Student Abilities During the Expansion of U.S. Education, 1950-2000*

Lutz Hendricks[†]   Todd Schoellman[‡]

October 18, 2010

Abstract

A large literature aims to understand variation in U.S. wages across school groups and over time. For this task, it is important to separate variation in skill prices from variation in worker characteristics that affect earnings ("abilities"). This is the goal of our analysis. Our main findings are:

1. Measured skill premiums substantially overstate skill price gaps across school groups. About 40% of the year 2000 college wage premium represents an ability premium.

2. About one-quarter of the increase in the college wage premium between 1950 and 2000 is due to the rising relative abilities of college graduates versus high school graduates.

JEL: I2, J24.

Key words: Education. Ability. Skill premium.

1 Introduction

Motivation.   A large literature documents the evolution of U.S. wages. Important facts high-lighted by this literature include the slow wage growth experienced, especially by less skilled workers, since 1973 (e.g., Levy and Murnane 1992) and the dramatic increase in the college wage premium since 1950 (Goldin and Katz, 2008). According to some estimates, the college wage premium has recently grown to levels that imply implausibly high rates of return to schooling.[1]

Each of these facts has motivated a literature exploring possible causes. For example, the rising skill premium has been attributed to skill-biased technical change (Katz and Murphy 1992; Bound and Johnson 1992; Autor, Katz, and Krueger 1998; Goldin and Katz 2008), international trade and migration, and changing labor market institutions (see Bound and Johnson 1992 and the survey by Levy and Murnane 1992).

When thinking about the causes of wage movements, an important question arises: to what extent do wages reflect skill prices as opposed to other worker characteristics that affect earnings? Much of the literature has abstracted from unobserved worker characteristics and treated wages as skill prices. We offer a complementary interpretation. Individual level data reveal a strong positive correlation between education, wages, and measures of cognitive skills (see Section 3.2). These correlations suggest that measured wages confound skill prices and worker abilities.

To fix ideas, write the mean log wage of school group s in year t as the sum of the unobserved skill price and the mean ability of the workers in that group: $w_{s,t} = z_{s,t} + E(a|s,t)$. In the data, mean worker ability increases with schooling: $E(a|s,t) > E(a|s-1,t)$.[2] It follows that measured skill premiums overstate the unobserved skill price premiums: $w_{s,t} - w_{s-1,t} > z_{s,t} - z_{s-1,t}$. Some part of the observed college wage premium is an ability premium.

As schooling expands over time, mean abilities within school groups decline: $E(a|s,t) < E(a|s,t-1)$.[3] As a result, measured wages grow more slowly than skill prices: $w_{s,t} - w_{s,t-1} < z_{s,t} - z_{s,t-1}$. Some part of the slow wage growth observed in recent U.S. data may be due to falling worker abilities.

The general point is that variation in measured wages confounds variation in skill prices and worker abilities. This poses a problem for attempts at understanding the evolution of U.S. wages. The purpose of this paper is to re-construct the U.S. post-war wage series to separate movements in skill prices from movements in worker abilities.

Approach.   Our task is complicated by the fact that neither skill prices nor worker abilities are directly observable. Abilities may be measured by test scores that capture cognitive skills.

---

[1] Heckman, Lochner, and Todd (2006) and Heckman, Lochner, and Todd (2008) estimate internal rates of return to schooling using generalized Mincerian earnings functions that far exceed typical interest rates on finanical assets or loans. Heckman, Lochner, and Todd (2006) argue that accounting for uncertainty and "psychic costs" of schooling may help account for these high estimated rates of return.

[2] The relationship between schooling and cognitive skills has been documented many times in literature. We document is again based on NLSY79 data in Section 3.2.

[3] We assume that mean abilities in the population are time invariant. The Flynn Effect (Flynn, 1984) suggests that mean abilities may rise over time. We discuss this possibility in Section 5.2.

However, test scores suffer from an unknown amount of measurement error and therefore contain limited information about the dispersion of abilities or their correlation with schooling. But these are key for measuring the wedge between observed wages and unobserved skill prices.

To solve this identification problem, we develop a model of school choice with heterogeneous abilities. The model features finitely lived individuals who choose from discrete school levels. Their objective is to maximize discounted lifetime utility. Choosing more education raises wages but incurs costs in the form of foregone earnings and direct costs. Individuals differ in two dimensions, which we call abilities and school costs. High ability enhances both wages and the payoffs to schooling. School costs are a stand-in for any individual traits that affect school choice but not wages.

A key question the model needs to resolve is: how large is the dispersion of abilities and how strong is its correlation with schooling? The gap between measured wages and skill prices is large when ability dispersion is high and school sorting by ability is strong. Heterogeneity in school costs introduces a friction, which permits the model to replicate the empirical correlation between abilities and schooling.

Calibration. We calibrate the model to match the joint distribution of schooling, wages, and cognitive test scores for the 1960 birth cohort, which we estimate from NLSY79 data (Bureau of Labor Statistics; US Department of Labor, 2002). We interpret Armed Forces Qualification Test (AFQT) scores as noisy measures of individual abilities. In contrast to much of the literature on school choice, we do not assume that AFQT scores measure abilities perfectly. Instead, we calibrate how precisely AFQTs measure abilities.

Results. Using the calibrated model, we simulate abilities, schooling and wages for the cohorts born in 1910 to 1960. We focus on wages at age 40 to avoid complications that arise when workers of different ages are compared. For each cohort, the model measures the distribution of abilities by level of schooling. We infer skill prices from abilities and measured wages, estimated from the 1950 to 2000 waves of the U.S. Census, according to $w_{s,t} = z_{s,t} - E(a|s,t)$.

We organize our results around the three features of U.S. wages mentioned earlier. Our main findings are:

1. Measured skill premiums overstate skill price differences across school groups by at least 50%. In particular, of the 36% college wage premium observed in the year 2000, only 23% reflect skill price differences.

2. Revisions to the time-series changes in skill premiums are small, except for college graduates. Adjusting for changing abilities reduces the increase in the college wage premium from 23% to 18%.

3. Measured wage growth rates and skill prices grow at similar rates.

We examine the robustness of our findings along three dimensions. First, we examine alternative calibration targets. Second, we consider the possibility that school sorting by ability may have improved over time, as suggested by Taubman and Wales (1972) and Herrnstein and Murray (1994). In both cases we find that our qualitative findings are robust. Finally, we relax the assumption that the distribution of abilities is time-invariant. A body of evidence known as the Flynn Effect (Flynn 1984; Flynn 2009) suggests that mean ability has increased steadily for several decades. Rising mean abilities introduce an additional wedge between measured wages and skill prices. This affects the wage growth rates, but not the evolution of skill premiums, implied by our model.

Related Literature. A small number of previous studies have studied whether student abilities have declined over time. Finch (1946) and Taubman and Wales (1972) collect aptitude or achievement test score data from several studies to identify changes in student abilities over time. We extend their work by quantifying the implications of changing abilities for the evolution of wages, taking into account that cognitive tests provide only noisy measures of individual abilities.

Juhn, Kim, and Vella (2005) study whether cohort mean abilities decline as education expands. Since they question the comparability of available aptitude test data, Juhn, Kim, and Vella (2005) propose an approach that avoids measuring abilities entirely. They regress cohort wages on measures of education using Census data and find a weak effect. This approach faces a number of challenges. Given that cohort education rises smoothly over time, it is difficult to disentangle the effects of experience, cohort quality and time varying skill prices. The identifying variation in their approach comes from the relative wage movements of young (educated) and old (less educated) cohorts. An alternative interpretation for such wage movements has been proposed by Card and Lemieux (2001). They show that the rising skill premium during the 1980s affected young and old workers differently and interpret this as evidence in favor of imperfect substitutability between young and old workers. We avoid this issue by focusing our analysis on workers of a fixed age.

A growing literature studies models of school choice with heterogeneous abilities. Part of this literature assumes that cognitive test scores measure abilities perfectly (e.g., Heckman, Lochner, and Taber 1998; Garriga and Keightley 2007). The variance of abilities can then be estimated by regressing wages on test scores. This sidesteps the identification problems we face. However, if test scores measure abilities with noise, regression estimates understate the variance of abilities and hence the ability gaps between school groups.

Laitner (2000) studies a model of human capital investment with workers of heterogeneous abilities. His model qualitatively accounts for changes in relative wages and in wage inequality observed in U.S. post-war data. We move beyond a qualitative analysis and quantify the importance of changing abilities for movements in measured wages. It is particularly important to quantify the bias in the skill premium. The model predicts that average ability of high school and college students should have dropped, so it is not clear what impact this will have on the skill premium.

Our work is also related to the large literature that documents the evolution of wages and skill premiums in the U.S. and proposes a range of explanations. We refer the reader to Goldin and Katz (2008) for references. Our analysis complements this literature. It suggests that the changing ability composition of workers masks some movements of relative wages during the post-war period.

4

Organization. The paper is organized as follows. Section 2 introduces our model of school choice. The calibration is described in Section 3. Our main findings are presented in Section 4. We consider extensions, such as improved school sorting by ability, in Section 5. The final Section concludes.

2 A Model of School Choice

Outline. We develop a model of school choice to measure the changing ability composition of workers with different educational attainment since 1950. A broad outline of the model is as follows.

The economy is inhabited by cohorts of finitely lived individuals. At birth, each individual is endowed with an ability a, which determines post-schooling earnings, and a school preference parameter p^*. Based on a and p^*, the worker chooses one of S schooling levels, indexed by s. Choosing more schooling raises earnings, but incurs higher schooling costs.

Ceteris paribus, students with higher abilities choose longer schooling. Since one of our main objectives is to measure the degree of sorting by ability, the model needs a friction that drives a wedge between abilities and school choices. The school preference p^* creates this friction.

After spending T_s years in school, the agent enters the labor market where she works an exogenous number of hours in each period of life, earning an exogenous hourly wage. More able workers earn more. Labor earnings can be consumed or saved. Consumption choice over the life-cycle is standard.

For now we focus on two driving forces for the increase in schooling attainment over time: changes in the relative costs of different education levels, and changes in the relative wages earned by different school levels. We consider other possibilities later. The details of the model are described next.

Demographics. Time is discrete and indexed by t. Each year, a cohort of new workers of unit measure is born. Individuals live for T periods.

Endowments. At birth, each person is endowed with a pair of scalars, $\hat{a}$ and p. $\hat{a}$ determines the worker's productivity in the labor market. p determines her cost of schooling.

In the population $\hat{a}$ and p are correlated. We model the correlation by assuming that both scalars share a common component, a. Specifically, we write $\hat{a}$ as the sum of two orthogonal components: $\hat{a} = a + a^*$ with $a \sim N(0, \sigma_a)$ and $a^* \sim N(0, \sigma_{a^*})$. Similarly, we assume that $p = a + p^*$ with $p^* \sim N(0, \sigma_{p^*})$. a, a^*, and p^* are mutually independent. The correlation between ability and school costs stems from the common component a.

We refer to a as worker "ability" as it affects both school choice and earnings. Since a^* does not affect school choice, we call it "luck" but point out that individuals may know their realizations

of a^* before choosing schooling. p^* affects school choice but not wages. We refer to it as the individual's school preference.

Preferences. At birth, individuals are indexed by their type $q = (a, a^*, p^*, \tau)$, where τ denotes the year of birth. Let $c_{s,q,v}$ denote the consumption of a person of type q with schooling s at age $v = t - \tau + 1$. Then lifetime utility is given by

$$\sum_{v=1}^{T} \beta^v \log(c_{s,q,v}) - ps - \chi_{s,\tau} \tag{1}$$

where $\beta > 0$ is the discount factor, and $\chi_{s,\tau}$ is a utility cost of schooling that is common to all members of cohort τ. ps denotes the individual specific disutility from schooling. School costs measure the relative preferences of workers for time spent in school versus work, the relative preferences of workers for college versus high school occupations, and the relative financial costs of different education levels.

Work. At ages 1 through T_s, students are in school and do not work. After graduation, workers supply e^a units of type s labor in each period. A worker with school type s earns a real wage of $e^{z_{s,t}}$ per efficiency unit of work time. The real wage varies by skill group and year.

2.1 Worker's Problem

Workers choose schooling s and a consumption path $c_{s,q,v}$ to maximize (1) subject a budget constraint which equates the present value of consumption to the value of lifetime earnings:

$$\sum_{v=1}^{T} \frac{c_{s,q,v}}{R^v} = Y(s,q) \tag{2}$$

$$= e^a \sum_{v=T_s+1}^{T} R^{-v} e^{z_{s,\tau+v-1}} \tag{3}$$

R is the exogenous gross interest rate.

2.2 Optimal Consumption and Schooling

We derive expressions that characterize the worker's consumption and schooling decisions. We can solve the worker's problem in two steps: first, we find the optimal allocation of consumption over time given school choice; then we find the school choice that maximizes lifetime utility.

The lifetime consumption profile obeys the standard Euler equation

$$c_{s,q,v+1} = \beta R c_{s,q,v} \tag{4}$$

which implies a present value of lifetime consumption given by $c_{s,q,1}\Lambda$ where $\Lambda = \sum_{v=1}^{T} \beta^{v-1}/R$ is a present value factor. The budget constraint then implies a level of consumption given by

$$c_{s,q,1} = \Lambda^{-1} Y(s,q) \tag{5}$$

Lifetime utility is then given by

$$V(s,q) = \sum_{v=1}^{T} \beta^v \left[\log\left(c_{s,q,1}\right) + (v-1)\log\left(\beta R\right)\right] - ps - \chi_{s,\tau} \tag{6}$$

$$= R\Lambda\beta \log\left(\Lambda^{-1} Y(s,q)\right) - ps - \hat{\chi}_{s,\tau} \tag{7}$$

where

$$\hat{\chi}_{s,\tau} = \chi_{s,\tau} - \sum_{v=1}^{T} \beta^v (v-1)\log(\beta R)) \tag{8}$$

is an aggregate of all the school-specific terms that are constant across workers. Optimal school choice satisfies

$$s = \arg\max V(s,q) \tag{9}$$

The model implies that school choice depends on p but not separately on a, p^* or a^*. To see this, note that the household prefers s over $\hat{s}$ if

$$V(s,q) - V(\hat{s},q) = R\Lambda\beta \log\left(Y(s,q)/Y(\hat{s},q)\right) - p(s - \hat{s}) - (\hat{\chi}_{s,\tau} - \hat{\chi}_{\hat{s},\tau}) > 0 \tag{10}$$

Since the ratio of lifetime earnings, $Y(s,q)/Y(\hat{s},q)$, is common to all individuals, the decision rule for schooling is fully characterized by a set of cutoff values $p_0 < ... < p_S$, such that agents with $p_{s-1} < p \leq p_s$ choose schooling level s. Variation in school costs (χ) or wages changes the cutoff values but not the ordering of individual school choices. This simplification improves the tractability and the transparency of the model, as discussed in Section 2.3.

The indifference condition (10) reveals two reasons why schooling may rise over time: changes in the relative costs of schooling ($\chi_{s,\tau} - \chi_{\hat{s},\tau}$) and changes in relative wages ($z_{s,t} - z_{\hat{s},t}$). Skill neutral changes in wages of school costs do not affect school choice.

Aptitude test scores. Our calibration approach relies on aptitude test scores as measures of ability. In particular, we use the AFQT scores reported for 94% of the NLSY79 sample (details below). In contrast to much of the literature, we assume that AFQT scores are *noisy* measures of a and write $AFQT = a + \epsilon_{AFQT}$. The noise term $\epsilon_{AFQT} \sim N(0, \sigma_{AFQT})$ is orthogonal to all other random variables.

2.3 Discussion of Modeling Choices

A number of our modeling assumptions deserve comment.

Ability. Our notion of ability is a broad one. We can think of a worker as endowed with various traits that affect either school choice or wages or both. We label these traits p^*, a^*, and a, respectively. For our purposes it is not important what traits each variable represents.

The language adopted in the paper reflects one particular interpretation: a contains abilities that raise worker productivity and reduce the cost of schooling. One may think of cognitive skills that enhance work and study efficiency. p^* represents preferences or skills that affect how costly schooling is without changing wages. Preferences and abilities are correlated.

An alternative interpretation follows the ideas of Manski (1989) where students learn about their abilities as they move through school. Abilities are cognitive and non-cognitive skills. Some are helpful in school (a), while others are not (a^*). When an agent enters the economy, she observes a noisy signal of her ability given by $p = a + p^*$ where p^* represents the signal noise. Students with better signals choose to stay in school longer. The alternative interpretation yields the same model equations as the one we adopt. However, it suggests to view AFQT as a noisy measure of the agent's signal p rather than as a noisy measure of a. Assuming that the agents know more about their own abilities than does the econometrician, we would define $AFQT = a + p^* + \epsilon_{AFQT}$.

Other labels could be attached to $\hat{a}$ and p without changing the model or the findings. The key feature of the model is sorting by a, but imperfect sorting due to a friction p^*. The degree of sorting is one of the central objects the model is designed to measure; it is important for the quantitative results (see Section 4). The exact nature of the friction may not be important.[4]

School choice depends only on p. The fact that agents' school choices depend only on p, not separately on a and p^*, greatly simplifies the model solution. Two assumptions generate this property. (i) Log utility implies that the value gap $V(s,q) - V(\hat{s},q)$ is a function of the ratio of lifetime earnings, $Y(s,q)/Y(\hat{s},q)$. (ii) The assumption that a worker of ability a supplies the same efficiency units of labor, regardless of school choice, implies that the ratio of lifetime earnings does not depend on ability.

The computational advantage of perfect sorting by p is substantial. In a more general model, we would have to find regions in (a, p^*) space where $V(s,q) > V(\hat{s},q)$ for all $\hat{s}$. With perfect sorting by p, we only need to solve for the cutoffs $p_1 < ... < p_S$ that satisfy the indiffence condition $V(s,q) = V(s-1,q)$. Another advantage is that perfect sorting renders identification transparent, as we discuss in Section 3.

Note that some model details, such as preferences and life-spans, do not affect the model outcomes we are interested in. We describe them mainly to show how school sorting by p could be the outcome of a fully specified model.

Other assumptions. The linear school cost specification, ps, could be relaxed without changing the findings. As long as agents with high p face a high marginal cost of increasing s, individuals sort by p and the model solution is unchanged.

---

[4]Borrowing constraints are commonly explored as a friction to educational sorting. The literature has not arrived at a consensus about their quantitative importance. Cameron and Taber (2004) find no evidence of borrowing constraints in the U.S. However, their evidence does not apply to the early cohorts contained in our data.

We abstract from variation in labor efficiency and hours worked across ages and school groups. The well-known collinearity of cohort, age and year prevents us from separately identifying the age efficiency profiles. This problem does not affect our main findings. Our goal is to measure by how much measured wage series need to be revised in order to account for time-varying abilities by school group. These revisions are calculated for fixed ages and do not depend on assumptions about age efficiency profiles.

Some authors argue that a 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). In assuming that earnings depend on human capital, but not directly on ability, we abstract from this possibility.

3 Calibration

Model parameters. The parameters to be calibrated determine the dispersion of abilities and preferences $(\sigma_a, \sigma_{a^*}, \sigma_{p^*}, \sigma_{AFQT})$ and the time paths of skill prices $(z_{s,t})$. The remaining parameter values do not affect the moments we use for calibration and therefore need not be determined.

Calibration targets. We summarize the data moments we use as calibration targets, deferring all details to Section 3.2:

1. From the NLSY79, we estimate the joint distribution of schooling and AFQT scores for the 1960 birth cohort. We also measure the wage "return" to AFQT by regressing log wages at age 40 on AFQT scores. The regression coefficients are called β_s.

2. From the PSID, we estimate the dispersion of the permanent component of wages, $Var(w|s)$.

3. From the 1950-2000 waves of the U.S. Census, we estimate the education attained and wages earned around age 40 $(w_{s,t})$ by the cohorts born between 1906 and 1965.

We further impose two constraints on the parameter values:

1. The model must imply correctly ordered skill prices $(z_{s+1,t} > z_{s,t})$ for all Census dates. This mainly bounds ability dispersion from above. If ability is very dispersed, mean ability can vary greatly across school groups. Since $w_{s,t} = E(a|s,t) + z_{s,t}$, this can imply negative skill premiums.

2. Based on the observed reliability of AFQT, we impose $\sigma^2_{AFQT} \geq 0.25\sigma^2_a$.

Two additional data moments have been suggested as calibration targets. Both turn out to be of limited usefulness.

1. If the fraction of college graduates expands over time while the fraction of high school dropouts declines, one might expect the model to predict a large increase in the variance of log wages of college graduates relative to dropouts. However, for the parameters we consider, the model does not imply strong trends in wage variances because the degree of sorting by ability is too weak.

2. When sorting by ability is strong, the distribution of wages for college graduates should be left truncated, while the distribution for high school dropouts should be right truncated. Again, for the parameters we consider, sorting by ability is not strong enough to generate significant deviations from symmetric wage distributions for any school group. We also did not see evidence of asymmetric distributions in the data.

Calibration algorithm. Our calibration algorithm simulates school choices, wages at age 40, and AFQT scores for the cohorts born between 1906 and 1965. The algorithm features two steps.

Step 1 calibrates the dispersion parameters $(\sigma_a, \sigma_{a^*}, \sigma_{p^*}, \sigma_{AFQT})$ to match the NLSY79 targets for the 1960 cohort. The algorithm varies σ_a on a grid. For each value, we find the pair $(\sigma_{p^*}, \sigma_{AFQT})$ that best matches the target moments. We simulate the school choices, wages, and AFQT scores for 1m individuals. We implicitly choose school costs $(\chi_{s,1960})$ to match the fraction of persons in each school group exactly. Since individuals sort perfectly by $p = a + p^*$, the values of $\chi_{s,1960}$ do not have to be computed.

For each school group, we regress log wage $(a + z_{s,t})$ on AFQT. We can do this without knowing $z_{s,t}$ because all regressions contain only persons with the same s. Conforming with the NLSY79 data, AFQT scores are transformed to be standard Normal. Next, we compute the fraction of persons in each school / AFQT quartile cell. We compute a weighed deviation between the model statistics and the corresponding data targets and find the pair $(\sigma_{p^*}, \sigma_{AFQT})$ that minimizes this deviation, subject to the two parameter constraints described above.

For some values of σ_a, all combinations of $(\sigma_{p^*}, \sigma_{AFQT})$ either imply too much permanent wage dispersion or too small wage returns to AFQT. These values are marked as inadmissible. The remaining values of σ_a are consistent with all data moments and constraints. The algorithm thus identifies an admissible range for σ_a rather than a unique value.

At this point, all the dispersion parameters have been determined for each admissible value of σ_a. The algorithm's Step 2 uses Census data to determine the skill prices $z_{s,t}$. For each Census year, we compute the distribution of schooling in the cohort aged 40. Given the dispersion parameters, we compute the mean ability in each school group, $E(a|s,t)$. Skill prices are then given by $z_{s,t} = w_{s,t} - E(a|s,t)$, where $w_{s,t}$ denotes mean log measured wages in Census year t and school group s. Again, we implicitly choose school costs $(\chi_{s,\tau})$ to match the fraction of persons in each school group exactly for every cohort.

	$Var(w|s)$	β_s	Corr. AFQT/school
σ_{a^*}	$+$	0	0
σ_a	$+$	$+$	0
σ_{p^*}/σ_a	$+$	$+$	$-$
σ_{AFQT}/σ_a	0	$-$	$-$

Table 1: Identification

3.1 Identification

The key identification problem is to decompose variation in measured wages ($w_{s,t}$) into the contributions of skill prices ($z_{s,t}$) and worker abilities ($E(a|s,t)$). In our model, the decomposition hinges on the dispersion of abilities (σ_a) and the degree of educational sorting, which is governed by σ_{p^*}/σ_a.

To identify these parameters, we rely on one data moment that relates to the dispersion of abilities, $Var(w|s)$, and one moment that measures educational sorting, the joint distribution of schooling and AFQT. However, the latter uses a noisy measure of a. Hence, we need another data moment that relates to the noise in AFQT: β_s. These considerations motivate our choice of calibration targets.

To understand how identification is achieved, it is useful to think about how each model parameter affects the three targets. This is summarized in table 1.

- Raising σ_{a^*} directly increases wage dispersion. Since agents do not consider a^* in their school choice, it does not affect the other moments.

- Consider raising σ_a while holding constant educational sorting (σ_{p^*}/σ_a) and the signal to noise ratio in AFQT (σ_{AFQT}/σ_a). A higher σ_a directly raises $Var(w|s)$. Since AFQT is normalized to have a unit standard deviation in the wage regression, a higher σ_a also raises β_s.

- Increasing σ_{p^*}/σ_a weakens educational sorting, which affects all target moments. The dispersion of a within school groups increases, which raises $Var(w|s)$ and β_s. Weaker sorting also reduces the correlation of AFQT and schooling.

- Increasing σ_{AFQT}/σ_a makes AFQT a noisier measure of ability and weakens the relationship between schooling and AFQT. Stronger attenutation bias lowers β_s.

Table 1 clarifies how identification works. Raising σ_a increases the return to AFQT, β_s. This can be offset by raising the noise in AFQT. This, in turn, reduces the correlation between schooling and AFQT, which must be offset through stronger sorting (lower σ_{p^*}/σ_a). The value of σ_{a^*} is chosen residually to match the variance of permanent wages.

Since we calibrate four parameters using only three moments, we can only identify a range of parameters. The range is determined by two constraints.

1. If σ_a is very high, a large amount of noise in AFQT is needed to keep β_s low enough. The noise weakens the relationship between schooling and AFQT, which needs to be offset by strong sorting (low σ_{p^*}/σ_a). But with large abiltiy dispersion and precise sorting $E(a|s)$ varies greatly across s, which implies incorrectly ordered skill prices ($z_{s+1,t} < z_{s,t}$). This implies an upper bound on σ_a.

2. If σ_a is very small, wage variation is largely unrelated to ability and AFQT, so that either β_s falls short of the data or AFQT needs to be more precise than data on its reliability suggest. This provides a lower bound for σ_a.

The calibration algorithm thus finds a range of σ_a values that is compatible with the constraints we impose on the parameters. For each σ_a a unique combination of the other parameters (or no combination at all) is consistent with the calibration targets. Fortunately, the acceptable range of σ_a turns out to be so narrow that the implied revisions to wage levels and growth rates also lie in a narrow range.

Note that we use the $Var(w|s)$ target only as an upper bound. If we chose a lower target, we could reduce σ_{a^*} to match it without disturbing any of the other data moments. Moreover, the upper bound never comes close to binding in our model. When σ_a is large, the model implies incorrectly ordered skill prices. In effect, we could drop σ_{a^*} and $Var(w|s)$ from the calibration without changing the results. The reason for keeping both is that we are interested in the fraction of wage variation that is due to schooling related traits (σ_a) versus traits that, in our model, are indistinguishable from luck (σ_{a^*}).

The remainder of this section describes the construction of the calibration targets in detail.

3.2   Data Moments

In this section, we describe how the calibration targets are constructed. The Appendix provides additional detail.

3.2.1   Cohort Education and Wages

We estimate educational attainment and wages by cohort from the 1950 to 2000 waves of the IPUMS database (Ruggles and Sobeck, 2007). We do not include 1940 because it is a war year. The sample includes all white men aged 35-44 who are not in school, who do not live in group quarters, and who report positive wage and salary income. The age range is chosen so that each birth cohort is observed exactly once at an age when schooling has likely been completed and labor market participation is high. We construct measures of educational attainment and of real hourly wages for each cohort born between 1906 and 1965.

Figure 1: Educational attainment by cohort
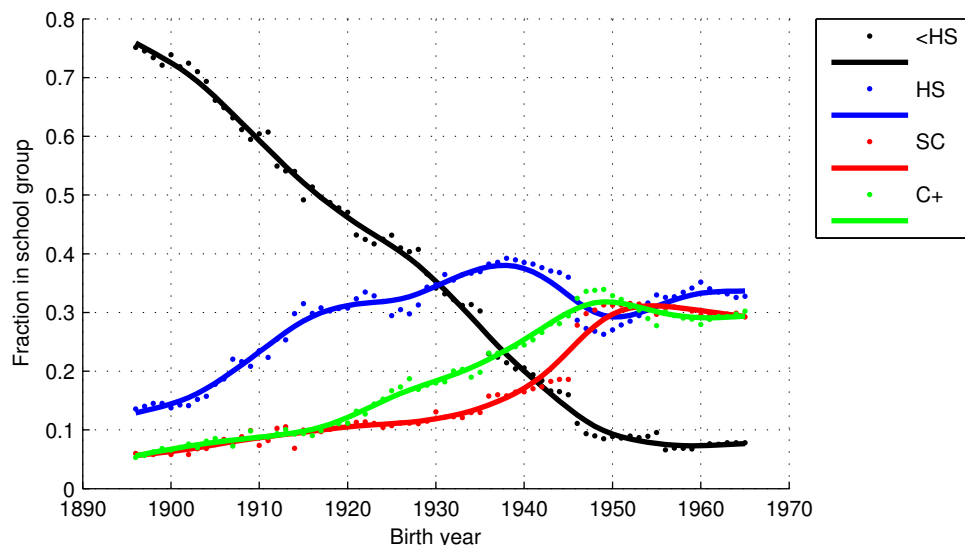
Cohort educational attainment. Figure 1 shows the fraction of persons in each birth cohort that reports a given schooling level. Similar data have been reported, for example, by Goldin and Katz (2008). The solid lines represent Hodrick-Prescott filtered data. To highlight the long-run trends, we include 1940 data in this figure, even though we do not use them in the calibration.

Relative wages. For each Census year, Figure 2 shows the mean log wage of each school group relative to high school graduates. Our data replicate the main features previously documented by Goldin and Katz (2008). Since 1950, we observe a sharp increase in the college wage premium and a decline in the relative wages of high school dropouts.

3.2.2 Education and Aptitudes

We use NLSY79 data to measure the degree of educational sorting by ability and the covariation of wages with ability. The NLSY79 is a representative, ongoing sample of persons born between 1957 and 1964. We retain all men who participated in the ASWAB battery of aptitude tests, which we interpret as a noisy signal of ability. We include members of the minority samples, but use weights to offset the oversampling of minorities. For each person, we construct measures of real hourly wages at age 40 and of educational attainment. The details are given in Appendix B.

Schooling and ability. Our proxy for ability is the 1980 Armed Forces Qualification Test (AFQT) percentile rank (variable R1682). The AFQT aggregates a battery of 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
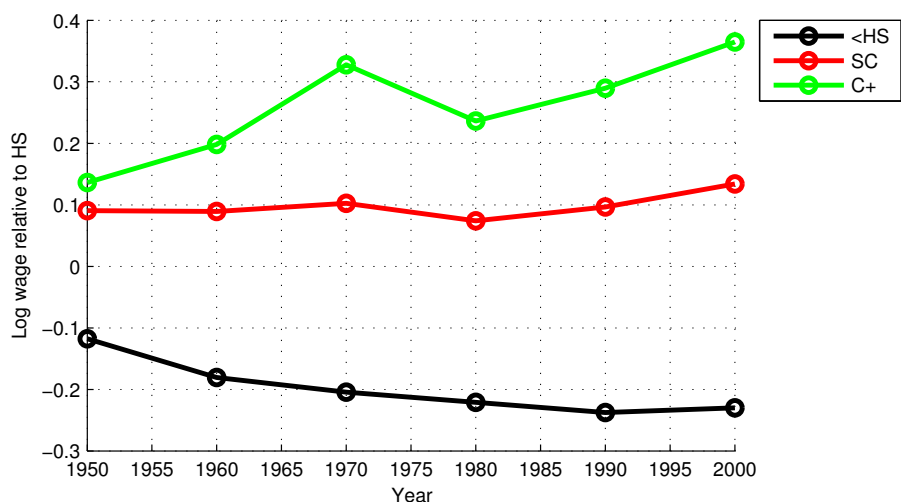
Figure 2: Skill Premia by Census Year

Table 2: Schooling and AFQT: NLSY79 data

AFQT quartile	¡HS	HS	SC	C+
1	0.861	0.417	0.184	0.010
2	0.118	0.338	0.317	0.105
3	0.021	0.194	0.309	0.294
4	0.000	0.051	0.190	0.591
Fraction	0.060	0.323	0.328	0.289
N	163	642	644	493

Note: Fraction of persons falling in each AFQT quintile, conditional on schooling. "Fraction" denotes the fraction of persons completing each school level. N is the number of observations.

by regressing AFQT scores on the age at which the test was administered (in 1980). We transform the residual so that it has a standard Normal distribution, which conforms with our model.

Table 2 characterizes educational sorting by ability. For each school class, the table shows the fraction of persons falling into each ability quintile. The table shows evidence of strong sorting. Half of high school dropouts fall into the lowest AFQT quintile, whereas half of college graduates fall into the highest quintile. This is consistent with Heckman and Vytlacil (2001).

Wages and ability. Table 3 reports the results from regressing log wages at age 40 on AFQT within school classes. AFQT is transformed so that it has a standard Normal distribution in the population. This makes the results comparable with the literature and conforms with our model. The regression coefficients are near $\beta_s = 0.11$, so that a one standard deviation increase in AFQT is associated, on average, with an 11% increase in wages.

Since the values of β_s are important for our finding, we compare our estimates with the literature.

Table 3: Wage regressions: NLSY79 data

	s_1	s_2	s_3	s_4
β	0.251	0.046	0.097	0.150
σ_β	0.082	0.026	0.027	0.038
R^2	0.05	0.02	0.03	0.04
N	163	642	644	493

Note: The table shows the results from regressing log wages at age 40 on AFQT score separately for each schooling group. β is the estimated return to schooling, σ_β is its standard error. N is the number of observations.

Previous results based on NLSY79 data yield estimates that are close to ours. Mulligan (1999, table 6) finds $\beta = 0.11$ in a regression that pools school groups but adds schooling regressors. Altonji and Pierret (2001, table I) find that β increases with experience. At 10 years of experience, a one standard deviation increase in AFQT is associated with a 10.5% increase in wages.

Estimates based on other data sources yield more diverse results. Based on NLS72 and High School and Beyond data, Murnane, Willett, Duhaldeborde, and Tyler (2000) find that a 1% increase in math scores is associated with 10% higher annual earnings. Bowles, Gintis, and Osborne (2001) collect 24 studies with a mean regression coefficient of 0.07. Their figure 6 suggests a wide dispersion of the estimated coefficients. The sensitivity analysis of section 4.2 explores how varying β_s affects our results.

Note that the regression coefficient is not given a structural interpretation in our analysis. We only use it to describe the data. We are interested in how the conditional mean of wages varies with measured ability, not in the "direct" effect of ability on wages, holding other characteristics constant. For this reason, we do not include controls in the wage regression. When we calibrate the model, we simulate AFQT scores and run a regression of exactly this form for the pool of workers who attain each education level.

Precision of AFQT. We wish to derive a lower bound on how precisely AFQT measures ability: σ_{AFQT}/σ_a. Aptitude scores suffer from two types of noise. The first relates to *validity*: does AFQT measure a or does it measure a different individual trait? Given that a is unobservable, we cannot bound validity and assume that AFQT measures a and a only.[5] The second type of noise relates to *reliability*: how precisely does AFQT measure whatever traits it measures? We can bound AFQT's reliability using its correlation with other measures of IQ.

Herrnstein and Murray (1994, Appendix 3) document the correlation between AFQT and six other IQ tests taken by some NLSY79 individuals. The correlations range from 0.71 to 0.9, with a median of 0.81. 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.

---

[5]Herrnstein and Murray (1994, chapter 3) summarize existing evidence supporting the notion that AFQT scores are correlated with job performance.

Table 4: Wage regressions: PSID data

Schooling	$\sigma_{Y(s)}$	σ_α	ρ	$\sigma_{\hat{\varepsilon}}$	σ_ε
<HS	0.389	0.335	0.887	0.171	0.296
HS	0.390	0.270	0.973	0.110	0.327
SC	0.359	0.285	0.881	0.192	0.268
C+	0.444	0.242	0.969	0.154	0.335

Note: The table shows the estimated coefficients obtained from wage regressions using PSID data. $\sigma_{Y(s)}$ is the standard deviation of lifetime earnings. $\sigma_{\hat{\varepsilon}}$ is the standard deviation of the shock in the process governing ζ.

To translate these correlations into a lower bound for σ_{AFQT}/σ_a, consider the following simple model of testing. Individuals take two IQ tests. Each measures a plus a measurement error term with standard deviation σ_{AFQT}. The correlation of the tests then equals $\left(1 + \sigma_{AFQT}^2/\sigma_a^2\right)^{-1}$, which we take to equal 0.8. The implied lower bound for the noise in AFQT is $\sigma_{AFQT}^2/\sigma_a^2 \geq 0.25$.[6]

This bound appears conservative. If AFQT suffers from additional measurement error due to imperfect validity, the bound would be tighter. Moreover, Jensen (1980) reports lower correlations between other IQ tests, ranging from 0.64 to 0.77, which would imply larger measurement errors.

3.2.3 Permanent wage dispersion

The last data moment used in the calibration characterizes the dispersion of the permanent component of wages, $Var(w|s)$. Since our model abstracts from luck and other transitory shocks to earnings, it is important that we purge transitory variation from the wage data. Following Guvenen (2007) and others, we do so by estimating the variance of permanent component of wages.

We think of log-wages of individual j at time t as being generated by an autoregressive earnings process with an individual-specific fixed component:

$$w_{j,t} = \alpha_j + X_{j,t}\beta + \zeta_{j,t} + \varepsilon_{j,t} \tag{11}$$

where X is the vector of the individual's characteristics, β is a vector of constants, ε is a transitory shock, and ζ is a persistent shock which evolves according to an AR(1) process. We estimate this income process using PSID data. Results are given in Table 4. Details are available in Appendix C.

In our model, the variance of log wages, conditional on schooling, equals $\sigma_{a*}^2 + Var(a|s)$. The corresponding data moment is $Var(\alpha_j)$. Note that our model does not constrain the share of the variance of permanent earnings that is related to school choice.

---

[6]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 a variance of measurement error equal to $0.53\sigma_a^2$.

Figure 3: Deviations from Calibration Targets

(a) Returns to AFQT   (b) Lowest Skill Price Difference
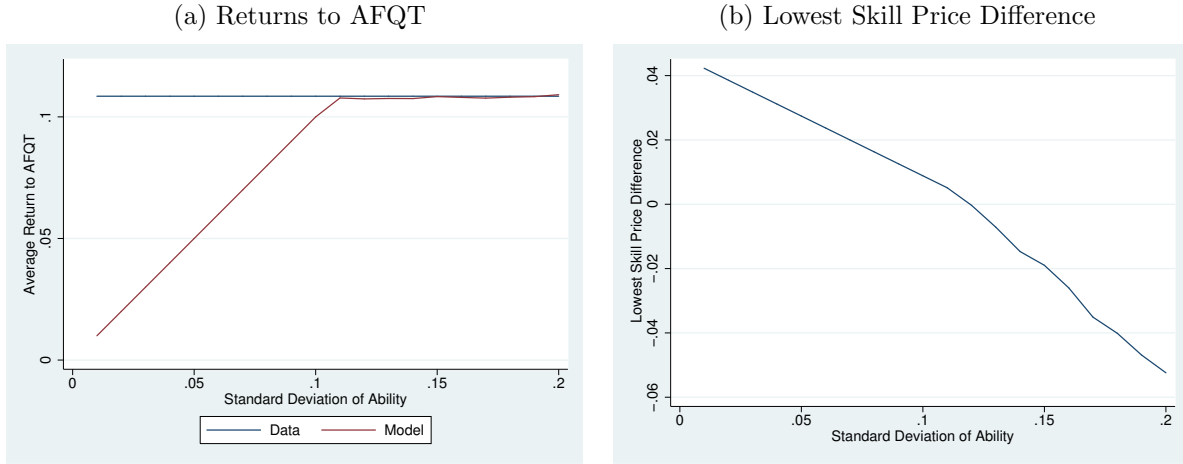
3.3   Model Parameters

Our calibration approach generates a range of acceptable parameter values. To see how that range is determined, Figure 3 show to what extent model can replicate the calibration targets for a grid of σ_a values.

Panel (a) shows the model implied wage returns to AFQT (the mean of β_s across school groups, weighted by the precision of the empirical estimates of β_s). For $\sigma_a < 0.11$ model cannot generate returns to AFQT that match the calibration target of 0.11. This is due to standard attenuation bias. When $\sigma_{AFQT} = 0$, AFQT measures ability perfectly and $\beta_s = \sigma_a$. Adding noise to AFQT reduces the regression coefficient below σ_a. The model therefore implies a lower bound of $\sigma_a \geq \beta_s$.

Panel (b) shows the smallest skill price premium, $z_{s,t} - z_{s-1,t}$, implied by the model across all years and school groups. Where this line drops below zero, the model implies incorrectly ordered skill prices. In this case, the direct effect of further schooling is to lower wages. This happens for all values of σ_a above 0.12, starting with the premium for college graduates in 1950. As σ_a increases above 0.12, the model quickly implies negative skill premiums for a larger range of years and school groups.

The model therefore pins down a very narrow range of acceptable parameter values. We focus on the case $\sigma_a = 0.12$. We show that for this value of σ_a we can calibrate the other parameters so that the model replicates the data targets we have discussed above. We then show the results for the bias in wages and wage premia. Other values of σ_a near 0.12 give similar answers. We also consider below the possibility that either of our targets may be relaxed.

Table 5 shows the calibrated model parameters for $\sigma_a = 0.12$. In this case the dispersion of preferences is larger than the dispersion of abilities, so that preferences drive much of school choices. Further, the dispersion of "luck" is large relative to the dispersion of ability, implying that "luck" drives much of wage dispersion. AFQT scores are relatively accurate measures of ability. Despite the fact that ability explains a minority of school choices or wage variation, we

17

Table 5: Model Parameters

Parameter	Role	Value
σ_a	Cognitive Ability Dispersion	0.12
σ_{p*}/σ_a	Relative Dispersion of Preferences	1.47
σ_{a*}/σ_a	Relative Dispersion of Other Ability	2.33
σ_{afqt}/σ_a	Accuracy of IQ Tests	0.37

show below that this calibration implies economically significant biases to wages.

3.4 Model Fit

In this section, we evaluate the model's ability to replicate the calibration targets. Since the model exactly replicates educational attainment by cohort, this is not shown.

Education and aptitudes. The first set of calibration targets characterizes the joint distribution of AFQT, schooling, and wages discussed in Section 3.2. Figure 4 shows the density of AFQT scores by schooling level for the model and the NLSY79 data. Overall, the model accounts reasonably well for the data. The main discrepancy is the too large fraction of low AFQT persons among high school graduates.

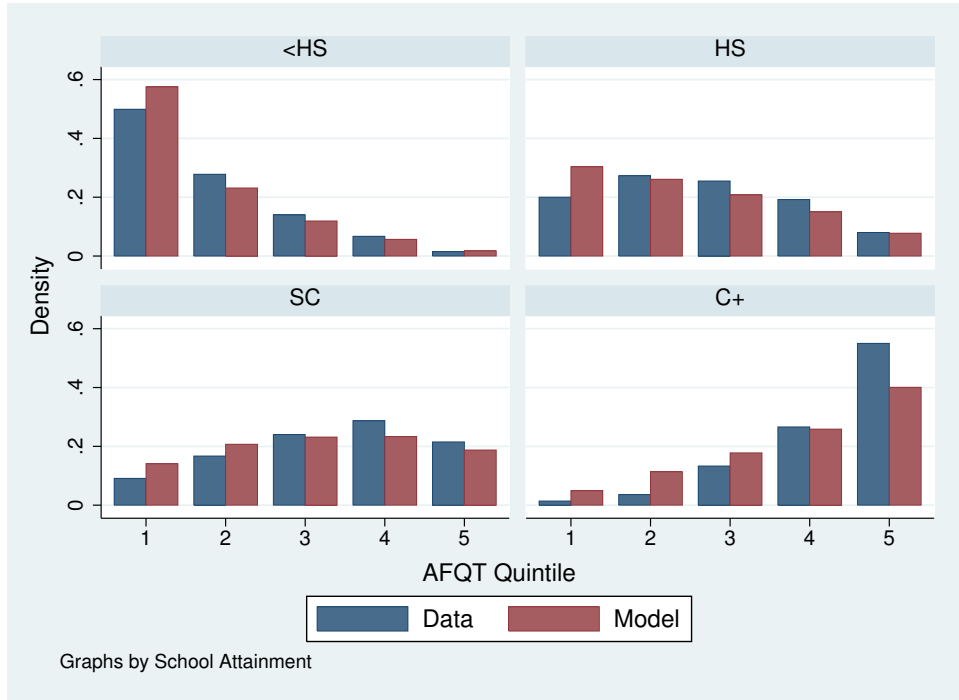Figure 4: AFQT Distribution for Different School Attainments

Figure 5 displays the coefficients β_s obtained by regressing log wages at age 40 on AFQT, which is scaled to have a standard Normal distribution. Separate regressions are estimated for each school group. The model fails to capture the pronounced U-shape in the empirical coefficients. Note, however, that the standard error band around β_1 is quite large.

Figure 5: Returns to AFQT
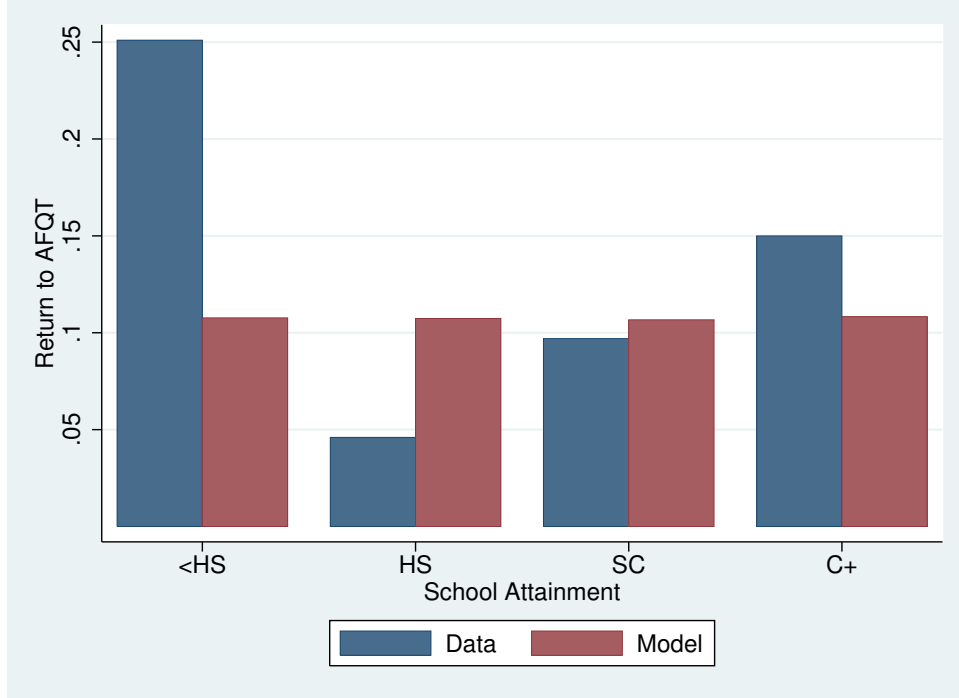
Wages. Figure 6 shows the variance of permanent wages for each school group. The model matches the mean variance by construction. The data show a slight tendency for the variance to decline with schooling, while it is roughly the same for all groups in the model. One reason is that the majority of wage variance is due to "luck" which is assumed not to vary by school group.

4 Results

4.1 Main Results

Our main objective is to adjust measured wages to account for the changing ability composition of workers in each school group. Figure 7 presents the results. Each panel compares the measured wage series for one school group, $w_{s,t}$, with the skill prices implied by the model, $z_{s,t}$. The model wages of high school graduates in the year 1970 are normalized to match the data. The model implies substantial revisions to both the growth rates and the levels of relative wages.

19
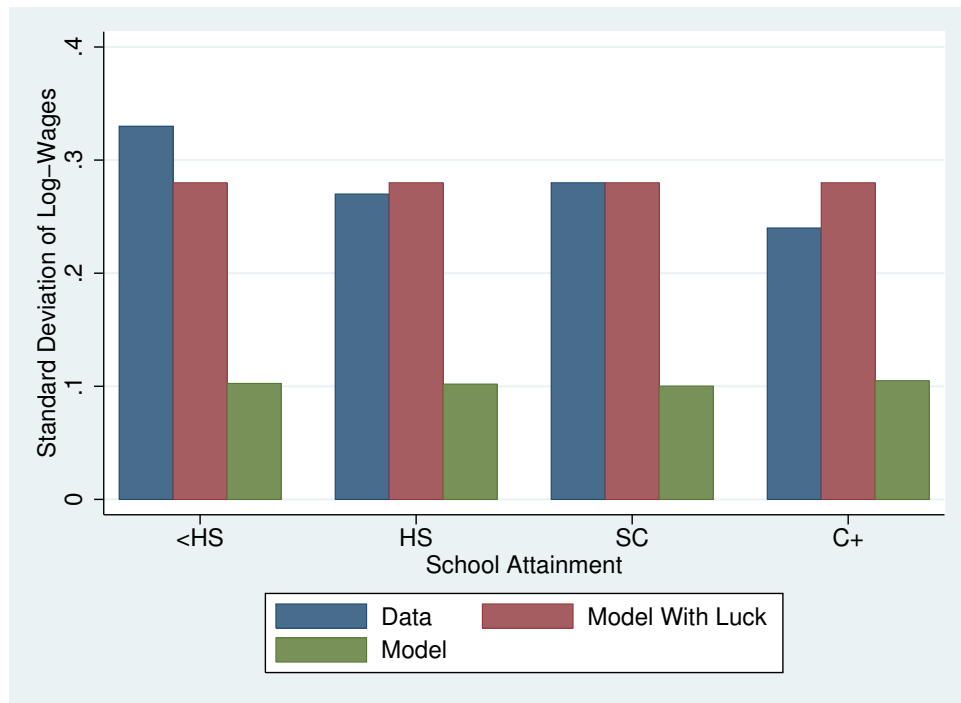
Figure 6: Standard Deviation of Permanent Wages

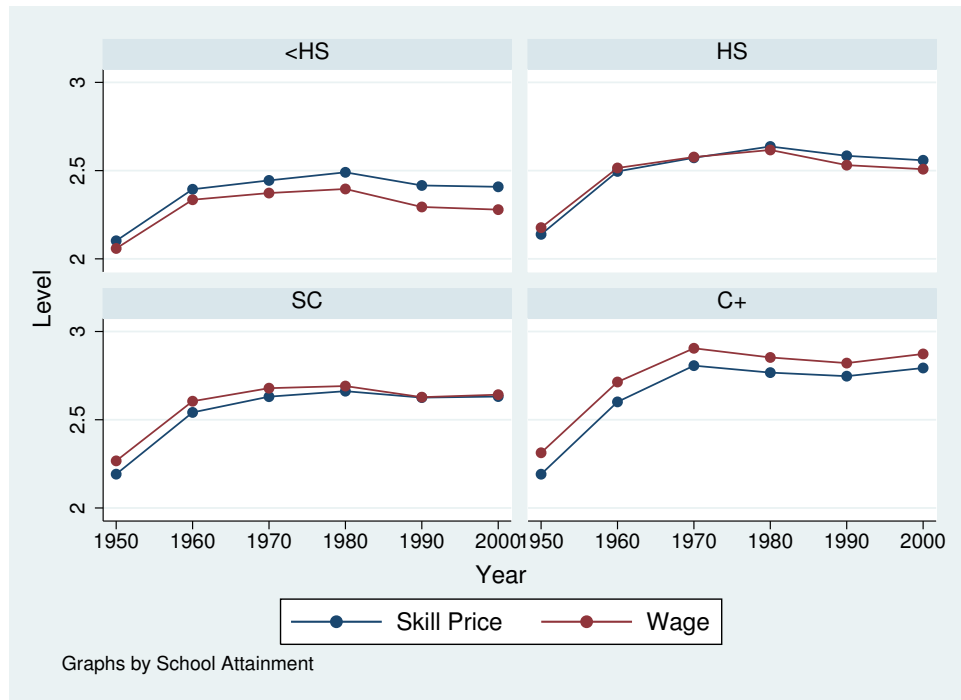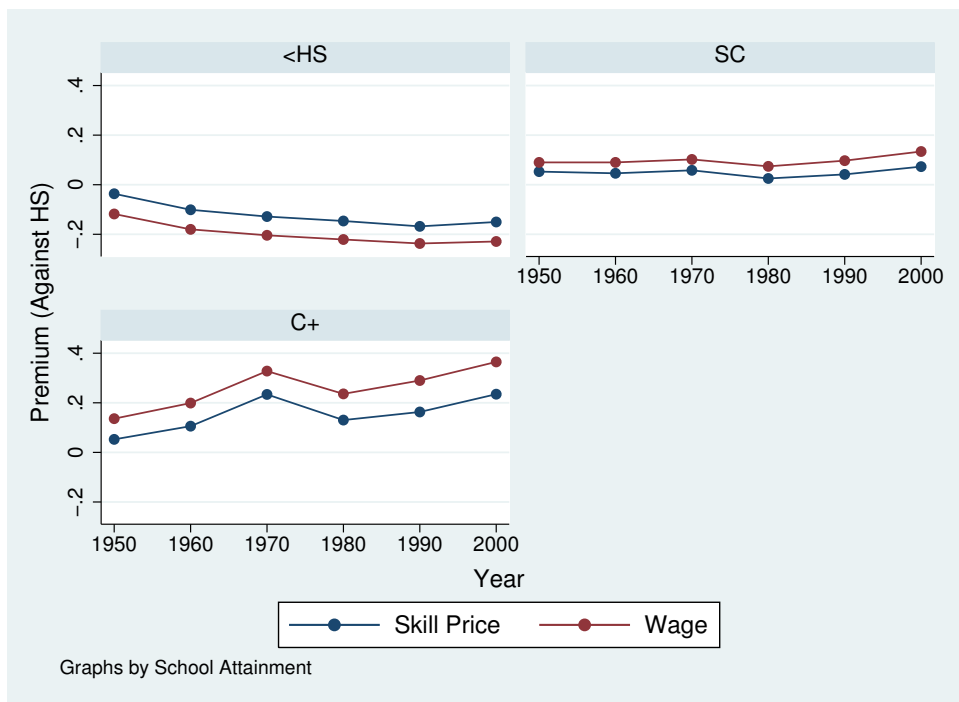Figure 7: Mean log wage by schooling and date

Figure 8: Skill premiums

Wage levels. Figure 8 displays how adjusting for changing abilities alters relative wage levels. Each panel represents one school group. It compares the measured skill premiums relative to high school graduates, $w_{s,t} - w_{2,t}$, with the relative skill prices implied by the model, $z_{s,t} - z_{2,t}$. Recall that the model implies $w_{s,t} = E(a|s,t) + z_{s,t}$, so that the gap between the measured wage series and the series depicting skill prices equals the ability gap of school group s relative to high school graduates.

Figure 8 displays our main finding: the model implies large downward revisions of all skill premiums.

Panel B of table 6 shows the numbers underlying Figure 8. Consider first the year 2000. The measured college wage premium is $w_{4,2000} - w_{2,2000} = 0.37$. Our model implies a mean ability gap between college graduates and high school graduates of 0.14 or about 1.2 standard deviations. This implies a skill price gap of $z_{4,2000} - z_{2,2000} = 0.23$. The measured college wage premium overstates the skill price gap by 55%.

Similar results are found for college dropouts and high school dropouts. In each case, measured wage premiums overstate skill price differences by more than 50%.

Skill premiums are considerably smaller in the 1950 data. Even though the model implies smaller absolute differences in mean abilities across school groups, measured skill premiums overstate skill price gaps by at least 70%. Notably, the college wage premium is reduced from 0.14 to only 0.05. Heckman, Lochner, and Todd (2006) point out that internal rates of return to schooling appear implausibly high. Our findings point towards a possible resolution of this puzzle.

21

Table 6: Results

A: Wage Level Results

	School Attainment			
	<HS	HS	SC	C+
Wage Growth	22%	33%	38%	56 %
Skill Price Growth	31%	41%	42%	60%
Bias, Percent	-28%	-21%	-15%	-7%

B: Skill Premium Results

	School Attainment Comparison		
	<HS-HS	SC-HS	C+-HS
Skill Premium, 1950			
Data	-12%	9%	14%
Skill Prices	-4%	5%	5%
Bias, Percent	224%	70%	159%
Skill Premium, 2000			
Data	-23%	13%	37%
Skill Prices	-15%	7%	23%
Bias, Percent	52%	84%	55%
Growth in Skill Premium, 1950-2000			
Data	-11%	4%	23%
Skill Prices	-11%	2%	18%
Bias, Percent	-3%	120%	26%

Wage growth. We turn next to the implications for wage growth rates. Panel A of Table 6 shows the changes in mean log wages by school group over the period 1950-2000. It compares measured wages with the model's skill prices.

We find upward revisions of wage changes between 0.04 and 0.09 for all school groups. The revisions are larger for the less educated groups. The intuition is that the fraction of high school dropouts has declined dramatically from about 60% for the 1910 cohort to less than 10% for the 1960 cohort. This reduction implies a decline in the mean ability of high school dropouts, which masks 29% of the true wage growth enjoyed by this group.

The share of college graduates expanded much less over the same period (from about one-quarter to one-third of the cohort). As a result, the mean ability decline among college graduates is smaller and masks only about 7% of true wage growth. Our findings account for only a small part of the slow wage growth experienced since 1973 (see Levy and Murnane 1992).

Since all school groups experienced declining mean abilities of similar magnitudes, our model implies modest revisions to the growth of measured skill premiums. The largest revision occurs for the college wage premium, which rose by 0.23 in measured wages compared with an increase in relative skill prices of 0.18.

Our findings may be summarized as follows:

1. Our model implies that between one-half and two-thirds of the measured wage gaps between school groups represent skill price gaps, while the remainder represents ability gaps.

2. About 25% of the measured growth in the college wage premium represents ability bias. For the other school groups, the changes in measured skill premiums and relative skill prices are similar.

3. While mean ability declines within each school group, the implied revisions to wage growth are modest.

4.2 Robustness

In this section we examine how robust our results are to changes in the calibration targets. Three main sets of data moments form the basis for our calibration: the wage returns to AFQT, the joint distribution of schooling and AFQT scores, and the bounds we impose on σ_{AFQT} and the variance of permanent wages.

Wage returns to AFQT, β_s. As pointed out in Section 3.2.2, empirical estimates of the relationship between aptitude scores and wages vary. We are mainly concerned about lower values of β_s for two reasons. First, among the studies summarized by Bowles, Gintis, and Osborne (2001) the mean value of β_s is only 0.07 compared with our estimate of 0.11. Second, lower values of β_s imply lower values of σ_a and therefore smaller ability corrections to the wage series.

Table 7: Robustness: $\beta_s = 0.08$

A: Wage Level Results

	School Attainment			
	<HS	HS	SC	C+
Wage Growth	22%	33%	38%	56 %
Skill Price Growth	29%	41%	43%	60%
Bias, Percent	-25%	-19%	-13%	-6%

B: Skill Premium Results

	School Attainment Comparison		
	<HS-HS	SC-HS	C+-HS
Skill Premium, 1950			
Data	-12%	9%	14%
Skill Prices	-5%	6%	6%
Bias, Percent	152%	56%	116%
Skill Premium, 2000			
Data	-23%	13%	37%
Skill Prices	-16%	8%	25%
Bias, Percent	43%	66%	45%
Growth in Skill Premium, 1950-2000			
Data	-11%	4%	23%
Skill Prices	-11%	2%	19%
Bias, Percent	-2%	92%	22%

Table 7 summarizes how our findings change when we set $\beta_s = 0.08$. The range of σ_a values that are consistent with the data now expands to [0.08, 0.12]. The lower bound allows the model to replicate the target value for β_s with minimal σ_{AFQT}. The upper bound is the same as in the baseline calibration. Since the range of acceptable σ_a values is narrow, we only show results for $\sigma_a = 0.1$ and note that other acceptable values yield similar findings.

Table 7 summarizes the revisions to measured wages and skill premiums. Comparing it with Table 6 reveals that all findings are similar to the baseline case. As expected, the lower value of σ_a implies slightly smaller revisions. However, the main conclusions remain unchanged.

Parameter bounds. For the baseline results the additional restrictions imposed on the parameters never bind. The lower bound on σ_a is obtained from the requirement that the model replicate the estimated wage return to AFQT, β_s. The lower bound on σ_{AFQT} does not bind. The upper bound on σ_a is obtained from the requirement that the skill prices are correctly ordered $(z_{s,t} > z_{s-1,t})$. The variance of permanent wages never exceeds our estimated value.

5   Extensions

5.1   Improved Educational Sorting

A number of studies suggest that educational sorting by ability increased during the 1950s. Taubman and Wales (1972) compile data from several previous studies that measure the cognitive abilities of students at various education levels between 1925 and 1963. Their data suggest that the fraction of high school graduates who attended some college increased disproportionately among the most able students. Most of the increase occurred in the 1950s (see figure 2 in Taubman and Wales 1972). Herrnstein and Murray (1994, chapter 1) show that sorting remained roughly unchanged between 1960 and 1980.

While the comparability of the test scores used and of the student populations covered is an issue, there are reasons to think that educational sorting may have increased.[7] Among the contributing factors may be the declining cost of long distance travel, the relaxation of parental borrowing constraints, and the spreading of standardized testing (Herrnstein and Murray, 1994, chapter 1). We explore the implications of increased sorting by ability in our model.

Calibration strategy.   All dispersion parameters are calibrated as in the baseline case described in section 3. However, we now treat σ_{p^*}, which governs the strength of educational sorting, as time-varying.

We calibrate the values of σ_{p^*} for the cohorts born between 1910 and 1960 as follows. For each cohort that is covered by Table 1 in Taubman and Wales (1972), we find the value of σ_{p^*} that best matches the mean AFQT percentiles for high school graduates who do and do not attempt college.[8] For the 1960 cohort we keep the value of σ_{p^*} that was calibrated from the NLSY79 data. Between these cohorts we interpolate σ_{p^*} linearly.

The range of admissible values for σ_a increases to [0.11, 0.14]. In the baseline calibration, $\sigma_a = 0.14$ implied negative skill price premiums during the early years. This is no longer the case because sorting by ability is now weaker during those years. For a given σ_a this reduces the ability bias corrections. The lower bound of 0.11 coincides with the baseline calibration. This is expected as the data moment that constrains σ_a from below is independent of the Census wage data. Given the narrow range of admissible parameters, we show results for $\sigma_a = 0.12$.

Results.   Ability sorting strengthens substantially between 1950 and 2000. The relative standard deviation of the school cost shock, σ_{p^*}/σ_a, declines from 2.5 to 1.5 over this period.

---

[7]Juhn, Kim, and Vella (2005) question the comparability of the test scores collected by Taubman and Wales (1972) on the grounds that they pool data based on different aptitude tests and covering different samples. Bishop (1989) addresses the comparability problem by using the Iowa Test of Educational Development, which has been administered to 95% of Iowa schools since 1940. Unfortunately, Bishop's data contain no information about the relative scores of different education groups.

[8]Their Table 2 reports in addition the fraction of high school graduates who attempt college for four AFQT percentiles. We do not use this data because it is interpolated from a regression model.
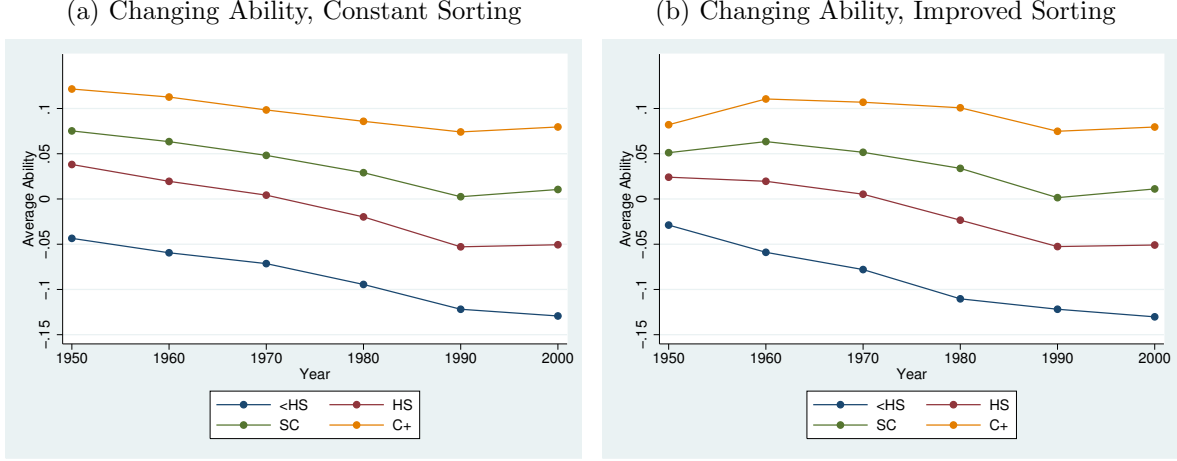
Figure 9: Changing Sorting by Ability

(a) Changing Ability, Constant Sorting (b) Changing Ability, Improved Sorting

Figure 9 shows how improved sorting changes the revisions to measured wages. It compares $E(a|s,t)$ for all years and school groups with the corresponding values for the baseline model where sorting is time invariant. For the year 2000, the two figures coincide. However, weaker sorting in 1950 compresses the ability gaps between school groups. As a result, the model with improved sorting implies no decline in the mean abilities of college graduates, while the ability decline is exacerbated for high school dropouts.

Changes in skill premiums. Table 8 summarizes the revisions to the measured wage series implied by the model with improved sorting. The revisions to the year 2000 skill premia are the same as those shown in Table 6. The revisions to the 1950 skill premiums are smaller than in the baseline case. For a given value of σ_a the model implies less ability bias in the earlier years because ability sorting is weaker.

Changes in wage growth. With time invariant sorting, the model implies upward revisions to wage growth for all school groups while education rises over time. When sorting improves over time, mean abilities decline by less for the highly skilled groups, while they decline by more for the least skilled groups. As a result, the revisions to wage growth are reduced for college graduates, while they are increased for high school dropouts (see Figure 9). For high school graduates and some college, the revisions change in ways that cannot be signed a priori.

This reasoning explains why the model now implies larger revisions to the growth rate of high school dropout wages, while there is almost no revision to the growth of college wages. Larger revisions to the growth rates of low skilled wages also imply that the model attributes a larger share of the observed increase in the college wage premium to ability bias (from 23% to 16%).

Overall, the qualitative findings of the baseline case remain valid. The model implies large reductions in the college wage premium and substantially faster growth of unskilled wages than the raw

26

Table 8: Results: Improved Ability Sorting

A: Wage Level Results

	School Attainment			
	<HS	HS	SC	C+
Wage Growth	22%	33%	38%	56 %
Skill Price Growth	32%	41%	42%	56%
Bias, Percent	-32%	-18%	-10%	0%

B: Skill Premium Results

	School Attainment Comparison		
	<HS-HS	SC-HS	C+-HS
Skill Premium, 1950			
Data	-12%	9%	14%
Skill Prices	-7%	6%	8%
Bias, Percent	81%	43%	74%
Skill Premium, 2000			
Data	-23%	13%	37%
Skill Prices	-15%	7%	23%
Bias, Percent	52%	84%	55%
Growth in Skill Premium, 1950-2000			
Data	-11%	4%	23%
Skill Prices	-8%	1%	16%
Bias, Percent	31%	384%	46%

wage data suggest. It also attributes a substantial fraction of the measured increase in the college wage premium to ability bias.

5.2   The Flynn Effect

Our results so far have assumed that the distribution of abilities is time-invariant. There is, however, substantial evidence that average scores on tests of cognitive skills have drifted up at a rate of about 1.5 standard deviations every 50 years. This observation is called the Flynn Effect (see Flynn 1984; Flynn 2009). There is some disagreement in the psychometric literature as to whether the Flynn Effect represents gains in cognitive skills or improvements in test taking skills (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 almost exactly to our time period. He also conjectures (based on somewhat impartial 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.

It is easy to see how rising mean abilities change the skill prices implied by the model. The indifference condition (10) implies that schooling is invariant against changes in mean cohort abilities. If mean abilities drift up by 1.5 standard deviations every fifty years, all of the skill price changes implied by the model over the period 1950 to 2000 are increased by $1.5\sigma_a$, or about 16% given our baseline parameters. The Flynn Effect does not alter the skill premiums implied by the model or their growth rates.

6   Conclusion

Measured wages confound skill prices and worker abilities. We develop a calibrated model of school choice in order to measure U.S. skill prices for four school groups over the period 1950-2000. We find that measured wages substantially overstate both the level and the growth rate of the college skill premium. Further, measured wage growth understates the growth of skill prices for workers with low schooling.

In our model, heterogeneity in school preferences accounts for most of the variation in school choice across individuals. This is a common feature of models of school choice. Since "explanations based on psychic costs are inherently unsatisfactory" (Heckman, Lochner, and Todd, 2006), future research should develop models of school choice with more explicit frictions that prevent perfect sorting by ability.

References

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.

BOWLES, S., H. GINTIS, AND M. OSBORNE (2001): "The Determinants of Earnings: A Behavioral Approach," *Journal of Economic Literature*, 39(4), pp. 1137–1176.

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., 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.

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.

GUVENEN, F. (2007): "Learning Your Earning: Are Labor Income Shocks Really Very Persistent?," *The American Economic Review*, 97(3), pp. 687–712.

HECKMAN, J., AND E. VYTLACIL (2001): "Identifying the Role of Cognitive Ability in Explaining the Level of and Change in the Return of Schooling," *The Review of Economics and Statistics*, 83(1), pp. 1–12.

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.

JENSEN, A. R. (1980): *Bias in Mental Testing*. 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, Y. DUHALDEBORDE, AND J. H. TYLER (2000): "How ImportantAre the Cognitive Skills of Teenagers in Predicting Subsequent Earnings?," *Journal of Policy Analysis and Management*, 19(4), 547–568.

Table 9: Summary statistics: Census data

	1950	1960	1970	1980	1990	2000
N	17619	75019	73216	380173	542070	564203
Fraction ¡HS	0.60	0.45	0.35	0.20	0.09	0.07
Fraction HS	0.22	0.32	0.35	0.39	0.27	0.34
Fraction SC	0.09	0.11	0.12	0.17	0.31	0.30
Fraction C+	0.09	0.12	0.18	0.25	0.33	0.29
$w_{1,40}$	7.8	10.3	10.7	11.0	9.9	9.8
$w_{2,40}$	8.8	12.4	13.2	13.7	12.6	12.3
$w_{3,40}$	9.7	13.5	14.6	14.7	13.8	14.0
$w_{4,40}$	10.1	15.1	18.3	17.3	16.8	17.7
College wage premium	0.14	0.20	0.33	0.24	0.29	0.36

Notes: The table shows summary statistics for the Census sample. N is the number of observations. The next four rows show the fraction of observations in each school group. $w_{s,40}$ denotes the average real wage of persons aged 40 in each school group. The college wage premium equals $w_{4,40} - w_{2,40}$.

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.

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.

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.

Appendix

A   Census Data

Samples. 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.

We restrict the sample to white men between the ages of 35 and 44, so that each cohort born between 1906 and 1965 is observed exactly once. The age range is chosen so that schooling is completed and most men participate full time in the labor market. We drop individuals who are in school or not employed, who live in group quarters, or who report no wage income. Table 9 shows descriptive statistics for each Census year.

Educational attainment. Our measure of educational attainment is the IPUMS variable EDU-CREC. It distinguishes nine levels of education, which we aggregate into four groups: less than high school, high school, some college, and at least college completed.

Before proceeding, it is useful to discuss a technical detail in the construction of the educational attainment data. Figure 1 shows discrete jumps between adjacent cohorts that are observed in different Census years. One reason is that the wording of the educational attainment question changed between 1980 and 1990. Prior to 1990, HIGRADE recorded years of schooling completed. Since 1990, EDUC99 asks for the highest degree attained. This affects in particular whether people report high school or some college.

We do not see a compelling way of correcting this problem. **?** use Current Population Survey data to estimate the changes in education between 1980 and 1990. Two problems prevent us from adopting their approach: (i) The magnitude of the mismeasurement likely changes from one Census year to the next. The reason is that differences in the educational attainment questions affect only a subset of the population. The size of this population changes with the distribution of educational attainment. (ii) We observe jumps in educational attainment also between 1970 and 1980, even though both years use the HIGRADE version of the attainment question.

The outstanding feature of the data is the large decline in the fraction of high school dropouts. The changes in the attainment questions affect mainly those who are the border between two degrees (e.g., high school vs. some college). Since most of those identified as dropouts in 1940 report less than 11 years of schooling, we are confident that they did not achieve a high school degree. We therefore believe that the decline in high school dropouts is real and not an artifact of the 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).

Our model abstracts from variation in demographic characteristics that are related to wages in the data, such as martial status or region of work. We therefore purge the data from the resulting variation in wages by regressing log wages on years of schooling, marital status, region of residence, and urban status. Using wages without adjusting for these characteristics changes the series of measured wages, but not the revisions implied by our model.

Aggregation. For consistency reasons we calculate all cohort and year aggregates from a matrix of summary statistics that is indexed by school group, birth year, and year (s, τ, t). For each cell, the matrix records mean log wages, aggregate earnings and hours, etc.

The educational attainment of a birth cohort is measured at the unique age, between 35 and 44, at which this cohort is observed. Since cohorts are observed at different ages, the educational attainment estimates are not fully comparable. However, data for pseudo-cohorts suggest that educational attainment does not change substantially between the ages of 35 and 44.

The log wage of school group s at date t is measured by the mean log wage of all persons who are exactly 40 years old in year t. In order to reduce measurement error, we impute this wage by regressing mean log wages on a quadratic in age.

B NLSY79 Data

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 and who did not participate in the ASWAB 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, race, marital status, and region of residence. Separate regressions are estimated for each year and schooling group.

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.

Summary statistics. Table 10 summarizes the data. For each school class, the table shows average years of schooling, the average AFQT percentile rank, the mean log wage at age 40, and the number of persons in the sample.

C PSID Data

This section describes how we estimate the variance of permanent wages. We use the 1968 to 2003 waves of the Panel Study of Income Dynamics (PSID). The sample contains all men who report

Table 10: Summary statistics: NLSY79 data

School class	¡HS	HS	SC	C+
Avg.school	9.5	11.7	13.3	17.0
Real wage at age 40	13.9	17.2	22.0	39.0
Adj. wage at age 40	12.7	14.2	17.5	32.6
AFQT percentile	0.13	0.34	0.50	0.75
N	163	642	644	493

Notes: The table shows summary statistics for each school group. Avg. school denotes average years of schooling. The last two rows show the average percentile in the AFQT distribution and the number of observations (unweighted).

at least 15 valid wage observations between the ages of 18 and 65. Wage observations are valid if hours worked fall in the interval $[520, 5110]$ and labor income is positive. Wages below 0.02 times the median or above 100 times the median are deleted. Labor income includes wage and salary income as well as the labor income share of self-employment income. The real wage is defined as total labor income divided by total hours worked, deflated by the Consumer Price Index.

Estimating the stochastic process governing wages. Our estimation strategy follows Guvenen (2007). The first step is to form a residual wage. We pool all observations within a given school group and regress the log real wage on a quartic in experience. We assume that the residual wage is governed by a process of the form

$$w_{j,t} = \alpha_j + X_{j,t}\beta + \zeta_{j,t} + \varepsilon_{j,t} \tag{12}$$
$$\zeta_{i,t} = \rho\zeta_{i,t-1} + \hat{\varepsilon}_{i,t} \tag{13}$$

where the error terms $\varepsilon_{i,t} \sim N(0, \sigma_\varepsilon)$ and $\hat{\varepsilon}_{i,t} \sim N(0, \sigma_\zeta)$ are independently distributed. $w_{j,t}$ is the log residual wage of person j at date t. It is composed of a fixed effect α_j, a persistent shock $\zeta_{j,t}$, and a transitory shock $\varepsilon_{j,t}$.

We estimate the parameters of the wage process by minimizing the sum of squared deviations between the empirical covariance matrix of wages and the one implied by the model (12). All deviations are equally weighted. Only elements of the empirical covariance matrix with at least 200 contributing individuals are retained.