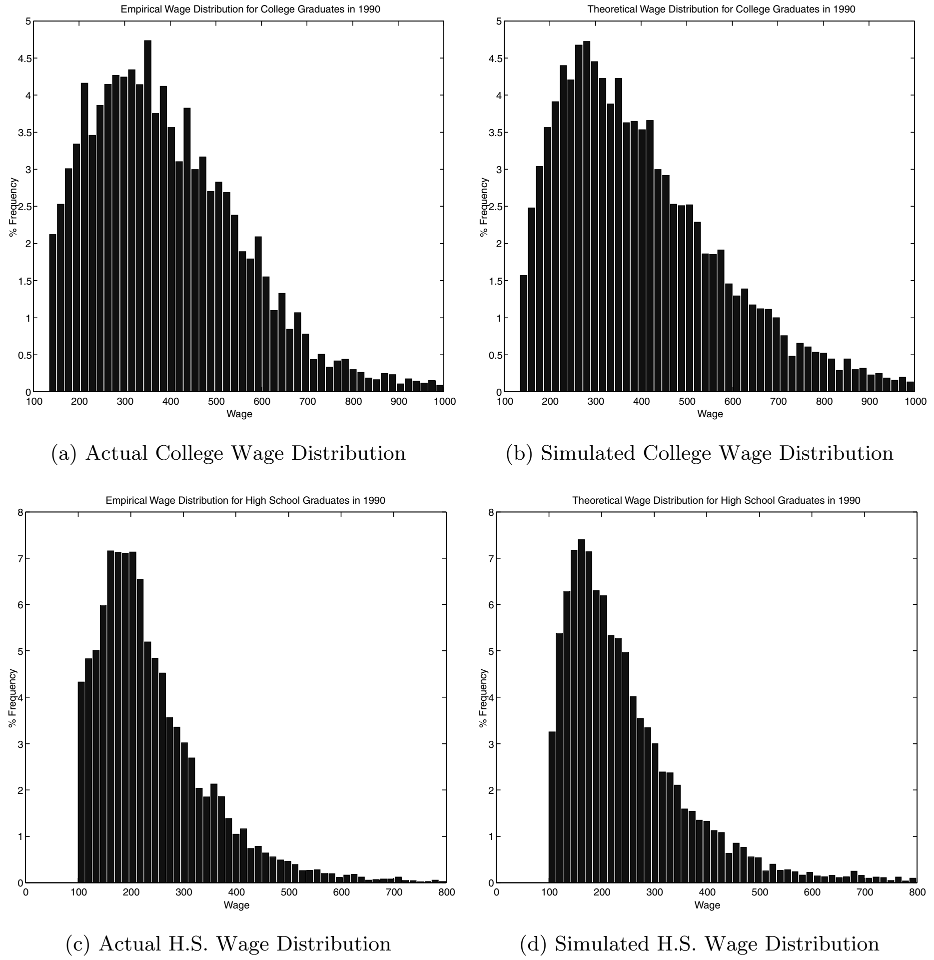
Panel A: Estimates from the Wage Distributions		
Variables	College	High School
π^j	0.315 (1.0E-03)	0.227 (0.167E-02)
$\tilde{\theta}^j$	0.877 (0.1E-06)	0.915 (0.57E-02)
η^j	1.55 (0.18E-02)	1.57 (0.11E-02)
$\rho^j \mathrm{E}^{j,s}[a]$	1,139.45 (8.94E-01)	920.37 (4.25)
x_s^j	1.086 (2.1E-04)	1.095 (3.98E-03)
$\rho^j \mathrm{E}^{j,u}[a]$	133.9 (8.92)	98.76 (7.04)
Panel B: Other Estimates		
Variables	Estimates	Standard Errors
λ^c	0.324	1.1E-04
$\tilde{v}_s$	182.95	0.85
$\tilde{v}_u$	146.48	0.63
$\tilde{z}^c$	143.35	9.8E-02
$\tilde{z}^h$	106.88	1.87E-01

and the four thresholds $\tilde{v}_i$ and $\tilde{z}^j$, $i \in \{s, u\}$, $j \in \{h, c\}$, in Figure 1.[31] The college enrollment rate is estimated to be 32.4% in my selected sample (recall that only college graduates are considered to be in the college group). This figure is in line with the college graduation rate reported elsewhere, for example, U.S. Department of Education (1997) estimates the college graduation rate, including male and female, at about 30% around 1990. This shows that my sample selection criterion does not distort the general education pattern in the United States among young men. Using formulas (8) and (9), I estimated the four thresholds in the equilibrium education and skill investment choices to be $\widehat{\tilde{v}_s} = 182.95$, $\widehat{\tilde{v}_u} = 146.48$, $\widehat{\tilde{z}^c} = 143.35$, and $\widehat{\tilde{z}^h} = 106.88$. These estimates have small standard errors, and they satisfy $\widehat{\tilde{v}_s} > \widehat{\tilde{v}_u}$, a condition required for Proposition 1.

B. *Evaluation of Step One Estimation.* Before presenting the results from steps two and three, I will evaluate how step one estimates replicate the actual wage distributions of the college and high school graduates. I simulate 10,000 wages according to the estimated wage offer schedules $\widehat{w^j(\theta)}$, and the estimated test technologies, $\widehat{f_s^j}$ and $\widehat{f_u^j}$, for each education group. Figure 2 compares the histograms of the simulated and the actual wages conditional on the wages being lower than the top-coded level, and Table 5 compares the first four centered moments for the simulated and actual wage distributions conditional on the wage being lower than the top-coded level. Both the histograms and moments suggest that step one estimates perform well in replicating the actual wage distributions.

C. *Step Two and Three Estimation Results.* In this section, I describe the results from steps two and three. As I mentioned in Section 5, I have to impose

[31] The standard errors of the four thresholds are simulated using a random sample of 1,000 draws of $\tilde{\theta}^j$, π^j, and η^j according to their asymptotic variance–covariance matrix.



(a) Actual College Wage Distribution

(b) Simulated College Wage Distribution

(c) Actual H.S. Wage Distribution

(d) Simulated H.S. Wage Distribution

FIGURE 2

SIMULATED AND ACTUAL WAGE DISTRIBUTIONS FOR H.S. AND COLLEGE GRADUATES IN 1989

parametric distributional specifications on the joint distribution of (a, v) in this stage. The results from three sets of distributional assumptions described in Section 5.C are presented in Table 6.

Panel A presents the estimates of $\mathcal{P}_2 = (\beta_1, \beta_2, \boldsymbol{\alpha})$ under the independent uniform distributions specification. The relationship between the skill investment cost and innate ability, $k(\cdot)$, is estimated to possess a negative slope of $\widehat{\beta_2} = -7.25$. The innate ability a is distributed on the support of $[-514.6, 1129.4]$. The negative lower support of the innate ability distribution a is an artifact of the uniform distributional specification of a when it attempts to match the relevant moments in Table 3 and should not be literally interpreted. Using the estimated function $\widehat{k(\cdot)}$, I can calculate that the threshold $\tilde{z}^c$ corresponds to an individual with innate ability in 706.3 efficiency units in dollars. Most importantly, I estimate, under the

TABLE 5

SUMMARY STATISTICS FOR ACTUAL AND SIMULATED WAGE DISTRIBUTIONS

		Mean	Standard Deviation	Skewness	Kurtosis
College	Actual	440.78	158.40	0.50	3.08
	Simulated	446.20	169.3	0.66	3.11
H.S.	Actual	279.07	112.44	1.02	3.98
	Simulated	280.03	114.96	0.99	3.90

TABLE 6

STEP TWO AND THREE ESTIMATES UNDER DIFFERENT SPECIFICATIONS OF G

Parameters	β_1	β_2	$\underline{\alpha}_a$	$\bar{\alpha}_a$	$\underline{\alpha}_v$	$\bar{\alpha}_v$	ρ^c
Panel A: Uniform Distributions Specification							
Estimate	5263.84	−7.25	−514.59	1129.4	35.8	406.3	1.399
Standard Error	683.21	0.93	3.25	4.89	1.05	2.32	0.00048
Panel B: Symmetric Triangular Distributions Specification							
Estimate	5654.84	−8.34	−347.8	1243.7	20.8	425.9	1.385
Standard Error	489.65	0.98	6.76	8.50	0.87	9.87	0.0025
Parameters	β_1	β_2	μ_a	σ_a^2	μ_v	σ_v^2	ρ^c
Panel C.1: Lognormal Distributions Specification, $\gamma = 0$							
Estimate	5012.48	−7.76	5.98	0.61	5.88	1.82	1.42
Standard Error	363.25	0.84	0.28	0.13	0.35	0.47	0.012
Panel C.2: Lognormal Distributions Specification, $\gamma = -0.3$							
Estimate	4986.18	−7.74	5.94	0.68	5.61	1.92	1.362
Standard Error	302.42	0.79	0.26	0.18	0.39	0.53	0.016

uniform distributions specification, that ρ^c is close to 1.4, that is, college attendance enhances an individual's general productivity by about 40%.

Panel B presents the estimates under the symmetric triangular distributions specification. The supports of the distributions of the innate ability a and the college attendance cost v change slightly from those under the uniform distributions specification. The means for the two variables are also slightly changed: Under the uniform specification, the means of a and v are, respectively, 307.4 and 221.5, whereas those means under the symmetric triangular distributions are 447.9 and 223.4, all in dollars. The individual with skill investment cost at the threshold $\tilde{z}^c$, using the estimated function $\widehat{k(\cdot)}$ under the triangular specification, has innate ability in 661.4 efficiency units. Under the uniform specification, about 74% of the individuals are below the threshold innate ability 706.3 for investing in skills if they attend college; that measure is 73.2% under the triangular specification. Similarly, about 33% of individuals have college attendance cost v above $\widehat{v_s}$ under the uniform specification, whereas that percentage is about 32 under the triangular specification. Importantly, the college productivity enhancement parameter ρ^c is estimated to be 1.385 under the triangular specification, similar to that estimated under the uniform specification.

Panels C.1 and C.2 present the estimates under lognormal specifications assuming that the correlation coefficients between a and v are, respectively, 0 and -0.3. The estimates of the mean of a and v under the lognormal specifications are quite different from those under either uniform or triangular specifications. The means of a and v are, respectively, 536.5 and 880.1 when $\gamma = 0$ and 533.8 and 713.4 when $\gamma = -0.3$. The sensitivities of the estimates of the means of a and v suggest that the welfare calculations I conduct below will be sensitive to distributional assumptions. Most surprisingly, however, is that the estimate of ρ^c seems to be quite insensitive to four distributional specifications: Under lognormal specifications, they are 1.42 and 1.36, respectively, when $\gamma = 0$ and when $\gamma = -0.3$.

8. DISENTANGLING THE COLLEGE WAGE PREMIUM

In this section, I consider how much of the estimated 58.64% college wage premium should be attributed to the productivity enhancement of college education. I conduct the following counterfactual experiment: I consider a hypothetical economy that is the same as the economy estimated above except that, counterfactually, college education is not productive (i.e., set $\rho^c = 1$), and investigate the equilibrium level of college wage premium in such an economy. I numerically solve for the equilibria of the hypothetical economy, which is a simple reparameterization of my structural model of education choices and wage determination. To the extent that the hypothetical economy differs from the estimated "actual" economy only in the productivity enhancing factor of college education, I can attribute the difference between the equilibrium college wage premium in the hypothetical economy and the actual college wage premium to productivity enhancement of college education.

Table 7 presents the summary statistics of the equilibrium of the hypothetical economies under four different parametric specifications, as well as those

TABLE 7

COMPARISON OF STATISTICS OF THE HYPOTHETICAL AND ACTUAL ECONOMIES

	Actual Econ.	Equilibrium of the Hypothetical Economies			
	Observed Equil.	Uniform	Triangular	Lognormal ($\gamma = 0$)	Lognormal ($\gamma = -0.3$)
Average wage	339.7	308.8	315.7	317.4	318.7
Col. attendance util. cost	31.4	2.33	1.93	6.87	6.32
Col. enroll. rate (%)	32.4	5.13	4.98	5.30	5.63
Skilled HS attendees (%)	22.68	25.08	26.88	27.04	26.76
Skilled col. attendees (%)	31.51	32.80	32.90	33.02	33.78
Ave. col. wage	452.90	369.45	371.82	376.58	382.95
Ave. HS wage	285.49	305.49	312.79	314.08	314.87
Col. wage premium (%)	58.64	20.93	18.87	19.90	21.62

from the observed equilibrium of the actual economy.[32,33] All of the hypothetical economies admit an equilibrium with a positive college enrollment rate, in spite of the counterfactual assumption that college attendance did not enhance productivity. The reason for attending college in the hypothetical economy is pure ability signaling. The equilibrium college wage premium in the hypothetical economies vary from 18.87% under the triangular specification to 21.62% under the lognormal specification with correlation coefficient $\gamma = -0.3$. The equilibrium college wage premium in the hypothetical economies accounts for 32.2% (i.e., 18.87/58.64) of the 58.64 percentage actual college wage premium under the triangular specification and accounts for up to 36.9% under the lognormal specification with $\gamma = -0.3$. As a way of summarizing, it seems that productivity enhancement of the college education (i.e., the fact that ρ^c is close to 1.4 instead of 1) explains close to two-thirds of the college wage premium under all four parameterizations.

The counterfactual experiment on the hypothetical economies also allows me to evaluate the output and welfare consequences of having a productivity enhancing college education. The effect on output is measured by the average wages since firms, in my model, earn zero profit. In measuring welfare cost, I also take into account the college attendance and skill investment costs. Relative to the observed equilibrium of the actual economy, outputs are, at the maximum, 9.1% lower under the uniform specification and, at the minimum, 6.2% lower under the lognormal specification with $\gamma = -0.3$. Once I take into account the college attendance and skill investment costs, however, welfare is about 1.8% higher than the actual outcome under the triangular specification and 0.3% lower under the uniform specification. The exact interpretation of this finding is not clear, but it seems to suggest that the extra cost of college attendance when ρ^c is around 1.4 (because more people attend college) roughly cancels out the beneficial effect on output due to increased productivity. The finding that under some distributional specifications (triangular and lognormal), the social welfare measure (average wage minus average college attendance utility cost) is higher in the hypothetical economy than in the actual economy may be counterintuitive. It arises because in our model individuals' private return to attend college differs from that of the social return, due to the informational frictions in the labor market. When the distribution of college attendance costs has a peak (as in a triangular or lognormal distributions), it is possible that the increase of ρ^c will attract a large fraction of individuals on the margin who have high college attendance costs to enroll in college. This would not happen under a uniform distributional assumption on v. Indeed, we find that the social welfare is higher in the actual economy than that in the hypothetical economy under the uniform distributional assumption.

[32] The values in the column labeled "Observed Equil." are obtained by simulating the model after estimation. In particular, in calculating the average wage for the actual economy, I used the estimated wage distributions to infer the distribution of wages over the top code.

[33] Each hypothetical economy admits another equilibrium in which no one attends college. However, such equilibrium is sustained with the belief by the employers that, if they were to observe a college attendee (which would not arise in the presumed equilibrium), they believe that he is of no higher ability than high school attendees. Such beliefs are not robust to standard equilibrium refinements such as the intuitive criterion of Cho and Kreps (1987).

TABLE 8

SIMULATED MAXIMAL EQUILIBRIUM COLLEGE WAGE PREMIUM FOR DIFFERENT VALUES OF ρ^c UNDER VARIOUS DISTRIBUTIONAL ASSUMPTIONS

	Distributional Assumptions			
ρ^c	Uniform	Triangular	Lognormal ($\gamma = 0$)	Lognormal ($\gamma = -0.3$)
1.0	20.93	18.87	19.90	21.62
1.1	32.88	34.21	31.08	35.13
1.2	41.59	46.32	40.98	44.09
1.3	50.17	53.29	49.73	52.10
1.4	58.65	58.67	58.41	58.66

So far, we have simulated the college wage premium in a hypothetical economy where the productivity enhancement factor $\rho^c = 1.0$. Table 8 presents the maximal sustainable college wage premium for a range of values for ρ^c from 1.0 to 1.4 under various distributional distributions on (a, v). In the simulations for each given distributional assumption, we take the corresponding parameter estimates reported in Table 6, except that we fix ρ^c at different values. Table 8 shows that the responses of the maximal sustain college wage premium to changes in ρ^c depend on distributional assumptions. Under triangular distribution, for example, an increase in ρ^c from 1.0 to 1.1 produces a much more significant increase in college wage premium than that under a uniform distribution. The response under a lognormal distribution with $\gamma = -0.3$ also differs quite substantially from that under independent lognormal distributions.

9. DISCUSSION: INTERPRETATION OF THE ESTIMATES

How should we interpret the estimates and judge whether they are reasonable? Recall that in the model the expected wage benefits of attending colleges are measured by the expected increase in *weekly* wages. Assume that an individual works about 40 weeks per year; my estimates suggest that (see the column labeled "Observed Equil." in Table 7) college graduates on average earn about \$(452.9 − 285.5) × 40 = \$6,696 *per year* more than high school graduates. From Section 8 we know that at most about one-third of \$6,696 is due to signaling. Let me for this informal discussion assume that signaling accounts for one-third of the \$6,696 (i.e., \$2,232) yearly college wage premium. Together with Lange (2004)'s finding that on average the signaling effects of education have a half-life of 3 years,[34] some simple arithmetic suggests that the undiscounted *lifetime* benefit in wages for college attendance *due to signaling* is about 5 times \$2,232, which gives an undiscounted expected wage gain for college attendance from signaling to be about \$11,160. The productivity enhancement accounts for about two-thirds of the college wage premium, which, differently from the signaling component, will never decline as firms learn about the workers' true productivity. Suppose that an agent

[34] It means that the signaling effects of education will decline by 50% within 3 years.

works for 40 years after finishing college. Thus the undiscounted lifetime benefit in wages for college attendance due to productivity enhancement is about \$6,696 $\times$ 2/3 $\times$ 40 = \$178,560. Thus the total undiscounted lifetime expected wage gain from attending college is \$11,160 + \$178,560, which is \$189,720. The college attendance costs I estimated in Table 6 are relative to the gains in weekly wages. In order to measure them in units that we are more familiar with, we have to translate them in relation to the lifetime gain in wages from attending college. Again assuming that an agent works for 40 weeks a year for 40 years after finishing college, we estimate that average attendance cost among college attendees (which has a cost of \$31.40 in relation to weekly wages in Table 7) is estimated to be \$31.40 $\times$ 40 $\times$ 40 = \$50,240. However, there is a large heterogeneity of attendance costs among college attendees. The marginal college attendees with the highest attendance costs are those with $v = \tilde{v}_s$, which is estimated in Table 4 to be \$182.95. This translates to a lifetime cost of \$182.95 $\times$ 40 $\times$ 40 = \$292,720. (Of course, in my model the expected wage lifetime wage gain for the marginal college attendee is higher than the \$189,720 average because they are skilled.) Unfortunately, given the broad interpretation that I give to the college attendance cost, there is no direct evidence to evaluate whether such estimates are plausible.

10. CONCLUSION AND REMARKS

This article proposes and empirically implements a structural model of education choices and wage determination to quantitatively evaluate the contributions of productivity enhancement and ability signaling in the college wage premium. Although I need to impose many simplifying assumptions to build a tractable structural model, the structural approach permits us to provide quantitative estimate of the importance of these two theories in explaining college wage premium. I estimated and simulated the model using 1990 U.S. Census 5% Public Use Micro Sample under a set of parametric specifications on the joint distribution of ability and college attendance costs. I find, quite robustly within the three sets of distributional parameterizations I consider, that college education enhances attendees' productivity by about 40%, and productivity enhancement accounts for close to two-thirds of the college wage premium. Less importantly, I find that, the output in a counterfactual economy with nonproductivity enhancing college ranges from 6.2% to 9.1% lower than the actual economy, but the lower outputs in the hypothetical economies are roughly compensated by the lower college attendance costs.

An important paper by Altonji and Pierret (2001) showed that employer discrimination based on observable characteristics such as education becomes less important as firms learn about workers' qualification over time. My paper ignores these important dynamic considerations by only looking at workers within three years of leaving schools. To seriously incorporate such considerations into the theoretical model and empirical implementation, I would need to address the difficult adverse selection problem in the rehiring labor market, and thus it is left for future research. I did, however, attempt to estimate the same model using data of workers who were at their prime working age, namely those with

21–24 years of working experiences. I find that the college wage premium in 1989 for this old cohort is 34.7%, smaller than the 58.6% premium for the cohort with 1–3 years of experience. This is consistent with Altonji and Pierret (2001) and Murphy and Welch (1989). Using the same estimation procedure, I estimated that, for the old cohort, the college education enhanced their productivity by about 21%, i.e., $\rho^c = 1.21$. The results from the counterfactual experiment indicate that ability signaling can account for at most 25% of the college wage premium for this old cohort, less than one-third that I found for the young cohort. Thus it does seem that ability signaling is less important for more experienced workers, consistent with the findings of Altonji and Pierret (2001).

Some final remarks about my model of education are in order. My model is stylized and some important features of education and labor markets are ignored. For example, I 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. I 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. More importantly, the cost of attending college is treated as a black box in my article. This simplifies the current analysis, but precludes me from analyzing the effect of tuition subsidy on the college enrollment and wage premium. It is probably not difficult to incorporate these issues into a theoretical model, but more data are needed to estimate such models. These problems notwithstanding, I believe that the model and its empirical implementation in the current study serve as a useful first step.

APPENDIX

A. *Proof of Proposition 1.* The proof of Proposition 1 rests on standard revealed preference arguments. It follows from the following three claims.

CLAIM 1: *In equilibrium*:

$$\begin{aligned}
&(i) \quad (z, v) \in \Omega_s^h \Rightarrow (z', v') \in \Omega_s^h \quad \text{if} \quad z' \le z, v' \ge v,\\
&(ii) \quad (z, v) \in \Omega_s^c \Rightarrow (z', v') \in \Omega_s^c \quad \text{if} \quad z' \le z, v' \le v,\\
&(iii) \quad (z, v) \in \Omega_u^h \Rightarrow (z', v') \in \Omega_u^h \quad \text{if} \quad z' \ge z, v' \ge v,\\
&(iv) \quad (z, v) \in \Omega_u^c \Rightarrow (z', v') \in \Omega_u^c \quad \text{if} \quad z' \ge z, v' \le v.
\end{aligned}$$

PROOF. I will only prove (i) since the other cases are similar. If $(z, v) \in \Omega_s^h$, then I have, by revealed preference, the following three inequalities:

$$V_s^h - z \ge V_u^h, \quad V_s^h - z \ge V_s^c - v - z, \quad V_s^h - z \ge V_u^c - v,$$

where the left-hand side is the expected payoff from being a member in Ω_s^h and the right-hand side of the above three inequalities are the expected equilibrium payoffs from being a member in, respectively, Ω_u^h, Ω_s^c, and Ω_u^c. The three equalities continue to hold if I replace z by z' and v by v' as long as $z' \le z, v' \ge v$. ■

CLAIM 2: *In equilibrium, for* $i \in \{s, u\}, j \in \{h, c\}$,

$$\sup_{\{v\}} \Omega_i^c = \tilde{v}_i, \quad \sup_{\{z\}} \Omega_s^j = \tilde{z}^j.$$

PROOF. I will prove $\sup_{\{v\}} \Omega_s^c = \tilde{v}_s$. Consider a worker of type $(z, \sup_{\{v\}} \Omega_s^c)$. Let z be sufficiently small (and possibly negative); I now prove that he must be indifferent between being in Ω_s^c and Ω_s^h. In order to show this, suppose, without loss of generality, that he strictly prefers to be in Ω_s^c. Then his revealed preference implies that $V_s^c - \sup_{\{v\}} \Omega_s^c > V_s^h$. This, in turn, implies that there exists $v' > \sup_{\{v\}} \Omega_s^c$ such that $(z, v') \in \Omega_s^c$, a contradiction to the definition of $\sup_{\{v\}} \Omega_s^c$. Hence $\sup_{\{v\}} \Omega_s^c = V_s^c - V_s^h = \tilde{v}_s$. The other equalities are proved analogously. ■

CLAIM 3: *In equilibrium with* $\tilde{v}_s > \tilde{v}_u$, *the border between* Ω_u^h *and* Ω_s^c *is a straight line connecting* $(\tilde{z}^h, \tilde{v}_s)$ and $(\tilde{z}^c, \tilde{v}_u)$.

PROOF. Suppose the border between Ω_u^h and Ω_s^c is not the mentioned straight line; then there must be a nonempty subset of Ω_u^h below it or a subset of Ω_s^c above it or both. Suppose there is a nonempty subset $\Omega \subset \Omega_u^h$ below the straight line. Pick any $(z, v) \in \Omega$. Since Ω is below the straight line connecting $(\tilde{z}^h, \tilde{v}_s)$ and $(\tilde{z}^c, \tilde{v}_u)$, I have $z + v < \tilde{z}^h + \tilde{v}_q$. From Claim 2, an individual with type $(\tilde{z}^h, \tilde{v}_q)$ is indifferent between being a member in Ω_u^h and Ω_s^c, that is, $V_u^h = V_s^c - \tilde{z}^h - \tilde{v}_q$. Together with the previous inequality, I obtain $V_u^h < V_s^c - z - v$, a contradiction to $(z, v) \in \Omega_u^h$. The other possibilities can be ruled out similarly. ■

B. *Omitted Details in Step One.* I provide some less important details of step one. First, I note that my model predicts that the lowest wage in education group j equals $w^j(0) = \rho^j \mathrm{E}^{j,u}[a]$. Hence I use the minimum wage observed in group j, namely, ω_1^j, as a consistent estimate of $\rho^j \mathrm{E}^{j,u}[a]$, i.e.,

$$\widehat{\rho^j \mathrm{E}^{j,u}[a]} = \omega_1^j. \tag{B.1}$$

Second, the census income is top coded at the maximum reported wage $\omega_{N^j}^j$. I deal with this issue as follows. First, I find the test signal, $\bar{\theta}^j$, for a group j member to earn this top-code wage by matching the right tail of the theoretical wage distribution with the fraction of observations at the top-code level. That is,

$$\pi^j (\bar{\theta}^j)^{\eta^j} + (1 - \pi^j)\big[1 - (1 - \bar{\theta}^j)^{\eta^j}\big] = 1 - \frac{\kappa_{N^j}^j}{\sum_{n=1}^{N^j} \kappa_n^j}, \tag{B.2}$$

where the left-hand side is the theoretical prediction of the proportion of group j workers receiving test signals above $\bar{\theta}^j$ and hence earning more than the top-code income; and the right-hand side is its empirical counterpart. Second, assuming that at least one individual in group j works on the complex job, the expected productivity of a worker with test signal exactly equal to $\bar{\theta}^j$ must be equal to the top-code wage level $\omega_{N^j}^j$. That is, (B.3) $$\omega^j_{N^j} = \rho^j \mathrm{E}^{j,s}[a]x_s \frac{\pi^j(\bar{\theta}^j)^{\eta^j-1}}{\pi^j(\bar{\theta}^j)^{\eta^j-1} + (1-\pi^j)(1-\bar{\theta}^j)^{\eta^j-1}}.$$

Combining (B.2) and (B.3), I can write $\rho^j \mathrm{E}^{j,s}[a]x_s$ as a function of π^j, η^j, and $\omega^j_{N^j}$.

Finally, I can rewrite Equation (5) using the parameterization of f^j_s and f^j_u given by (11) to obtain

(B.4) $$\rho^j \mathrm{E}^{j,s}[a]x_s = \rho^j \mathrm{E}^{j,s}[a] + \frac{1-\pi^j}{\pi^j}\left(\frac{1-\tilde{\theta}^j}{\tilde{\theta}^j}\right)^{\eta^j-1} \rho^j \mathrm{E}^{j,u}[a].$$

Hence I can, using the estimate of $\rho^j \mathrm{E}^{j,u}[a]$ given by (B.1) and the expression of $\rho^j \mathrm{E}^{j,s}[a]x_s$ in terms of π^j, η^j, and $\omega^j_{N^j}$ derived in the above paragraph, write $\rho^j \mathrm{E}^{j,s}[a]$ as a function of π^j, η^j, $\tilde{\theta}^j$, ω^j_1 and $\omega^j_{N^j}$.

To summarize, in the actual estimation of step one of the estimation procedure, I can write the likelihood function (13) as a function of only three variables π^j, $\tilde{\theta}^j$, and η^j.[35]

REFERENCES

ALBRECHT, J., "A Procedure for Testing the Signaling Hypothesis," *Journal of Public Economics* 15 (1981), 123–32.

ALTONJI, J., AND C. R. PIERRET, "Employer Learning and Signaling Value of Education," in I. Ohashi and T. Tachibanaki, eds., *Internal Labour Markets, Incentives and Employment* (New York: St. Martin's Press, 1998), 159–95.

——, AND ——, "Employer Learning and Statistical Discrimination," *Quarterly Journal of Economics* 116 (2001), 313–50.

ANGRIST, J., AND A. KRUEGER, "Does Compulsory School Attendance Affect Schooling and Earnings?" *Quarterly Journal of Economics* 106 (1991), 979–1014.

ANTONOVICS, K., "Persistent Racial Wage Inequality," Mimeo, University of California, San Diego, 2002.

ASHENFELTER, O., AND A. KRUEGER, "Estimates of the Economic Returns to Schooling from a New Sample of Twins," *American Economic Review* 84 (1994), 1157–73.

BECKER, G., *Human Capital* (New York: Columbia University Press for the National Bureau of Economic Research, 1964).

BEDARD, K., "Human Capital versus Signaling Models: University Access and High School Dropouts," *Journal of Political Economy* 109 (2001), 749–75.

CARD, D., "Estimating the Returns to Schooling: Progress on Some Persistent Econometric Problems," *Econometrica* 69 (2001), 1127–60.

CARNEIRO, P., K. HANSEN, AND J. HECKMAN, "Estimating Distributions of Treatment Effects with an Application to the Returns to Schooling," *International Economic Review* 44 (2003), 361–422.

CASELLA, G., AND R. L. BERGER, *Statistical Inference* (Pacific Grove, CA: Brooks/Cole Publication, 1990).

[35] I also estimated the model treating those earning below the minimum wage as being unemployed (i.e., their signal θ tells the employers that they are not productive enough to be hired at the minimum wage) and thus treat the bottom of the income distribution similar to how I treated the top coding. I obtain similar results.

Cho, I.-K., and D. M. Kreps, "Singaling Games and Stable Equilibria," *Quarterly Journal of Economics* 102 (1987), 179–221.

Davidson, R., and J. MacKinnon, *Estimation and Inference in Econometrics* (New York: Oxford University Press, 1993).

Donald, S. G., and H. J. Paarsch, "Identification, Estimation, and Testing in Parametric Empirical Models of Auctions within the Independent Private Values Paradigm," *Econometric Theory* 12 (1996), 517–67.

Fang, H., "Social Culture and Economic Performance," *American Economic Review* 91 (2001), 924–37.

Guerre, E., I. Perrigne, and Q. Vuong, "Optimal Nonparametric Estimation of First Price Auctions," *Econometrica* 68 (2000), 525–74.

Heckman, J., L. Lochner, and C. Taber, "Explaining Rising Wage Inequality: Explorations with a Dynamic General Equilibrium Model of Labor Earnings with Heterogeneous Agents," *Review of Economic Dynamics* 1 (1998), 1–58.

Hendel, I., J. Shapiro, and P. Willen, "Educational Opportunity and Income Inequality," *Journal of Public Economics* 89 (2005), 841–70.

Ichimura, H., and C. Taber,"Semiparametric Reduced Form Estimation of Tuition Subsidies," *American Economic Review Papers and Proceedings* 92 (2002), 286–92.

Johnson, N. L., S. Kotz, and N. Balakrishnan, *Continuous Univariate Distributions*, Vol. 2, 2nd ed. (New York: John Wiley & Sons Inc., 1995).

Juhn, C., K. M. Murphy, and B. Pierce, "Wage Inequality and the Rise in Returns to Skill," *Journal of Political Economy* 101 (1993), 410–42.

Keane, M., "Financial Aid, Borrowing Constraints, and College Attendance: Evidence from Structural Estimates," *American Economic Review Papers and Proceedings* 92 (2002), 293–97.

——, and K. Wolpin, "The Effect of Parental Transfers and Borrowing Constraints on Educational Attainment," *International Economic Review* 42 (2001), 1051–1103.

Lang, K., Does the Human-Capital/Educational-Sorting Debate Matter for Development Policy?" *American Economic Review* 84 (1994), 353–58.

——, and D. Kropp, "Human Capital Versus Sorting: The Effects of Compulsory Attendance Laws," *Quarterly Journal of Economics* 101 (1986), 609–24.

Lange, F., "The Returns to Schooling and Ability during the Early Career: Evidence on Employer Learning and Post-school Investment," Mimeo, Yale University, 2004.

Mincer, J., *Schooling, Experience and Earnings* (New York: Columbia University Press, 1974).

Moro, A., "The Effect of Statistical Discrimination on Black-White Wage Differentials: Estimating a Model with Multiple Equilibria," *International Economic Review* 44 (2003), 467–500.

Murphy, K., and F. Welch, "Wage Premium for College Graduates: Recent Growth and Possible Explanations," *Educational Researcher* 18 (1989), 17–26.

Paarsch, H. J., "Deciding between the Common and Private Value Paradigms in Empirical Models of Auctions," *Journal of Econometrics* 51 (1992), 191–215.

Riley, J., "Testing the Educational Screening Hypothesis," *Journal of Political Economy* 87 (1979), Part 2, S227–52.

Schultz, T. W., "Investment in Human Capital," *American Economic Review* 51 (1961), 1–17.

Solon, G., "Intergenerational Income Mobility in the United States," *American Economic Review* 82 (1992), 393–408.

Spence, A. M., "Job Market Signaling," *Quarterly Journal of Economics* 87 (1973), 355–74.

Stiglitz, J., "The Theory of 'Screening', Education, and the Distribution of Income," *American Economic Review* 65 (1975), 283–300.

Taber, C. R., "The Rising College Premium in the Eighties: Return to College or Return to Unobserved Ability?" *Review of Economic Studies* 68 (2001), 665–91.

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 DC, 1997.

WEISS, A., "A Sorting-cum-Learning Model of Education," *Journal of Political Economy* 91 (1983), 420–42.

——, "Human Capital vs. Signaling Explanations of Wages," *Journal of Economic Perspectives* 9 (1995), 133–54.

WILLIS, R., AND S. ROSEN, "Education and Self-Selection," *Journal of Political Economy* 87 (1979), Part 2, S7–S36.

WOLPIN, K. I., "Education and Screening," *American Economic Review* 67 (1977), 949–58.

ZIMMERMAN, D. J., "Regression Toward Mediocrity in Economic Status," *American Economic Review* 82 (1992), 409–29.