Student Abilities During the Expansion of US Education*

Lutz Hendricks† Todd Schoellman‡

January 17, 2012

Abstract

The US experienced two dramatic changes in the structure of education in a fifty year period. The first was a large expansion of educational attainment; the second, an increase in IQ test score gaps between college bound and non-college bound students. We study the impact of these two trends on the composition of school groups by innate ability and the importance of these composition effects for wages. Our main finding is that there is a growing gap between the abilities of high school and college-educated workers that accounts for one-half of the college wage premium for recent cohorts and the entire rise of the college wage premium for the 1910–1960 birth cohorts.

JEL: I2, J24.

Key words: Education. Ability. Skill premium.

*For helpful comments we thank Berthold Herrendorf, Richard Rogerson, Guillaume Vandenbroucke, and seminar participants at the Federal Reserve Banks of Atlanta and Cleveland, North Carolina State University, the University of Georgia, the University of Iowa, the University of Pittsburgh, the Clemson University Bag Lunch, the Triangle Dynamic Macro Workshop, the 2009 Midwest Macroeconomic Meeting, the 2009 NBER Macroeconomics Across Time and Space Meeting, the 2009 North American Summer Meeting of the Econometric Society, the 2009 Society for Economic Dynamics, and the 2010 Canadian Macro Study Group. The usual disclaimer applies.

†University of North Carolina, Chapel Hill; lutz@lhendricks.org

‡Arizona State University; todd.schoellman@gmail.com

1 Introduction

The twentieth century witnessed an extraordinary expansion of education in the United States. Figure 1a illustrates this trend. For the birth cohorts born every ten years between 1910 and 1960, it displays the fraction of white men in four exhaustive and mutually exclusive education categories: high school dropouts (<HS), high school graduates (HS), those with some college but not a four-year degree (SC), and college graduates with at least a four-year degree (C+). Of the men born in 1910, only one-third finished high school. By the 1960 cohort, high school graduation had become nearly universal and the median man attended at least some college.[1]

At the same time that high school completion and college enrollment were expanding, there were also large and systematic changes in who pursued higher levels of education. The general trend was for education to become more meritocratic, with ability and preparation becoming better predictors of educational attainment. Figure 1b gives evidence of this point. It shows for several birth cohorts between 1907 and 1960 the average percentile score on IQ tests for those who continue on to college, versus the same score for those who stop with high school completion.[2] Continuing to college here pools together the categories we call SC and C+. The trend is striking. For the very earliest cohorts, students who do not continue on to college score nearly as well on IQ tests as students who do. By the cohorts born in the mid-1940s a considerable gap had grown between the two. A recent literature has attempted to explain this trend, with the introduction of the SAT, a decline in discrimination, and the GI Bill often cited as major forces (Bowen, Kurzweil, and Tobin 2005, Herrnstein and Murray 1994).

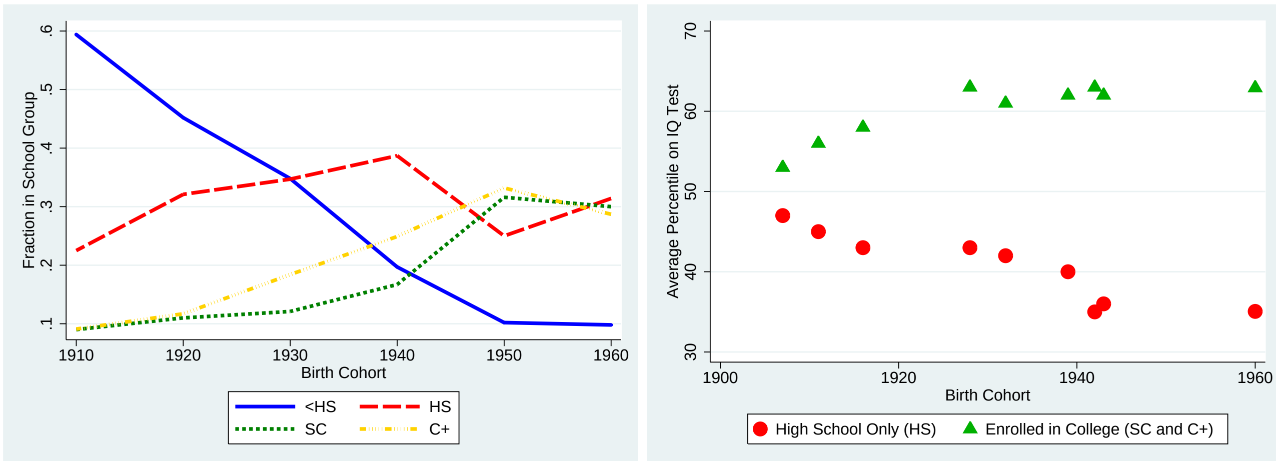
Our main idea is that these two trends have combined to change the composition of innate abilities by educational attainment for different cohorts. For example, it is unlikely that the motivation and talent of high school dropouts is the same for the 1910 and 1960 cohorts, given that more than half of the 1910 cohort dropped out but less than ten percent of the 1960 cohort did. Likewise, the innate characteristics of college graduates are likely to have changed given the large expansion of college enrollment and the changes in how college students are selected.

[1] These data are derived from the 1950–2000 population censuses. We focus on cohorts born at ten year intervals to match with the ten year intervals between censuses. Each data point represents average schooling at age 40 for the relevant cohort. For more details on the construction of the data in Figure 1, see Appendix A.1.

[2] Data points for the pre-1950 cohorts all come from Taubman and Wales (1972), who collected the results of a number of earlier studies on IQ and schooling and converted them into this common metric. The data point for the 1960 cohort comes from our own calculations based on the NLSY79.

Figure 1: Changes in US Education in the Twentieth Century

(a) The Expansion of Education (b) Changes in IQ Test Scores



Our primary motivation for studying compositional effects is to understand their importance for the evolution of wage patterns over the course of the twentieth century. To be concrete, we focus on two well-known features of the college wage premium. First, the college wage premium rose by 15 percentage points between the 1910 and 1960 cohorts.[3] Second, the current college wage premium is 50 percentage points, which is difficult to reconcile with the low college completion rate in human capital models.[4] We establish in this paper that changes in the composition of student abilities by educational attainment between the 1910 and 1960 cohorts can quantitatively explain the entire rise in the college wage premium while simultaneously making it easier to reconcile the current college wage premium with human capital theory.

To fix ideas, we think of the average log-wages of workers with a particular educational attainment as being a function of the price of skills specific to that education group and the quantity of those skills the average worker provides. The quantity is in turn determined by workers' innate abilities and the human capital they acquire over the course of their life. Much of the previous literature seeking to explain the college wage premium holds the quantity of skills fixed and focuses on reasons why skill prices may have changed – for

[3]Katz and Murphy (1992), Bound and Johnson (1992), Autor, Katz, and Krueger (1998), and Goldin and Katz (2008) propose skill-biased technological change as an explanation for the rising skill premium. Bound and Johnson (1992) and the survey of Levy and Murnane (1992) propose other explanations including international trade or migration.

[4]Heckman, Lochner, and Todd (2006) and Heckman, Lochner, and Todd (2008) make this point. They extend the basic model to allow for a psychic cost of schooling or risk in human capital investments, which is an alternative way to make the model consistent with the data.

example, due to skill-biased technological change. We allow for either component of wages to change. The primary challenge we face is that while mean wages are observed directly, the other terms – skill prices, human capital, and ability – are not. Our approach to this problem is to use the information provided by IQ test scores. We treat IQ test scores as potentially noisy proxies for ability and use them to disentangle the role of ability from the other two factors. Our methodology does not allow us to separate skill prices from human capital.

We begin by writing down a simple model of school choice with heterogeneous ability that formalizes the challenge we face. We show that the quantitative impact of compositional effects on wages are controlled by two parameters. The first governs how strongly sorted the different school groups are by their innate ability; more sorting means larger gaps in mean ability between school groups. The second parameter governs the mapping from ability to wages; a higher value for this parameter means that mean ability gaps have larger implication for wages. We take this model to the data in two steps.

First, we calibrate the model to the NLSY79 (Bureau of Labor Statistics; US Department of Labor 2002). The NLSY79 is a representative sample of cohorts born around 1960 that includes information on their wages, education, and IQ test scores. We construct two key moments from this data set: the relationship between wages and IQ test scores, and the degree of educational sorting by IQ scores. We begin by following the previous economic literature and consider the special case where IQ test scores measure ability exactly (Heckman, Lochner, and Taber 1998, Garriga and Keightley 2007). In this case our two empirical moments identify the two key parameters of the model and we can provide some simple results. However, we also draw on evidence from the psychometric literature to establish that IQ test scores are likely a noisy measure of ability. We show how to bound the plausible amount of noise in IQ test scores and recalibrate our model. Our first key result is that differences in mean ability between college and high school graduates likely account for half of the observed college wage premium.

Then we calibrate the model to fit the historical changes in schooling over the course of the twentieth century from figure 1. Our second key result is that changes in the ability gap between high school and college graduates explain the *entire* rise in the college wage premium between the 1910 and 1960 cohorts. We provide decompositions to show that the expansion of education and the increase in sorting each explain about half of the total result. Finally, we provide a number of robustness checks on the calibration exercise and for the key empirical moments that identify our model.

Our paper is most closely related to two existing literatures. The first concerns the role of changing ability, measured by IQ test scores. Finch (1946) and Taubman and Wales (1972) use aptitude and achievement test score data from several studies to identify changes in student abilities over time, but do not study the implications of this process for changes in wages. Laitner (2000) formulates a model of human capital investment with workers of heterogeneous abilities that qualitatively generates predictions for relative wages and wage inequality consistent with US post-war data. However, he does not attempt to quantify the implications of his model.

A second literature consists of papers that decompose wage premiums or changes in wage premiums into the underlying effects of skill prices versus skill quantities. A well-known set of papers use instrumental variable regressions to measure the private return to schooling as opposed to the portion that arises from differences in mean ability. However, IV approaches measure the return to schooling for a particular group; if there is heterogeneity in this return, interpreting these coefficients is not straightforward (Card 2001, Heckman, Lochner, and Todd 2006). A smaller set of papers compares the wage premiums of men from different cohorts observed in the same year (Juhn, Kim, and Vella 2005, Acemoglu 2002). It is not clear whether such wage gaps can be attributed to mean ability gaps or to gaps in skill prices for workers who are also of different ages (Card and Lemieux 2001). Probably the most related paper to ours is Bowlus and Robinson (2010). They use the predictions of the human capital accumulation model of Heckman, Lochner, and Taber (1998) to derive conditions that enable them to disentangle changes in skill prices from changes in skill quantities, given observed wage changes. They focus more on recent cohorts; for this time period, we find answers consistent with theirs, as we will show in the paper. However, our methodology is very different. Further, our exercises extend to an earlier period than do theirs.

The rest of the paper is organized as follows. Section 2 introduces our model of school choice. Section 3 calibrates the model to the NLSY79 and derives cross-sectional results. Section 4 calibrates the model to the time series data and derives further results. Section 5 provides robustness checks and the final section concludes.

2 A Model of School Choice

Our first goal is to specify a parsimonious model of school choice that formalizes the intuition from the introduction. We show that the quantitative magnitude of our results depends on two key parameters. The model guides our subsequent empirical work.

The basic environment is a discrete time overlapping generations model. Each year a cohort of unit measure is born. Individuals are indexed by their year of birth τ as well as their age v, with the current period given by $\tau + v - 1$. Individuals live for a fixed T periods.

2.1 Endowments

At birth, each person is endowed with a variety of idiosyncratic, time-invariant traits that affect their wages and schooling. We assume that these traits are captured by a two-dimensional endowment (a, p). a represents cognitive ability. Cognitive ability is useful for both work and school, because it makes it easier to learn and process new information or perform new tasks. p represents the taste for schooling. It is a preference parameter that captures the relative (dis)utility that a person derives from spending time in school instead of working. The two traits are assumed to be independent. We assume that cognitive abilities are drawn from a time-invariant standard normal distribution. Given the assumptions that we make below, both the mean and the standard deviation of this distribution can be normalized in this way. An individual's tastes for schooling are also drawn from a normal distribution with mean 0 and a standard error $\sigma_{p,\tau}$. Given that these are the only two endowments in the model, we can denote by $q = (a, p, \tau)$ the type of an agent, their endowment and their birth cohort.

2.2 Preferences

Let $c(q, v)$ denote the consumption of a person of type q at age v, and let $\beta > 0$ be the common discount factor. Then lifetime utility is given by:

$$\sum_{v=1}^{T} \beta^v \log[c(q, v)] + (p + a)\chi(s, \tau). \quad (1)$$

Workers value consumption in the standard way. They also place a direct utility value on their time spent in school, which is determined by the interaction between a worker-specific component $(p + a)$ and a cohort and school-specific component $\chi(s, \tau)$. The former term captures how enjoyable (p) and easy (a) a particular individual finds schooling to be. The latter captures how desirable school type s and its associated career paths are for cohort τ. It varies by cohort to capture changes in school and work, such as the amount of studying required to succeed in college or the career paths open to those with a particular educational attainment. We restrict χ to be positive and increasing in s. In this case, workers with a higher taste for schooling or who find school easier derive more utility from school and go to school longer, all other things equal. This complementarity is essential for our results. We could adopt alternative functional forms that preserve complementarity and our results would obtain; we have chosen this functional form as the simplest.

2.3 Budget Constraint

School type s takes $T(s)$ years to complete. While in school, students forego the labor market. After graduation, workers receive earnings $w(s, q, v)$ that depend on their school attainment, age, and ability. Their budget constraint requires them to finance lifetime consumption through lifetime earnings,

$$\sum_{v=1}^{T} \frac{c(q, v)}{R^v} = \sum_{v=T(s)+1}^{T} \frac{w(s, q, v)}{R^v}, \qquad (2)$$

where R is the exogenous interest rate.

In keeping with much of the literature, we assume that workers with different educational attainments provide different labor inputs.[5] We assume that wages are given by

$$\log[w(s, q, v)] = \theta a + z(s, \tau + v - 1) + h(s, v).$$

Wages have three determinants in our model. As mentioned before, cognitive ability affects wages directly. Since we have assumed that ability is distributed standard normal, θ is an important parameter. It measures the increase in wages that comes from a one standard deviation rise in cognitive ability. $z(s, \tau + v - 1)$ is the price per unit of type s labor supplied by cohort τ at age v. Finally $h(s, v)$ captures the human capital accumulated by workers of education s at age v through experience or learning-by-doing, which has a systematic effect on wages.

2.4 Characterization of School Choice

Workers choose their school attainment s and a consumption path $c(q, v)$ to maximize preferences (1) subject to their budget constraint (2). We characterize the solution in two

[5] For example, our setup is consistent with the literature that allows high school and college-educated workers to be imperfect substitutes in aggregate production. However, we do not take a stand on the demand side of the market since doing so is not essential to our model. One channel that we are implicitly ruling out is that the rising skill premium may reflect an increase in the rental price of high ability labor relative to low ability labor (Juhn, Murphy, and Pierce 1993, Murnane, Willett, and Levy 1995).

steps: first, we find the optimal allocation of consumption over time given school choice; then we find the school choice that maximizes lifetime utility.

Consumption in this model satisfies the standard Euler equation, $c(q, v+1) = \beta R c(q, v)$. If we combine this equation with the budget constraint and then plug into the utility function, we can rewrite lifetime utility as:

$$\theta a \sum_{v=1}^{T} \beta^v + \sum_{v=1}^{T} \beta^v \log \left[\frac{R(\beta R)^{v-1}}{\sum_{u=1}^{T} \beta^{u-1}} \sum_{u=T(s)+1}^{T} \frac{e^{h(s,u)+z(s,\tau+u-1)}}{R^u} \right] + (p+a)\chi(s,\tau). \tag{3}$$

This equation has three additive terms. The first term captures the effect of ability on lifetime utility: higher ability allows for higher lifetime consumption. The second term captures the impact of school attainment on lifetime utility: more schooling means fewer years in the labor market but also changes the skill price and the rate of human capital accumulation. Finally, the last term captures the direct utility effect of schooling.

A key property of our model is that school choices depend only on the sum $p + a$, and not on other individual-specific attributes or on p or a independently. To see this, note that the first term of our indirect utility function depends on ability but does not interact with school choices, so that it drops out of the individual's optimization problem. The second term does not depend on p or a. So endowments interact with school choice only through the third term, which includes the linear combination $p+a$. Our model includes the common property that ability does not affect school choice through the earnings channel, because it raises both the benefits of schooling (higher future wages) and the opportunity cost (higher foregone wages today) proportionally. Instead, ability, tastes, and school choice interact through preferences in the third term. Given our assumptions on $\chi(s, \tau)$, school attainment in our model is increasing in $p + a$. The individuals who have the highest combination of $p + a$ will choose college; those with middling values will choose high school graduation or some college; and those with the lowest values will choose to drop out of high school.

Since cognitive ability is one component of the sum $p + a$, the model generates positive but imperfect sorting by cognitive ability into school attainment. Further, since the standard deviation of ability is normalized to 1, the degree of sorting by ability into educational attainment is controlled by a single parameter, $\sigma_{p,\tau}$. As $\sigma_{p,\tau}$ rises, more of the variation in $p+a$ comes from variation in p. In this case, workers are less sorted by ability across school groups and mean ability gaps are smaller. In the limiting case of $\sigma_{p,\tau} = \infty$, educational choices are explained entirely by tastes for schooling. In this case, $E(a|s) = E(a) = 0$ for all school groups.

2.5 Implications for Mean Ability and Wages

Since the model allows for positive sorting by cognitive ability, it generates composition effects that matter for wages. The average wage of workers from cohort τ with education s at age v is given by:

$$E[\log(w)|s, \tau, v] = \theta E[a|s, \tau] + z(s, \tau + v - 1) + h(s, v).$$

In our model, these wages are affected by three terms: by $\theta E[a|s, \tau]$, which we call effective ability; by skill prices, z; and by human capital, h. Our goal is to separate out the role of effective ability in explaining wage patterns from the other two terms. We make no attempt to separate out skill prices from human capital endowments in this paper.

The quantitative magnitude of our results depends on two key model parameters. The first is $\sigma_{p,\tau}$, which determines the strength of sorting by ability into different school groups, which is reflected in $E[a|s, \tau]$ in the average wage equation. The second is θ, which determines the impact of cognitive ability on wages. In general, the smaller is $\sigma_{p,\tau}$ and the larger is θ, the larger is the quantitative role for mean ability in explaining observed wage patterns. Other parameters such as β or R matter little or not at all for our quantitative results. Perfect sorting by $p + a$ is critical for this simplification.

2.6 Model Discussion

Our model admits other interpretations that yield similar results. One useful reinterpretation follows Manski (1989). Students still possess cognitive ability a, which makes school easier and raises wages, just as in our baseline interpretation. However, students have no tastes for schooling. If they knew their own ability, they would perfectly sort by ability into school attainment. Imperfect sorting in this model comes from the assumption that students are imperfectly informed about their cognitive ability, with p representing signal noise and $p+a$ representing their signal of their own cognitive ability. Students with better signals of cognitive ability further their education, because they anticipate that schooling will be relatively painless. This reinterpretation generates the same prediction of perfect sorting by $p + a$. Because of this the calibration and results from this alternative model would be *identical* to those derived from our baseline model.

Our model does assume only a single stand-in friction that prevents perfect sorting by cognitive ability. An alternative approach taken elsewhere is to model multiple frictions in detail (Cunha, Heckman, and Navarro 2005, Navarro 2008). Doing so would complicate our model and identification. However, the primary impediment is that we lack sufficient historical data to calibrate multiple frictions in detail.

An alternative friction to perfect sorting by ability that is not nested by our setup is borrowing constraints. Borrowing constraints differ from tastes because they are asymmetric: they prevent some high-ability students from furthering their education, but have no effect on low-ability students. By contrast, variation in tastes causes some high-ability students to drop out, but it also causes some low-ability students to attain high levels of education. The literature has not arrived at a consensus about the quantitative importance of borrowing constraints. Cameron and Taber (2004) and Stinebrickner and Stinebrickner (2008) find no evidence of borrowing constraints in the United States for recent cohorts of college attendees. We have little evidence as to whether credit constraints were quantitatively important for earlier cohorts. However, evidence gathered in Herrnstein and Murray (1994) suggests that low-ability students are becoming less likely to attend college over time. This information is consistent with a decline in the dispersion of tastes, but not a model featuring only a relaxation of borrowing constraints over time.

3 Calibration to the NLSY79 and IQ Test Scores

Our model is a parsimonious formalization of the basic challenge. Mean wages are affected by skill prices, human capital, and mean ability, none of which are directly observable. In the model, two key parameters determine how important mean ability is for explaining wage patterns. The first is $\sigma_{p,\tau}$, which determines the size of mean ability gaps between school groups; the second is θ, which determines the impact of ability on wages. Now we turn to the question of how IQ test scores can help us calibrate these parameters and quantify the role of ability for wage patterns.

Our primary data source is the NLSY79. The NLSY79 has two properties that make it ideal for our purposes. First, it is a representative sample of persons born between 1957 and 1964. Second, it includes information about the wages, school choices, and IQ test scores of individuals in the sample. Most other data sets are deficient along one of these dimensions. For example, information on SAT scores are drawn from a non-representative sample, while common data sets such as the population census do not include information on IQ test scores.

We restrict our attention to white men. We exclude women for the typical reason that only a selected sample of women work. Further, the selection process itself may be changing over time. We also exclude minorities because we eventually want to turn our attention to earlier cohorts, for whom discrimination limited school attainment choices and wages. We include members of the supplemental samples, but use weights to offset the oversampling of low income persons. Since everyone born in the NLSY79 is from a narrow range of cohorts, we group them together and call them jointly the 1960 cohort. In this section we focus on the 1960 cohort and provide some initial cross-sectional results; in the next section we generate time series results.

We use as our measure of IQ test score their Armed Forces Qualifying Test (AFQT) score. The AFQT is a widely recognized as a cognitive test and AFQT scores are highly correlated with the scores from other IQ tests. For each person, we construct real hourly wage at age 40, educational attainment, and AFQT score. We use regressions to remove the age effects from AFQT scores, then standard normalize them. Details are available in the Appendix.

Since IQ test scores play a central role in our analysis, it is important to be precise about how we interpret them. We think of IQ test scores as noisy, scaled proxies for ability, $\tilde{IQ} = \eta\,(a + \tilde{\varepsilon}_{IQ})$, where η is an unknown scaling factor and ε_{IQ} is a normal random variable with mean 0 and standard deviation σ_{IQ}. We standard normalize IQ test scores to remove the scaling factor. Once standard normalized, IQ test scores and the noise term are given by:

$$IQ = \frac{a}{\sqrt{1+\sigma_{IQ}^2}} + \varepsilon_{IQ}$$

$$\varepsilon_{IQ} \sim \mathcal{N}\left(0, \frac{\sigma_{IQ}}{\sqrt{1+\sigma_{IQ}^2}}\right)$$

We now turn to using IQ scores to quantify the role of ability in wage patterns.

3.1 Results When IQ Test Scores Measure Ability Exactly

IQ test scores provide us with useful information on the role of ability in school choices and wages. Intuitively, we can use the degree of sorting by IQ test scores into educational attainment as a proxy for the degree of sorting by ability into educational attainment, which helps identify $\sigma_{p,1960}$. Likewise, we can use the effect of IQ test scores on wages as a proxy for the effect of ability on wages, which helps identify θ. To see how this process works, we begin with a special case: $\sigma_{IQ} = 0$. In this special case, IQ test scores measure ability exactly, and our identification and results are straightforward.

Table 1: Returns to IQ in the NLSY79

Dependent variable: log-wages	
β_{IQ}	0.104
	(0.017)
γ_{HS}	0.17
	(0.06)
γ_{SC}	0.35
	(0.06)
γ_{C+}	0.69
	(0.07)
Observations	1942
R^2	0.24

We begin by identifying θ. The wage generating process in our model is:

$$\log[w(s, q, v)] = \theta a + z(s, \tau + v - 1) + h(s, v).$$

Generally, we do not have direct information on a. Instead, we have measured IQ test scores. Our empirical counterpart to this regression is to regress wages at age 40 on IQ test scores and a full set of school dummies:

$$\log(w) = \beta_{IQ} IQ + \sum_s \gamma_s d_s + \varepsilon_w. \tag{4}$$

IQ is the individual's standard normalized IQ test score, and β_{IQ} is the coefficient associated with that score. d_s is an indicator variable that takes a value of 1 if the individual has school attainment s. Since we focus on wages at age 40, γ_s captures the joint wage impact of skill prices and human capital; we are unable to separate the two. ε_w is assumed to be a normal random variable that captures factors such as shocks or luck that affect wages but are not associated with IQ test scores, skill prices, or human capital.

Table 1 shows the results of our regression of log-wages on IQ test scores as implemented in the NLSY79. The return to IQ is $\beta_{IQ} = 0.104$. This will be our baseline estimate of the return to IQ test scores for the remainder of the paper; what will change is how we interpret it. In the case where IQ measures ability exactly, the interpretation is straightforward: $\beta_{IQ} = \theta$. A one standard deviation rise in ability (which is the same as IQ test score) raises log-wages by 10.4 percentage points. This is the first key parameter for determining the importance of composition effects.

Table 2: Conditional Distribution of IQ Test Scores Given Schooling in the NLSY79

	IQ Quartile			
School Attainment	1	2	3	4
<HS	86%	12%	2%	0%
HS	42%	34%	19%	5%
SC	18%	32%	31%	19%
C+	1%	11%	29%	59%

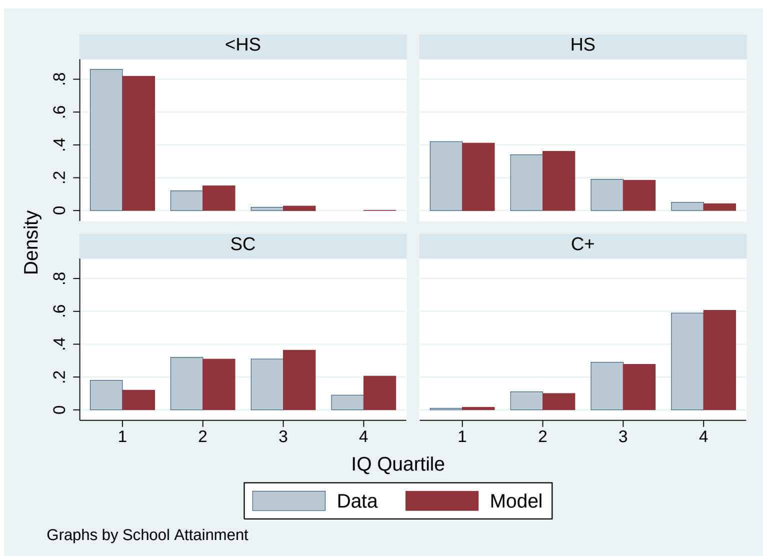
The second feature of the data that is important for our results is the degree of sorting by ability into educational attainment. Table 2 provides some evidence that school groups are strongly sorted by IQ test score. Each row of the table corresponds to one of our four school groups. The four columns give the conditional probability of someone with that school level having an IQ test score in each of the four quartiles of the distribution. The vast majority (86%) of high school dropouts are from IQ quartile 1, while 76% of high school graduates have below-median IQ test scores. On the other hand, 88% of college graduates have above-median IQ test scores.

In the case where IQ measures ability exactly, these facts imply that school groups are strongly sorted by ability. It is again straightforward to use this information. We can compute $E(IQ|s)$ from the NLSY79. In this special case, $E(a|s) = E(IQ|s)$. When combined with our estimate that $\theta = 0.104$, we can calculate the role of effective ability gaps in explaining wage premiums, which is given by $\theta\left[E(a|s) - E(a|s')\right] = \beta_{IQ}\left[E(IQ|s) - E(IQ|s')\right]$.

We also calibrate our model to the NLSY79 and use it to produce results on the role of effective ability gaps in explaining wage premiums. At this point the exercise is not strictly necessary because, as highlighted above, estimates of β_{IQ} and $E(IQ|s)$ are sufficient for these results. However in doing this exercise we hope to build some intuition for the general calibration procedure, which will be necessary for subsequent exercises. We also want to show that in this case the calibrated model produces results very similar to to the simpler calculation, which suggests to us that the calibration passes a basic test of reasonableness.

We now outline our calibration procedure. In this and all subsequent calibrations we treat $\chi(s, 1960)$ as a set of free parameters that we vary so that we fit schooling by cohort exactly. The reason is that our quantitative results are sensitive to getting educational attainment right; small deviations in the model-predicted school attainment can generate important differences in the quantitative predictions. We find it more straightforward to fit the attainment exactly. Given that we do so, we use our model as a measurement device to study the implied importance of effective ability. Then there are three remaining parameters that do not drop out of the model: σ_{IQ}, θ, and $\sigma_{p,1960}$. We have set $\sigma_{IQ} = 0$. Our calibration method chooses $\theta = 0.104$ so that the return to IQ test scores in the model matches the same statistic in the data. Finally, we choose $\sigma_{p,1960}$ so that the model fits the sorting by IQ test scores into educational attainment as closely as possible. Our model is deliberately parsimonious, and yet this approach is quite successful. Figure 2 compares the sorting in the data and the sorting predicted by the model for the best fit of $\sigma_{p,1960} = 0.87$. The model is able to generate sorting quite comparable to the data. The only significant discrepancy is that the model-generated distribution for those with some college has too many people with above-average IQ test scores and too few with below-average IQ test scores. Otherwise the fit between model and data is quite close, which suggests that the model will generate mean IQ test score gaps comparable to the data. We now verify that this is the case.

Figure 2: Model-Predicted and Actual Distribution of IQ Given Schooling



Our cross-sectional results measure the role of effective ability gaps in explaining observed school wage premiums. Table 3 shows our results. Each row contains the results from comparing high school graduates to one of the three remaining school groups. Of these, we are particularly interested in the college–high school comparison, since the college wage premium receives so much attention in the literature. Columns 2 and 3 gives the effective ability gaps that we find via direct calculation as well as those generated by the calibrated model; they are nearly identical. We view this fact as a useful check on the model and a way to show that the model does not generate any unusual predictions. Further, the

Table 3: Results when IQ Tests Measure Ability Exactly

School	Effective Ability Gap		Wage Gap
Comparison	Calculation	Model	Data
<HS–HS	-0.08	-0.08	-0.24
SC–HS	0.06	0.07	0.18
C+–HS	0.14	0.15	0.52

gaps are economically large. To help make this point, we provide in the final column the actual school wage premiums from the US Census.[6] Differences in mean ability account for roughly one-quarter to one-third of the wage premiums, with slightly smaller results for the college wage premium.

These initial results help address the puzzle of Heckman, Lochner, and Todd (2006) and Heckman, Lochner, and Todd (2008). They find that in standard human capital models, the college wage premium for recent cohorts is difficult to reconcile with less than one-third of the recent cohorts graduating from college, unless one incorporates some substantial uncertainty or a large "psychic cost" of attending schooling. Our results help reduce this puzzle modestly by pointing out that some of the apparently high college wage premium at age 40 is actually attributable to the gap in mean ability between college and high school graduates; the true private return to college is smaller than the observed wage gap.

3.2 Results When IQ Test Scores Measure Ability With Noise

Our initial results can be derived without calibrating the model. However, the model enables us to undertake two additional exercises. The first is to consider the case where IQ test scores measure ability with noise. In this case, the mean IQ test score for different school groups is not the same as the mean ability, so we cannot measure mean ability gaps directly. Instead, we use the calibrated model to quantify the role of ability.

Before discussing the exact calibration procedure, it is useful to see why allowing for noise in IQ test scores is likely to be important. The main reason is that we use the log-wage return to IQ test scores to help identify θ. In the case where IQ test scores measure ability exactly, then in fact $\theta = \beta_{IQ} = 0.104$. However, if IQ test scores measure ability with noise, then our empirical regression suffers from attenuation bias. The standard result that applies in this case is that $\theta > \beta_{IQ} = 0.104$. Hence, a given gap in mean abilities

[6]We use the US Census for wages to be consistent with the results of Section 4. There we examine the importance of mean ability for earlier cohorts, for which NLSY79 wage data are not available. Details of the wage measurement are available in the Appendix.

will lead to a larger effective ability gap, which in turn accounts for more of observed wage premiums. Our goal now is to quantify this analysis: how much noise is there likely to be in IQ test scores, and how much more important is effective ability in accounting for wages?

The primary challenge of implementing a model where IQ test scores are noisy is that we do not have direct evidence of how well IQ test scores measure ability. The obvious reason is that ability itself is not measured; if it were, we would not need to use IQ test scores as a proxy for ability. However, we will establish that we can make inferences that enable us to bound usefully the noise in IQ test scores. We begin by demonstrating how to construct a lower bound on the noise in IQ test scores.

IQ test scores could measure ability with noise for two basic reasons. First, they could lack *validity*: it could be that IQ test scores measure a mixture of traits, of which a is only one component, or that IQ test scores measure only a subset of the components of a. Given that a is unobservable, there is little we can do to quantify this point. The second reason for noise is that IQ test scores are not perfectly *reliable*: repeatedly administering similar or identical IQ tests yields different results. We can make progress in this direction. We construct our lower bound on the noise in IQ test scores by requiring an IQ test score not be a better predictor of ability than it is of other IQ test scores.

To quantify this statement, recall that we think of IQ test scores as noisy, scaled proxies for ability. We ignore any potential issues stemming from a lack of validity and focus instead on a lack of reliability. In this context it is natural to think of the noise in IQ tests ε_{IQ} as being an independent, test-specific draw. Then the correlation between two IQ test scores for a given individual is $\left(1+\sigma_{IQ}^2\right)^{-1}$. We have ample evidence on the magnitude of this correlation. Herrnstein and Murray (1994, Appendix 3) document the correlation between AFQT scores and scores from six other IQ tests taken by some NLSY79 individuals. The correlations range from 0.71 to 0.9, with a median score of 0.81.[7] Cawley, Conneely, Heckman, and Vytlacil (1997) show that the correlation between AFQT scores and the first principal component of the ASVAB scores is 0.83.

Putting these correlations together, we use $\left(1+\sigma_{IQ}^2\right)^{-1}=0.8$. In turn this suggests a lower bound $\sigma_{IQ}\geq 0.5$.[8] If IQ test scores were any more precise as measures of ability,

[7]A slight complication arises from the fact that Herrnstein and Murray compute correlations between percentile ranks rather than raw scores. We conducted simulations to verify that this has only a minor quantitative effect on the resulting correlation.

[8]A similar approach is taken by Bishop (1989) to estimate the measurement error in the PSID's GIA score. Based on the GIA's KR-20 reliability of 0.652, Bishop's result implies $\sigma_{IQ}=0.73$, which would imply a larger role for ability than what we find here. In fact, we construct an upper bound for σ_{IQ} that is lower than this value below, suggesting that Bishop's results may have counterfactual implications for wages.

then the correlation between scores from different tests should be higher. This calculation produces a lower bound because we have ignored any possible issues with validity. We use this lower bound by fixing $\sigma_{IQ} = 0.5$ in the model, then calibrating θ and $\sigma_{p,1960}$ to fit the return to IQ and the school-IQ sorting as well as possible. We are able to hit the former moment exactly. We showed earlier that even with only a single parameter $\sigma_{p,1960}$ we are able to replicate the school-IQ sorting closely (Figure 2); that continues to be the case here and throughout the remainder of the paper. We do not show the remaining figures to conserve space.

We also want to establish an upper bound on the plausible noise in IQ test scores. The purpose of this bound is not to argue that the true results are at some midpoint of the lower and upper bounds. Instead, we establish an upper bound to show that a bounding argument in this case is effective in the sense that the range of results is fairly narrow. Given this fact, it is innocuous to use the lower bound as our benchmark, which we do. Further, we will show that the results for the lower bound are already large relative to the wage patterns in the data, which reinforces our decision.

To derive an upper bound, we impose plausible limits on the size of the effects that we find. More noise in IQ test scores implies a larger attenuation bias in the regression of wages on IQ, a larger value for θ, and larger effective ability gaps between school groups. At some point the implied effective ability gaps become implausibly large. One natural benchmark is that an effective ability gap should not be bigger than the corresponding wage premium. If it were, this would imply a negative private return to going to school longer, which would seem inconsistent with simple optimization on the part of the students who achieve that attainment in the data.

Implementing the upper bound requires us to iteratively calibrate the model. We guess a particular value of σ_{IQ}. We then calibrate the θ and $\sigma_{p,1960}$ to fit the return to IQ and the school-IQ sorting as well as possible. Finally we compute the model's predicted effective ability gaps $\theta(E[a|s] - E[a|s-1])$ and compare them to the corresponding wage premiums in the data. If all effective ability gaps are smaller than the corresponding wage premiums then we guess a larger value for σ_{IQ} and restart the process; if any effective ability gap is larger than the corresponding wage premium then we guess a smaller value for σ_{IQ} and restart the process. We repeat until we find the σ_{IQ} so that one effective ability gap is exactly equal to the corresponding wage premium and all other effective ability gaps are smaller than their corresponding wage premiums.

Table 4 summarizes the results of our bounding exercises. Rows 2 and 3 give the value of the calibrated parameters while rows 4–6 give the results from different school

Table 4: Cross-Sectional Results when IQ Test Scores Measure Ability With Noise

	Model: Effective Ability Gap			Wage Gap
	IQ = a	LB	UB	
σ_{IQ}	0.00	0.50	0.68	
θ	0.104	0.155	0.228	
<HS–HS	-0.08	-0.14	-0.22	-0.24
SC–HS	0.07	0.11	0.18	0.18
C+–HS	0.15	0.25	0.39	0.52

comparisons. Column 2 repeats the results for the case where IQ measures ability exactly, for reference. Column 3 gives the results for the lower bound. We find this lower bound by fixing $\sigma_{IQ} = 0.5$; in this case a modest increase in θ is required for the model to generate a return to IQ test scores of 0.104 as seen in the data. This larger value of θ in turn yields larger effective ability gaps, 57–75% higher than in the case where IQ measures ability exactly. An alternative way to judge the size of effective ability gaps is by comparing them to observed wage premiums, given in column 5. Effective ability gaps account for at least 48% of observed wage premiums, and more than half of the wage premium for high school dropouts and those with some college. These large results go further towards reducing the puzzle that it is hard to reconcile the high college wage premium with a low college completion rate in a human capital model (Heckman, Lochner, and Todd 2006, Heckman, Lochner, and Todd 2008). Our results are consistent with those of Bowlus and Robinson (2010). They find that most of cross-sectional wage premiums can be accounted for through differences in skill quantities rather than skill prices. We focus on one component of skill quantities, namely innate abilities, and find that they can account for about half of cross-sectional wage premiums.

Finally, column 4 includes the results at the upper bound. We find that this upper bound binds for the SC–HS comparison at a value of $\sigma_{IQ} = 0.68$. Comparing column 3 to column 4 shows that the lower bound and the upper bound are already fairly similar in terms of the parameterizations and the results. In particular, the results for the upper bound are less than twice those for the lower bound. In the next section we will tighten the upper bound even further so that the difference between the lower and upper bounds is even smaller. We now turn to the time series calibration.

4 Calibration to the Time Series

The first contribution of the model is that it enables us to generate results for the case where IQ test scores measure ability with noise. The second contribution of the model is that it enables us to generate results for the time series. While the NLSY79 provides us with excellent data on schooling, wages, and IQ test scores for the 1960 cohort, no comparable data set exists for earlier cohorts. At the same time, the dramatic expansion of education and the growing IQ test score gap between those who enroll in college and those who do not lead us to believe that composition effects may play a large role in the wage patterns of the twentieth century. In this section we calibrate the model to see if this is the case and, if so, to quantify the magnitude of the effects.

4.1 Calibration

Our time series calibration follows the same basic outline as the cross-sectional calibration. What this means is that we calibrate both a lower bound and upper bound for σ_{IQ}, and provide the parameters and the results for each case. We now discuss these calibrations in more detail.

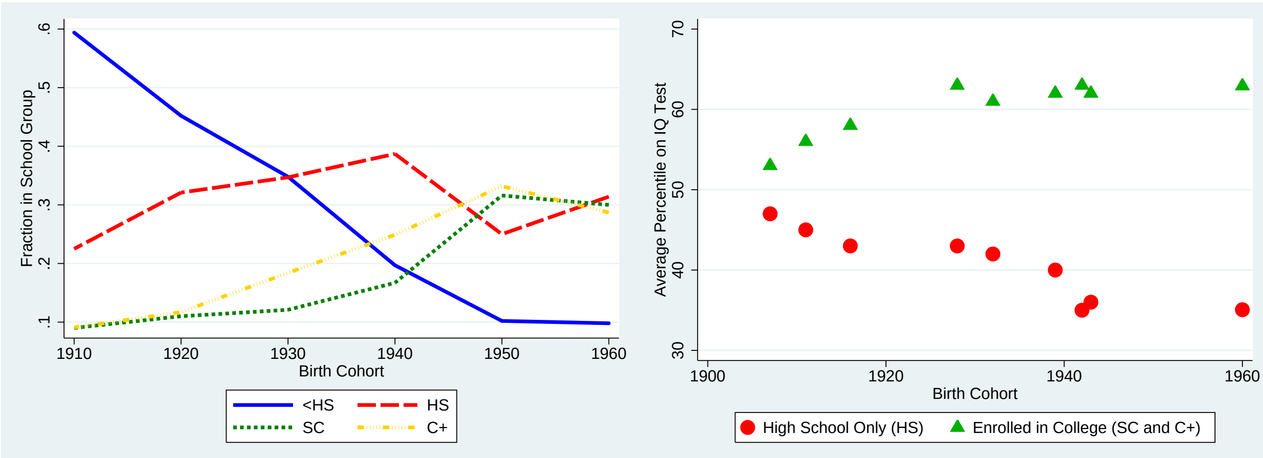
Our lower bound is still $\sigma_{IQ} = 0.5$. Given this moment, we use the remaining parameters to fit the model to the data. In particular, we still choose $\chi(s, \tau)$ as a free parameter to fit the expansion of schooling shown in figure 3a. Likewise, we use $\sigma_{p,\tau}$ to fit the degree of sorting by IQ test score into educational attainment, shown in figure 3b. For the 1960 cohort this information comes from the NLSY79. For earlier cohorts, we calibrate $\sigma_{p,\tau}$ to the standardized data on sorting provided by Taubman and Wales (1972). Since they do not report IQ test scores for all birth cohorts, we interpolate between their data points. Note that the Taubman and Wales data suggest that students are becoming more strongly sorted over time. Our model can replicate this observation if the dispersion of tastes is declining over time, so that ability plays a larger role in school choices for later cohorts.[9] Finally, we calibrate θ so that the model-predicted return to schooling matches that of the data, $\beta_{IQ} = 0.104$.

We use the same basic iterative procedure as before to find the upper bound. We guess a particular value of σ_{IQ}. We then calibrate the remaining parameters to fit the model to the data. Given the full set of parameters, we look at the model's predictions for effective ability gaps. If all effective ability gaps are smaller than their corresponding wage gaps, we

[9] An alternative interpretation is that students are imprecisely informed about their own ability, but that they are becoming more precisely informed over time; see section 2.6 for further discussion.

Figure 3: Time Series Calibration Strategy

(a) Use $\chi(s,\tau)$ to Fit Expansion (b) Use $\sigma_{p,\tau}$ to Fit Sorting



start again with a larger σ_{IQ}; if any effective ability gap is larger than its corresponding wage gap, we start again with a smaller σ_{IQ}. We expect this upper bound to be closer to the lower bound than in the previous section. The reason is that in the previous section we checked this bound only for the 1960 cohort, whereas now we check it for the 1910–1960 cohorts, which gives more wage premiums that may potentially bind the size of our effective ability gaps.

4.2 Results for Lower Bound

We begin by presenting the results for the lower bound in detail; we take these results to be our benchmark findings and show the comparison to the upper bound in the next section. Table 5 shows the full set of calibrated parameters for the lower bound. The value for θ is the same as in the cross-sectional calibration. The main new point to note is that the calibrated dispersion of tastes declined substantially between the 1910 and 1960 cohorts, indicating that ability played a much greater role in determining who continued to college for the 1960 cohort. Workers in this model sort perfectly by $p + a$ and the variance of a is set at 1 throughout. Then for the 1910 cohort variance in cognitive ability accounted for just 16% of the variance in $p + a$, while for the 1960 cohort it accounted for 72%.

These changes in sorting, along with the expansion of education, imply large changes in the mean ability of the four school groups. Figure 4 shows the model-implied evolution of the distribution of ability conditional on schooling. Figure 4a illustrates the degree of

Table 5: Calibrated Parameters for the Lower Bound

Parameter	Role	Value
σ_{IQ}	Noise in IQ Tests	0.50
θ	Effect of Ability on Wages	0.155
$\sigma_{p,1960}$	Dispersion of Preferences	0.62
$\sigma_{p,1950}$	Dispersion of Preferences	0.81
$\sigma_{p,1940}$	Dispersion of Preferences	1.04
$\sigma_{p,1930}$	Dispersion of Preferences	1.19
$\sigma_{p,1920}$	Dispersion of Preferences	1.39
$\sigma_{p,1910}$	Dispersion of Preferences	2.27

sorting found for the 1960 cohort in the NLSY79. There are clear differences in the mean of the ability distribution between each of the four school groups, and almost no overlap between the distributions for high school dropouts and college graduates.

Figure 4b illustrates a particular counterfactual: it shows the distribution of ability conditional on schooling that would have applied if we held the 1960 cohort's attainment fixed, but gave them the dispersion of tastes that we calibrated for the 1910 cohort. By comparing figures 4a and 4b we can see the effect of the increase in sorting isolated from the effect of the expansion of education. The distributions in figure 4b have very small mean differences, particularly for those who at least graduate high school. Further, the distributions overlap substantially.

Finally, figure 4c illustrates the model-implied distributions for the 1910 cohort. By comparing figures 4b and 4c we can see the effect of the expansion of education. The mean of each distribution is shifted left in figure 4b by the rise in schooling. To see why this happens, consider the distribution for high school graduates. Over time, attainment rises. In the model, this happens because high school graduates with relatively high levels of $p+a$ in later cohorts start attempting college. At the same time, some people with relatively low $p + a$ in later cohorts will complete high school instead of dropping out. Both of these effects act to reduce the average ability of high school graduates.

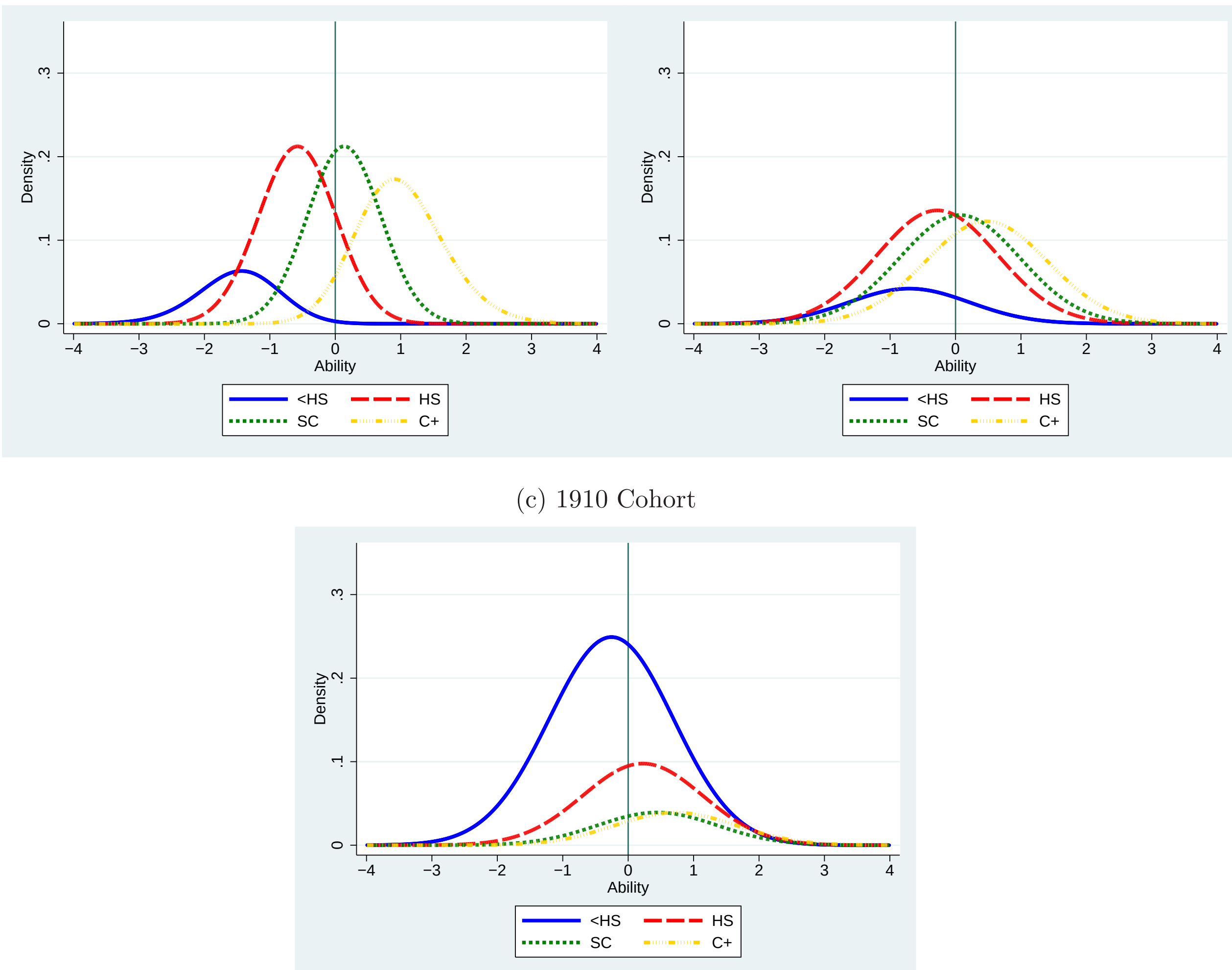
Comparison of figures 4a and 4c shows the combined effect of the expansion of education and the change in sorting. The leftward shift in ability for high school dropouts is particularly pronounced because both effects move in the same direction, toward a decline in ability. On the other hand there is hardly any change in the peak of the distribution for college graduates, as the expansion of college is in large part offset by the change in sorting. Intuitively, it is possible to expand college enrollment without lowering the mean

Figure 4: The Distribution of Ability Conditional on Schooling

(a) 1960 Cohort

(b) Counterfactual: 1960 Cohort with 1910 Sorting

(c) 1910 Cohort



ability of college graduates if stronger sorting by IQ test scores makes it possible to identify high-ability students who in earlier cohorts did not attend college. This point will be central to our subsequent results.

We have two main sets of time series results. First, we examine how changes in effective ability between cohorts have affected wage growth. These results are presented in table 6. The second column gives the model-predicted change in effective ability between the 1910 and 1960 cohorts for each of the four school groups. In the fourth column we give the measured wage growth conditional on schooling, taken from census data. Given the observed wage growth and the model-implied change in mean ability, we back out the implied growth in $h + z$ in the third column. This column measures the growth in skill

Table 6: Changes in Mean Ability and Wages, 1910–1960 Cohorts

	Model-Predicted Change		Data Change
	Effective Ability	$h+z$	Wage
<HS	-0.19	0.41	0.22
HS	-0.13	0.41	0.29
SC	-0.05	0.35	0.30
C+	0.04	0.39	0.43

Table 7: Changes in Mean Ability Gaps and Wage Premiums, 1910–1960 Cohorts

	Model-Predicted Change		Data Change
	Effective Ability Gap	$h+z$ Gap	Wage Premium
<HS–HS	-0.07	0.00	-0.06
SC–HS	0.08	-0.06	0.02
C+–HS	0.17	-0.02	0.15

prices and human capital, which is also the wage growth that would have been observed if mean ability had remained constant. Changing ability had the largest effect for high school dropouts: the 19 percentage point decline in effective ability caused observed log wage growth to be roughly one-half of the growth in $h + z$. The effect for high school graduates was smaller and for those with some college smaller still. For college graduates we find effective ability actually *rose* between the 1910 and 1960 cohorts, resulting in wage growth that overstates the growth in $h+z$. Hence our model can generate a wage slowdown that affects the less educated groups more.

For our second set of time series results we examine how changes in effective ability gaps have affected wage premiums between the 1910 and 1960 cohorts. These results are presented in table 7. The second column gives the model-predicted change in effective ability gaps as compared to high school graduates for each of the three remaining school groups. Given the observed wage premium growth in the fourth column, we again back out the implied growth in $h+z$ gaps in the third column. Changing ability had the largest effect for college graduates, relative to high school graduates. In fact, we find that the *entire* rise in the college wage premium can be attributed to the fact that effective ability for college graduates rose modestly and for high school graduates declined between the 1910 and 1960 cohorts. Likewise, we find that the entire change in the high school dropout-high school graduate premium can be attributed to changes in mean ability for the two groups.

To summarize, our results suggest that changes in ability have slowed observed wage

Table 8: Results for Lower and Upper Bounds of IQ Noise

	Model: Effective Ability			Data: Wages
	$IQ = a$	LB	UB	
σ_{IQ}	0.00	0.50	0.66	
θ	0.104	0.155	0.217	
C+–HS	0.15	0.25	0.37	0.52
ΔHS	-0.08	-0.13	-0.19	0.29
ΔC+–HS	0.10	0.17	0.25	0.15

growth for most school groups as mean ability has declined. Further, our most important result is that the entire rise in the college wage premium can be explained by changes in the relative ability of college and high school graduates. Our results use a different methodology but arrive at a similar conclusion as Bowlus and Robinson (2010), who find that 72% of the rise in the college wage premium between the years 1980 and 1995 can be attributed to changes in the quantity of labor services provided by college relative to high school graduates. We conclude this section by noting that these results are actually the lower bound of what is plausible, which we take as our benchmark. We now turn to showing the entire range of possible results.

4.3 Range of Results

We use the iterative procedure outlined in the previous section to find the upper bound on σ_{IQ}, calibrate the remaining parameters, and derive the model predictions. Rather than present all possible results, we focus on a few select results, presented in table 8. These results are for the college wage premium for the 1960 cohort; the growth in wages for high school graduates between the 1910 and 1960 cohorts; and the change in the college wage premium between the 1910 and 1960 cohorts, presented in rows 5–7. The fifth column shows the data on wages, while columns 2–4 show the model-implied role for effective ability given different values of σ_{IQ}. The main message from these rows is that the upper bound is roughly 50 percent larger than the lower bound in terms of θ and in terms of each of the three key wage statistics. This means that it accounts for roughly 50 percent more of the college wage premium, the slowdown in observed high school wage growth, and the rise in the college wage premium. The upper bound is overall quite close to the lower bound, suggesting that our bounding argument restricts the range of potential results successfully. We now turn to robustness analysis to test what key assumptions drive our results.

Table 9: Robustness: Constant Sorting by Cohort

	Model: Effective Ability				Data: Wages
	Baseline		Constant Sorting		
	LB	UB	LB	UB	
σ_{IQ}	0.50	0.66	0.50	0.53	
θ	0.155	0.217	0.155	0.165	
C+–HS	0.25	0.37	0.25	0.27	0.52
ΔHS	-0.13	-0.19	-0.16	-0.18	0.29
ΔC+–HS	0.17	0.25	0.08	0.09	0.15

5 Robustness

In the previous section we established our three key results. At the lower bound of the range, our model accounts for about half of the college wage premium for the 1960 cohort as well as the entire rise of the college wage premium between the 1910 and 1960 cohorts. It also predicts a slowdown in wages conditional on schooling that has a stronger effect on less educated groups. We now perform robustness analysis. We focus on two key mechanisms. First, we provide results for the case where sorting is held constant. These results can be viewed as a decomposition, but they also show that more than half of our key results survive even in the absence of changes in sorting. Second, we provide results for the case where $\beta_{IQ} < 0.104$. This is the key moment for our calibration so we find it worthwhile to consider smaller values. We also briefly consider the possible importance of changes in the ability distribution over time.

5.1 Constant Sorting

For our first robustness check we re-calibrate the model with constant sorting. That is, we fix $\sigma_{p,\tau} = \sigma_{p,1960}$ for all cohorts. We do so both to separate the quantitative results due to the expansion of education from the results due to the increase in sorting, and to test the robustness of our model to information derived from potentially mismeasured historical data. Other than holding $\sigma_{p,\tau}$ fixed, the details of the calibration are as in section 4. We continue to calibrate θ and $\sigma_{p,1960}$ to the β_{IQ} and IQ–school sorting from the NLSY79. We again provide results for both the lower and the upper bound. The lower bound is still given by $\sigma_{IQ} = 0.50$, but we have to recalibrate the upper bound since constant sorting changes the effective ability gaps for earlier cohorts.

The results are presented in table 9 in the same format as table 8. We present again

Table 10: Robustness: Lower Returns to IQ Test Scores

	Model: Effective Ability				Data: Wages
	Baseline		$\beta_{IQ} = 0.07$		
	LB	UB	LB	UB	
σ_{IQ}	0.50	0.66	0.50	0.78	
θ	0.155	0.217	0.104	0.205	
C+–HS	0.25	0.37	0.17	0.37	0.52
ΔHS	-0.13	-0.19	-0.08	-0.19	0.29
ΔC+–HS	0.17	0.25	0.11	0.25	0.15

the results for the baseline model as well as those for the model for constant sorting. We note two key findings. First, the model with constant sorting generates smaller time series results. The quantitative reduction is modest for the change in wage levels and stronger for the change in the college wage premium; for the latter, our results are roughly one-half of those in the baseline model. This finding indicates that half of the model's predictions for the time series of the college wage premium stems from changes in sorting and half from the expansion of education; each is important. Likewise, half of our results would survive even if we adopted the extreme view that the historical data of Taubman and Wales (1972) should be disregarded entirely.

The second main finding of this table is that the upper bound collapses to lie almost exactly at the lower bound for the case with constant sorting. This happens because the constant sorting experiment assumes more sorting and larger effective ability gaps in earlier cohorts than does the baseline experiment. Because of this the model with constant sorting hits its upper bound for much smaller values of σ_{IQ}. The range of plausible results in this case is extremely narrow.

5.2 Lower Return to IQ

The key moment for our calibration is the log-wage return to IQ in the NLSY79, which we measure as $\beta_{IQ} = 0.104$. Our estimate is similar to other estimates in the literature that use the NLSY79 (see for example Mulligan (1999) table 6, or Altonji and Pierret (2001) table I). However, estimates based on other data sources differ. Bowles, Gintis, and Osborne (2001) collect 24 studies using different data sources. The mean return across studies was 7%, with substantial dispersion. In this section we examine the robustness of our results to using $\beta_{IQ} = 0.07$ as an input to our calibration.

Our calibration strategy is the same as the baseline case, except that for each possible

σ_{IQ} we calibrate θ to replicate $\beta_{IQ} = 0.07$. Table 10 gives the results in the same format as the previous two tables. The main finding is that the results for the lower bound are about one-third smaller than in the baseline case, while the results for the upper bound are the same. Note, however, that even for the lower bound of the robustness check $\beta_{IQ} = 0.07$, we still account for roughly three-quarters of the rise in the college wage premium and nearly one-third of the current college wage premium.

The fact that our results change at the lower but not the upper bound may seem unusual, but it arises from the difference in how we calibrate our lower and upper bounds. To find the lower bound we fix $\sigma_{IQ} = 0.5$ and calibrate θ so that the model generates a return to IQ test scores consistent with the data. If we lower the empirical target from $\beta_{IQ} = 0.104$ to $\beta_{IQ} = 0.07$, then we calibrate a lower θ and find smaller results. On the other hand, to find the upper bound we calibrate θ so that the model produces the largest plausible mean ability gaps, then calibrate the noise σ_{IQ} that would also allow the model to be consistent with the observed return to schooling. Hence if we lower the observed return to schooling this has only a small effect on θ; the calibration primarily adjusts σ_{IQ} upward. The higher level of noise ($\sigma_{IQ} = 0.78$ intend of 0.66) allows the model to be consistent with the lower target of β_{IQ} while still having the largest plausible effective ability gaps. Hence, the change in results at the lower bound but not the upper bound is a consequence of our calibration methodology.

5.3 The Flynn Effect

Our results so far have assumed that the distribution of abilities is time-invariant. There is, however, substantial evidence of a sustained rise in cognitive test scores throughout our time period, a phenomenon known as the Flynn effect (Flynn 1984, Flynn 2009). There is disagreement in the psychometric literature as to whether the Flynn effect represents real gains in cognitive skills, improvements in test-taking skills, or some other possibility (see Flynn (2009)). Here we explore the implications for our measurements if the Flynn Effect captures actual rises in cognitive ability.

Although it is still somewhat controversial, the evidence now seems to suggest that the rise in ability is a mean shift that affects all parts of the distribution more or less equally. In this case, our approach is simple. Flynn (2009) documents that average test scores on the WISC, a broad-based IQ exam, rose 1.2 standard deviations between 1947 and 2002, which corresponds roughly to our cohorts. He also conjectures (based on incomplete evidence) that test scores on the Raven's Progressive Matrix Exam, a test of spatial recognition, rose 1.83 standard deviations over the same years. We measure the implications in our model if these two changes represent real gains in cognitive ability.

The Flynn effect has modest implications for our work. An increase in the entire distribution of ability changes the mean ability of workers $E(a|s)$ by a constant amount, but does not affect the mean ability gaps $E(a|s) - E(a|s')$. In particular, a rise in ability of 1.2–1.83 standard deviations implies a rise in effective ability $\theta E(a|s)$ by 19–28 percentage points, if we use the benchmark $\theta = 0.155$. Any of these results implies that the mean ability of all school groups actually rose between the 1910 and 1960 cohorts, so that our model does not help explain the wage slowdown. Otherwise, the Flynn Effect has no important implications for our results about wage premiums because it affected different school groups equally.

6 Conclusion

Between the 1910 and 1960 cohorts the college wage premium widened substantially. Today the college wage premium is sufficiently large that it may be difficult to reconcile with a model of individual human capital investment. Most papers have tried to understand these movements as the result of changes in skill prices, roughly the wage per unit of labor. We break with this literature by asking instead whether changes in the units of labor per worker may be responsible. Large changes in the school attainment of workers and the degree of sorting by IQ test score into educational attainment suggest that the mean ability of workers with different education levels may have changed. The main purpose of this paper is to quantify these compositional effects and their impact on wages.

Our results suggest that much of the most important wage questions can be attributed to changes in the mean ability of students by school attainment. Our benchmark results can explain all of the rise in the college wage premium as well as half of the currently high college wage premium. Additionally, our model can help explain some of the wage slowdown as the result of declining mean ability conditional on schooling. Our robustness checks indicate that roughly half of our time series results come from the expansion of education and half from the increase in sorting. Our results would still be economically significant even if IQ test scores measured ability exactly or if we used more conservative moments for our calibration.

We have relied on a simplified model with one dimension of ability and one generic friction to sorting, the tastes of workers. Future work could make progress by developing a more detailed model of ability or the frictions that act to prevent stronger sorting by ability, and by finding more historical data on these forces for empirical use.

References

Acemoglu, D. (2002): "Technical Change, Inequality, and the Labor Market," *Journal of Economic Literature*, 40(1), 7–72.

Altonji, J. G., and C. R. Pierret (2001): "Employer Learning and Statistical Discrimination," *The Quarterly Journal of Economics*, 116(1), pp. 313–350.

Autor, D. H., L. F. Katz, and A. B. Krueger (1998): "Computing Inequality: Have Computers Changed the Labor Market?," *The Quarterly Journal of Economics*, 113(4), 1169–1213.

Bishop, J. H. (1989): "Is the Test Score Decline Responsible for the Productivity Growth Decline?," *The American Economic Review*, 79(1), pp. 178–197.

Bound, J., and G. Johnson (1992): "Changes in the Structure of Wages in the 1980's: An Evaluation of Alternative Explanations," *The American Economic Review*, 82(3), pp. 371–392.

Bowen, W. G., M. A. Kurzweil, and E. M. Tobin (2005): *Equity and Excellence in American Higher Education*. University of Virginia Press.

Bowles, S., H. Gintis, and M. Osborne (2001): "The Determinants of Earnings: A Behavioral Approach," *Journal of Economic Literature*, 39(4), pp. 1137–1176.

Bowlus, A. J., and C. Robinson (2010): "Human Capital Prices, Productivity, and Growth," Mimeo. University of Western Ontario.

Bureau of Labor Statistics; US Department of Labor (2002): "National Longitudinal Survey of Youth 1979 cohort," Produced and distributed by the Center for Human Resource Research, The Ohio State University. Columbus, OH.

Cameron, S. V., and C. Taber (2004): "Estimation of Educational Borrowing Constraints Using Returns to Schooling," *The Journal of Political Economy*, 112(1), pp. 132–182.

Card, D. (2001): "Estimating the Return to Schooling: Progress on Some Persistent Econometric Problems," *Econometrica*, 69(5), 1127–1159.

CARD, D., AND T. LEMIEUX (2001): "Can Falling Supply Explain the Rising Return to College for Younger Men? A Cohort-Based Analysis," *The Quarterly Journal of Economics*, 116(2), pp. 705–746.

CAWLEY, J., K. CONNEELY, J. HECKMAN, AND E. VYTLACIL (1997): "Cognitive Ability, Wages, and Meritocracy," in *Intelligence, Genes, and Success: Scientists Respond to The Bell Curve*, ed. by B. Devlin, S. Fienberg, D. Resnick, and K. Roeder. New York: Springer.

CUNHA, F., J. J. HECKMAN, AND S. NAVARRO (2005): "Separating uncertainty from heterogeneity in life cycle earnings," *Oxford Economic Papers*, 57, 191–261.

FINCH, F. H. (1946): *Enrollment Increases and Changes in the Mental Level*. Stanford University Press.

FLYNN, J. R. (1984): "The Mean IQ of Americans: Massive Gains 1932 to 1978," *Psychological Bulletin*, 95, 29–51.

——— (2009): *What Is Intelligence? Beyond the Flynn Effect*. Cambridge University Press.

GARRIGA, C., AND M. P. KEIGHTLEY (2007): "A General Equilibrium Theory of College with Education Subsidies, In-School labor Supply, and Borrowing Constraints," Mimeo. Federal Reserve Bank of St Louis.

GOLDIN, C., AND L. F. KATZ (2008): *The Race between Education and Technology*. Harvard University Press.

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

HECKMAN, J. J., L. J. LOCHNER, AND P. E. TODD (2006): "Earnings Functions, Rates of Return and Treatment Effects: The Mincer Equation and Beyond," in *Handbook of the Economics of Education*, ed. by E. Hanushek, and F. Welch, vol. 1, chap. 7, pp. 307–458. Elsevier.

——— (2008): "Earnings Functions and Rates of Return," *Journal of Human Capital*, 2(1), 1–31.

HERRNSTEIN, R. J., AND C. A. MURRAY (1994): *The Bell Curve: Intelligence and Class Structure in American Life*. New York: Free Press.

Juhn, C., D. I. Kim, and F. Vella (2005): "The expansion of college education in the United States: Is there evidence of declining cohort quality?," *Economic Inquiry*, 43(2), 303–315.

Juhn, C., K. Murphy, and B. Pierce (1993): "Wage Inequality and the Rise in the Return to Skill," *Journal of Political Economy*, 101(3), 410–442.

Katz, L. F., and K. M. Murphy (1992): "Changes in Relative Wages, 1963-1987: Supply and Demand Factors," *The Quarterly Journal of Economics*, 107(1), 35–78.

Laitner, J. (2000): "Earnings within Education Groups and Overall Productivity Growth," *The Journal of Political Economy*, 108(4), pp. 807–832.

Levy, F., and R. J. Murnane (1992): "U.S. Earnings Levels and Earnings Inequality: A Review of Recent Trends and Proposed Explanations," *Journal of Economic Literature*, 30(3), pp. 1333–1381.

Manski, C. F. (1989): "Schooling as experimentation: a reappraisal of the postsecondary dropout phenomenon," *Economics of Education Review*, 8(4), 305–312.

Mulligan, C. B. (1999): "Galton versus the Human Capital Approach to Inheritance," *The Journal of Political Economy*, 107(6), pp. S184–S224.

Murnane, R. J., J. B. Willett, and F. Levy (1995): "The Growing Importance of Cognitive Skills in Wage Determination," *The Review of Economics and Statistics*, 77(2), pp. 251–266.

Navarro, S. (2008): "Using Observed Choices to Infer Agent's Information: Reconsidering the Importance of Borrowing Constraints, Uncertainty and Preferences in College Attendance," Mimeo: University of Wisconsin.

NLS User Services (1992): "NLS79 Profiles of American Youth. Addendum to Attachment 106," http://www.nlsinfo.org/ordering/display_db.php3.

Ruggles, S., and M. Sobeck (2007): "Integrated Public Use Microdata Series: Version 2.0," Minneapolis: Historical Census Projects, University of Minnesota.

Stinebrickner, R., and T. R. Stinebrickner (2008): "The Effect of Credit Constraints on the College DropOut Decision: A Direct Approach Using a New Panel Study," *American Economic Review*, 98(5), 2163–2184.

Taubman, P., and T. Wales (1972): *Mental ability and higher educational attainment in the 20th Century*. Berkeley: National Bureau of Economic Research and Carnegie Commission on Higher Education.

Table 11: Summary statistics: Census data

	Census Year					
	1950	1960	1970	1980	1990	2000
Number of Observations	17,503	74,744	72,873	379,087	539,145	562,262
Fraction <HS	59%	45%	35%	20%	10%	10%
Fraction HS	23%	32%	35%	39%	25%	31%
Fraction SC	9%	11%	12%	17%	32%	30%
Fraction C+	9%	12%	18%	25%	33%	29%
$w_{<HS}$	9.6	13.0	15.3	12.4	12.7	12.0
w_{HS}	11.4	15.7	18.3	16.4	16.7	15.2
w_{SC}	13.4	18.0	21.1	18.6	19.3	18.1
w_{C+}	16.5	22.7	28.5	25.1	26.5	25.4
College wage premium	0.37	0.37	0.44	0.42	0.47	0.52

A Appendix

A.1 Census Data

Samples. The census is taken every ten years; we have data from many birth cohorts spanning multiple censuses. We focus our attention on the cohorts that are exactly 40 years of age in the 1950–2000 censuses, or the cohorts born every ten years from 1910–1960. Focusing on one age eliminates any problems associated with comparability of educational data and wages at different ages. We use the public-use micro data files available from Ruggles and Sobeck (2007). We use 1% samples for 1950–1970 and 5% samples thereafter. In 1950, only sample line individuals report wages and hours worked. This reduces the effective sample size to only one quarter of the 1960 sample. Table 11 shows descriptive statistics for each census year.

Educational attainment. Our measure of educational attainment is derived from the variables EDUCREC and HIGRADE (both detailed). For the 1990 and 2000 censuses we use the variable EDUCREC, which records the information on degrees obtained. We include those with GEDs among high school graduates and those with 2 year degrees among those with some college. For earlier censuses we have only the variable HIGRADE, which records the number of years of education a person has obtained. We classify persons into four school groups as follows: we call those with fewer than 12 years of schooling high school dropouts; those with exactly 12 years of schooling high school graduates; those with 13–15 years of schooling have some college; and those with at least 16 years of schooling, college graduates.

Cohorts that respond to both the HIGRADE and EDUCREC questions (in two different censuses) typically have different measured attainment for the two questions, even if they are old enough that significant changes in actual school attainment are unlikely. Goldin and Katz (2008) use Current Population Survey data to produce a more detailed concordance between EDUCREC and HIGRADE questions that they use between 1980 and 1990. We use the raw responses as discussed above, for two reasons. First, the concordance is likely to vary by year as the structure of education changes, especially within each of our four discrete categories. Second, our focus here is on the large-scale movements, such as the near-universality of high school graduation and the increase in college attendance. Since most of those identified as dropouts in the 1910 census report less than 11 years of schooling, we are confident that they did not achieve a high school degree, let alone start college. By contrast the 1960 cohort answered directly about degree completion, raising our confidence in these estimates. We therefore believe that these major trends are real and are not likely the artifact of changing data collection.

Wages. We calculate hourly wages as the ratio of wage and salary income (INCWAGE) to annual hours worked. Annual work hours are the product of weeks per year times hours per week. For consistency, we use intervalled weeks and hours for all years. Where available we use usual hours per week. Wages are computed only for persons who report working "for wages" (CLASSWKR) and who work between 520 and 5110 hours per year. All dollar figures are converted into year 2000 prices using the Bureau of Labor Statistics' consumer price index (CPI) for all wage earners (all items, U.S. city average).

A.2 NLSY79 Data

Sample. The sample includes white males. We drop individuals with insufficient information to determine their schooling. We also drop individuals who completed schooling past the age of 34 or who did not participate in the ASVAB aptitude tests. Observations are weighted.

Schooling. We divide persons into four school groups (less than high school, high school, some college, and college or more) according to the highest degree attained. Persons who attended 2-year colleges only are assigned the "some college" category. The last year in school is defined as the start of the first three year spell without school enrollment.

Wages. We calculate hourly wages as the ratio of labor income to annual hours worked. Labor income includes wages, salaries, bonuses, and two-thirds of business income. We delete wage observations prior to the last year of school enrollment or with hours worked outside the range $[520, 5110]$. We also delete wage observations outside the range $[0.02, 100]$ times the median wage. Wages are deflated by the CPI.

We remove from the wage data variation that is due to demographic characteristics not captured by our model. This is done by regressing log wages on schooling, experience, and region of residence. Separate regressions are estimated for each year and schooling group. The adjusted wage removes the effects of years of schooling (within school groups) and region.

For consistency with the Census data, we focus on wages earned at age 40. Since not all persons are interviewed at age 40, we interpolate these wages using data for ages 39 to 41.

AFQT. The measure of IQ in the NLSY79 is the 1980 Armed Forces Qualification Test (AFQT) percentile rank (variable R1682). The AFQT aggregates various ASVAB aptitude test scores into a scalar measure. The tests cover numerical operations, word knowledge, paragraph comprehension, and arithmetic reasoning (see (NLS User Services 1992) for details). We remove age effects by regressing AFQT scores on the age at which the test was administered.