CHANGES IN THE CHARACTERISTICS OF AMERICAN YOUTH: IMPLICATIONS FOR ADULT OUTCOMES

JOSEPH G. ALTONJI, PRASHANT BHARADWAJ & FABIAN LANGE†

† ALTONJI: YALE UNIVERSITY, NBER, & IZA; BHARADWAJ: YALE UNIVERSITY; LANGE: YALE UNIVERSITY & IZA

Date: **September 2007, Preliminary Draft**.

1. Introduction

How will the adult labor market outcomes of American youth compare to the previous generation? Will gaps between race and ethnic groups narrow or widen? How will other key outcomes, such as marriage and fertility and incarceration rates differ across generations? The answers to these questions will of course hinge in part on the broad changes in social processes, culture, government policy, and the economy that would be very difficult to forecast decades in advance. However, they will also be influenced by differences across generations in the distribution of cognitive and noncognitive skills that are already observable, at least in part. With information on the parental background, race and ethnicity, cognitive tests, educational outcomes, and early labor force outcomes and other behaviors, one can explain a substantial portion of the variance across people in employment rates, hourly wage rates, and hourly earnings and other outcomes at age 40. We will examine the implications of changes in the characteristics of American youth for a set of adult outcomes. Our focus on adult outcomes provides a natural way in which to aggregate variation across cohorts in observed skill distributions. We therefore generate a metric to compare and aggregate the variation in the skill distribution along the various dimensions of skill. The metric we choose are labor market outcomes (primarily wages and employment) of individuals of the NLSY79 cohort during the 1998-2004 survey years. To avoid misunderstanding, we should note that throughout the paper, we use the term "skill indicator" to refer to variables that are correlated with labor market outcomes. In the case of race and gender, part of that relation may be due to discrimination rather than to skill differences that correlated with race and gender.

The specific cohorts that we will compare are determined by the availability of panel data from National Longitudinal Survey of Youth, 1997 (NLSY97) on individuals who were aged 12 to 16 in 1997 and of panel data from the National Longitudinal Survey of Youth, 1979 (NLSY79) on individuals who were aged 14 to 22 in 1979. The most recent release of NLSY97 is 2005, when the sample was aged 20 to 24. The most recent release of NLSY79 is for 2004, when the cohort[1] was 39 to 47 years old. At this point the respondents typically have more than 10 years of experience[2].

The first step in our study is to create a set of measures of characteristics that are relevant to adult outcomes and are comparable across NLSY97 and NLSY79. The second step is to examine the consequences of differences between the characteristics of the 1979 and 1997 cohorts for various adult outcomes. Basically, we assess what the adult outcomes of the 1997 cohort will be if the relationship between characteristics and adult outcomes turns out to be the same for the 1997 cohort as it has been for 1979. To accomplish this, we draw on Dinardo, Fortin, and Lemieux ((1995), hereafter, DFL). Basically, we re-weight the 1979

[1]In this study we use the word cohort to refer to either the NLSY79 or the NSLY97. We will use the word birth-year to refer to groups of individuals defined by their birth year. [check that we actually do this consistently in the paper]

[2]Mean wages typically rise rapidly during the first 10 years of experience but show little change with experience subsequently.

sample to have the same distribution of characteristics as the 1997 sample, taking population weights into account. By comparing the distribution of outcomes in the re-weighted 1979 sample with the distribution prior to re-weighting, we can assess the effects of differences in the characteristics of 1979 and 1997 cohorts on the distribution consequences for the outcome of interest. For example, we can estimate what the distribution of the adult wage rates of the 1997 cohort will be if it faces the same labor market environment as the 1979 cohort. Furthermore, we can decompose the shift in the distribution into the contributions of various subsets of characteristics. An advantage of this general approach is that it permits one to assess the effects of particular characteristics without having to specify a detailed parametric model relating outcomes to characteristics.

As we will discuss in detail below, there are several key stumbling blocks that must be faced. The first is the comparability of the family background and education measures from the two data sets. The second is the comparability of the AFQT tests, which were based upon different forms of the test and were administered at somewhat different ages. The final key issue is the representativeness of the 1979 and 1997 cohorts, particularly after attrition and for missing data on two key variables are accounted for.

Our main results to date are as follows.

(1) First, the 1997 cohort is stronger than the 1979 cohort in most dimensions that matter for wages. In particular, the 1997 cohort is stronger than 1979 cohort in education, parental education, and test scores. However, the fraction of individuals who lived with both parents at age 14 declined. Increases in the representation of African American and Hispanics imply a slight reduction in adult wages, but are not a big factor.
(2) Second, the increase in skills of the younger cohort implies on average a wage increase of about 7% for whites and more for minorities.
(3) Third, the implied difference in employment rates are small, but they show an increase for males and drop for females, and Blacks and Hispanics gain relative to White males.
(4) Fourth, we observe a decline in skill gaps across race and gender. Black and hispanic males and females gain relative to their white counterparts. White women gain more than white men. The most important sources of the gains vary across race/gender groups.
(5) Finally, the results show a widening of the skill distribution with in race/sex groups.

Our work relates to large literature in macroeconomics (eg, Denison (1974) and Jorgensen et al (1987)) that measure changes in the characteristics of the American labor force on the basis of education, work hours, and change in the age and gender mix of the labor force for purposes such as growth accounting, the analysis of productivity, and the estimation of aggregate production functions and factor demand models. Many papers in economics, demography and sociology have examined trends in the skill characteristics and labor market outcomes of particular groups, such as blacks relative to whites and men versus women. Others examine trends in test scores, school quality, and field of study of various cohorts[3]. Below we provide a brief discussion on evidence on trends in some of the key skill indicators that play a role in our analysis.

The paper continues in section 2, where we present our methodology. In section 3, we describe the data. We also present evidence on and ways of accounting for biases that may arise due to problems with the NLS97 base year sample, missing data on key variables, and attrition. In section 4 we discuss the specifications of the probit models used to adjust the 1979 sample to match the characteristics of 1997. In sections 5 and 6 we present our basic results. In Section 7 we extend the analysis to incarceration and fertility and consider the role of immigration. (Not available for this draft.) In the final section, we summarize our main findings and outline the next steps in our project.

2. Econometric Methods

We now describe our procedure for assessing the changes in the skill distribution across the NLSY79 and NLSY97. There are many dimensions of skill one could look at. Skill measures such as years of education and test scores are not reported in a natural metric that allows us to aggregate variations in observed skill distributions. We measure and aggregate the contributions of the various skills using the labor market outcomes (primarily wages and employment) of the NLSY79 cohort during the 1998-2004 survey years. At this point the 1979 cohort has reached the peak of its life-cycle earnings profile[4]profile.

Our estimates of counterfactual wage distributions answer the question, "What will the wages of the NLSY97 cohort be if the wage distribution at the peak of life-cycle earnings conditional on skills is unchanged between the NLSY79 and NLSY97 cohort?" To answer this question we reweight the NLSY79 to have the same distribution of skills as the 1997 cohort and then use the reweighted data to generate the counterfactual wage distribution for the NLSY97 cohort. Our procedure is a straightforward modification of the decomposition procedure developed by DiNardo, Fortin, and Lemieux (1996, hereafter DFL).

2.1. **Basic Approach.** For each observation from the NLSY79 we obtain a realization (z, w) of the random vectors Z and W. Observations from the NLSY97 consist of realizations of Z only. We lack realizations of W for NLSY97 and strive to obtain counterfactual distributions of W. Let stand for the populations that the NLSY79 and NLSY97 are drawn from, respective[5]. We will sometimes refer to the cohorts by "1979" and "1997" rather than t and t'.

Wages in the economy faced by 1979 are determined by $w = W^{79}(z, u)$, where the vector z is observed and the vector u is not. The function $W^{79}(z, u)$ describes the mapping

[3]For example, see Bishop (1989), Smith and Welch ( ), Blau and Beller (1992), Grogger (19??), , Card and Kreuger (1996) and Neal (2005)

[4]Mean wages typically rise rapidly during the first 10 years of experience but show little change with experience subsequently. In 1998 even the youngest respondents in the NLSY79 typically have more than 10 years of labor market experience. At this point in time the respondents in the NLSY79 are between 33 and 41 years old.

[5]The birth years are 1957-1964 for NLSY79 and 1980-1984 for the NSLY97.

from observable and unobservable skills into wages for the 1979 cohort. It serves as our metric for aggregating components of the skill vector z.

Let $g(u|z,t)$ and $g(u|z,t')$ be the conditional densities of u given z for the two cohorts. We make the following key assumption on the relation between observed on unobserved skills:

Assumption A1: *The distribution of u conditional on z is the same for 1979 and 1997 cohorts:*

$$(A.1) \qquad g(u|z,t) \quad = \quad g(u|z,t')$$

This assumption allows us to construct a counterfactual distribution of wages using $W^{79}(z,u)$ and the observed distribution of Z for the 1997 cohort. A.1 is not likely to hold exactly. Behavioral responses to differences between t and t′ in skill prices, unobserved differences across cohorts in school quality, neighborhood environment, or family environment might lead the assumption to fail. Furthermore, changes in compulsory schooling laws, college tuition subsidies, or race and gender discrimination would alter the relationship between parental education and innate characteristics that are transmitted to children.

Let $f(w|t,z) = \int f(W^t(u,z)|t,z)g(u|z,t)du$ be the density of adult wages of the cohort t (the 1979 cohort) conditional on z. Assumption (A.1) implies that the conditional wage density for cohort t and t' are the same:

$$\text{(2.1)} \qquad \begin{aligned} f(w|t,z) &= \int f(W^t(u,z))g(u|z,t)du \\ &= \int f(W^t(u,z))g(u|z,t')du \\ &= f(w|t',z) \end{aligned}$$

In the remainder of the paper, we will suppress the "t" superscripts on the wage function W because we will always be considering the 1979 wage function[6].

Note that $f(w|t) = \int f(w,z|t)dz$. Equation (2.1) implies that[7]

$$\text{(2.2)} \qquad f(w|t') \quad = \quad \int f(w,z|t)\psi(z)dz$$

where

$$\text{(2.3)} \qquad \psi(z) = \frac{f(z|t')}{f(z|t)} = \frac{p(t'|z)}{p(t|z)}\frac{p(t)}{p(t')}$$

and $p(t'|z)$ and $p(t|z)$=1-$p(t'|z)$ are the probabilities or "propensity scores" of appearing in sample t' and sample t, respectively, conditional on z. The ratio $\frac{p(t)}{p(t')}$ is the unconditional odds that the observation is from cohort t. Thus the second equality in (2.3) says that $\psi(z)$ is

[6]If one were to assume a parametric form $W^{79}(z,u)$=$W(z,u;\beta_{79})$ and make assumptions about how β_{97} will relate to β_{79}, then one could use $W(z,u;\beta_{97})$ to forecast the wage distribution for the 97 cohort. We *do not* explore such assumptions and instead simply focus on the counterfactual distribution generated by changing endowments of observed skills between the 1979 and 1997 cohort.

[7]First note that $f(w|t') = \int f(w,z|t')dz = \int f(w|z,t')f(z|t')dz$. Then apply assumption (A1) to get $f(w|t') = \int f(w|z,t)f(z|t')dz$. Substitute $f(w|z,t) = \frac{f(w,z|t)}{f(z|t)}$ and then use $\frac{f(z|t')}{f(z|t)} * \frac{p(t')}{p(t)} = \frac{f(z,t')}{f(z,t)} = \frac{p(t'|z)}{p(t|z)} * \frac{f(z)}{f(z)} = \frac{p(t'|z)}{p(t|z)}$

also equal to the product of the odds that an observation comes from cohort t' conditional on z multiplied by $\frac{p(t)}{p(t')}$, the unconditional odds that the observation is from cohort t. The term following the first equality says that the weight function $\psi(z)$ may also be expressed as the relative frequency (density) of the skill vector z in 1997 versus 1979.

Equation (2.2) shows that under Assumption 1 one can obtain the density of adult wages for a population that faces the 1979 wage function but has the observed characteristics of the 1997 sample. One simply multiplies the density from t by the weight function $\psi(z)$. Clearly, multiplying by $\psi(z)$ also ensures that the reweighted distribution of skills for 1979 is identical to the actual distribution of skills in 1997.

We implement (2.2) as follows. First, we pool data from NLSY79 and NLSY97 and estimate the propensity score $p(t'|z)$ using skill measures Z that are observed for both the NLSY79 and the NLSY97 cohort. We then generate the 'propensity weight' $\psi(z)$ and apply these weights to the NLSY79 data. The reweighted data are used to generate various statistics of the counterfactual wage distribution $f(w|t')$. In particular, we estimate $f(w|t')$ itself and compare it to $f(w|t)$. We also derive the counterfactual wage distribution conditional on various sub-components of Z. Note, that in applying these procedures we must account for the sampling weights provided by the NLSY79 and NLSY97 to achieve population representative samples[8].

2.2. **Identifying the Contribution of Subsets of Variables to Differences between the 1979 and 1997 Wage Distributions.** The above procedure takes the observed wage distribution of the NLSY79 sample and transforms it into the counterfactual distribution that is implied by assumption (A.1) and by a shift from the distribution $f(z|t)$ to $f(z|t')$. It is natural to ask about the contribution of various partitions of the random vector Z (say Z_1 and Z_2) to the overall changes from $f(w|t)$ to $f(w|t')$. That is, one would like to decompose the differences between the observed wage distribution of the NLSY79 sample and the counterfactual wage distribution of the NLSY97 sample into variation due to Z_1 or Z_2.

The first thing to note is that it is in general not possible to arrive at a unique decomposition of the change in $f(w|t)$ or even the changes in any statistic of $f(w|t)$ into "contributions of Z_1 and Z_2". Any decomposition is of interest only in so far that is describes the data in a way that corresponds to an implicit or explicit understanding of the underlying structure of the economy.

Below we propose a decomposition of the change from $f(w|t)$ to $f(w|t')$ that is based on a partition of the skill vector Z into subgroups that can be interpreted as determined hierarchically in the sense that Z_1 is determined prior to Z_2 . We believe this decomposition provides useful insights into the changing skill distribution between 1979 and 1997.

[8]We also generate weights to account for attrition and non-response for crucial variables. Details are provided in Section 3.

To present the decomposition, we need some additional notation. Partition Z into the subvectors Z_1 and Z_2. Let $f(z_1|t)$ and $f(z_1|t')$ denote the pdf for Z_1 conditional on membership in t and in t', respectively, and let $F(z_1|t)$ and $F(z_1|t')$ denote the corresponding cdfs. Let $f(w|t, Z_2(t'))$ denote the density function of w at t if the distribution of Z_2 is that of t' but the pdf of z_1 conditional on z_2 is that of t, $f(z_1|z_2, t)$. It is defined as:

$$\begin{aligned} f(w|t, Z_2(t')) &= \int\int f(w, z|t) f(z_2|t') dz_1 dz_2 \\ &= \int\int f(w|z, t) f(z_1|z_2, t) f(z_2|t') dz_1 dz_2 \end{aligned}$$

Note that by definition, $f(w|t)$=$f(w|t, Z(t))$ and by Assumption 1:

$$\begin{aligned} f(w|t') &= f(w|t, Z(t')) \\ &= \int f(w|z, t) f(z|t') dz \\ &= \int\int f(w|z, t) f(z_1|z_2, t') f(z_2|t') dz_1 dz_2 \\ &= \int f(w|z, t) \psi(z) f(z_1|z_2, t) f(z_2|t) dz_1 dz_2 \end{aligned} \tag{2.4}$$

We can similarly define:

$$\begin{aligned} f(w, z_1|t, Z_2(t')) &= \int f(w, z_1|z_2, t) f(z_2|t') dz_2 \\ &= \int f(w, z_1|z_2, t) \psi(z_2) f(z_2|t) dz_2 \end{aligned} \tag{2.5}$$

and

$$\begin{aligned} f(w|t, Z_1(t), Z_2(t')) &= \int\int f(w|t, z_1, z_2) f(z_1|z_2, t) f(z_2|t') dz_1 dz_2 \\ &= \int\int f(w|t, z_1, z_2) f(z_1|z_2, t) \psi(z_2) f(z_2|t) dz_1 dz_2 \\ &= \int\int f(w|t, z_1, z_2) f(z_1, z_2|t) \psi(z_2) dz_1 dz_2 \end{aligned} \tag{2.6}$$

The expression (2.6) gives the distribution of w if (i) the distribution of (w, z_1) conditional on z_2 is that of period t and (ii) the marginal distribution of z_2 is that of period t'. To obtain $f(w|t, Z_1(t), Z_2(t'))$ using (2.6), we estimate propensity weights based on z_2 only. We can then decompose the difference in the 1997 and 1979 wage densities as:

$$\begin{aligned} f(w|t') - f(w|t) &= f(w|Z_1(t'), Z_2(t')) - f(w|Z_1(t), Z_2(t')) \\ &+ f(w|Z_1(t), Z_2(t')) - f(w|Z_1(t), Z_2(t)) \end{aligned} \tag{2.7}$$

The second term in brackets on the RHS of (2.7) is the contribution of the change between t and t' in the distribution of Z_2, holding the distribution of Z_1 given Z_2 constant. The first term is the consequence of the change in the distribution of Z_1 given Z_2 evaluated at $f(z_2|t')$.

Comparing the terms in (2.7) with the expressions (2.4) and (2.6) shows that we can estimate the components in (2.7) by sequentially applying the propensity weights. One obtains the second term $f(w|Z_1(t), Z_2(t')) - f(w|Z_1(t), Z_2(t))$ by first applying $\psi(z_2)$ to the NLSY79 data, as in (2.6), to obtain $f(w|Z_1(t), Z_2(t'))$ and then subtracting $f(w|t) = f(w|Z_1(t), Z_2(t)$. To obtain the first term, one applies $\psi(z_1, z_2) = \psi(z)$ to the NLSY79 data to obtain $f(w|t')$ and subtracts $f(w|Z_1(t), Z_2(t'))$. Note that this decomposition procedure can be undertaken sequentially with as many subvectors as desired.

Note that in general the decomposition will be sensitive to the order in which it is performed. An alternative decomposition to (2.7) is given by:

$$\begin{aligned} f(w|t') - f(w|t) &= f(w|Z_1(t'), Z_2(t')) - f(w|Z_1(t'), Z_2(t)) \\ &+ f(w|Z_1(t'), Z_2(t)) - f(w|Z_1(t), Z_2(t)) \end{aligned} \tag{2.8}$$

The first term of this alternative decomposition and the second term of (2.7) both provide a measure of the contribution of the change in the distribution of Z_2 to the change in the distribution of w between t and t'. However, these measures are not identical. The first term of (2.8) will be equal to the second term of (2.7) only under special circumstances (such as independence between Z_1 and Z_2).

The fact that the decomposition is sensitive to order is not a "disadvantage" of the decomposition method. Rather, it is a consequence of the assumptions implicit in performing a decomposition based on particular hierarchical structure of Z. If we vary the order of Z and include for instance schooling before parental education, then we make the implicit assumption that schooling is responsible for part of the shift in parental education. It is natural that variations in implicit assumptions are reflected in the results of the decomposition.

We start by including race and gender in the prediction model, then add parental background variables and then add variables capturing individual characteristics such as education and cognitive ability scores. Finally we add variables describing the transition into the work force. The decompositions based on (2.7) assume that the relation between parental background variables and individual characteristics is the same in 1979 and 1997. Thus, changing distributions of parental background will entail changes in the resulting individual education and ability distributions. The decomposition therefore implicitly assumes that the cross-sectional relation between background variables and education and ability in 1979 is causal in the sense that changes in the distribution of parental background result in changes in the individual variables. Similar assumptions are made regarding the relation between parental background, individual education and ability scores on the one hand and the variables describing the speed with which individuals transition into the workforce on the other hand. Clearly these are strong assumptions.[9]

A limitation of the DFL type reweighting technique that we employ is that it does not permit one to easily distinguish whether a variable is important because it has a strong

[9]Belley and Lochner (2007) present evidence that the link between parental background and educational attainment is stronger for the NLSY97 cohort and the NLSY79 cohort.

link with wages or because the change in its distribution between the 1979 and 1997 cohorts is large. Furthermore, (2.7) does not permit one to examine the effect of a shift in the marginal distribution of Z_1 holding the marginal distribution of Z_2 constant. Even when there is a natural causal ordering to the variables, as is the case with race and gender and parental background "causing" HGC, afqt, and the school to work transition measures, it is interesting to ask how the NLSY97 cohort would stack up against NLSY79 if parent education had the 1997 distribution the distributions of the other variables was as in 1979. Unless the feature of w that is of interest is and additively separable function of Z_1 and Z_2, one cannot analyze the effect of a change in the distribution of Z_2 without specifying the change in the joint distribution of Z_1 and Z_2 [10]. This is true even for the mean of the wage. The models discussed in Section 2 do not restrict the functional form relating w to z_1, z_2, and u. However, if one is willing to impose additive separability restrictions, then one can analyze the effects of a shift in the distribution of z_1 holding z_2 fixed, and vice versus. In Section 5.4 we impose additive separability restrictions on the conditional mean function and use multivariate linear regression as the basis for a decomposition of the overall shift in the mean.

3. Data

The above procedure requires comparable skill measures across surveys. The NLSY79 and NLSY97 surveys are designed for the same purpose: examining the transition of young Americans into the work-place. Consequently, these surveys provide many variables that are comparable across both the 1979 and 1997 cohorts. Nevertheless, the surveys vary sufficiently to pose challenges to achieving comparability. In this section we describe the samples and the variables used in the analysis.

3.1. The Samples.

We use survey years 1979-2004 for the NLSY97 and 1997-2004 for the NLSY97, which were the latest available at the time we created the data sets for this paper. To maximize sample sizes for minority groups we utilize both the cross-sectional samples and the supplemental samples in the NLSY79 and NLSY97 and use the base year weights provided by the BLS to achieve representativeness of the population[11]. We omit oversampled whites and the military sample from the supplemental sample of the NLSY79. These samples have been discontinued by the BLS in 1990 and 1984 respectively, and so do not provide labor

[10] The Oaxaca-Blinder decomposition applies to the case when the statistic of interest is the mean and $W(Z,u)$ is linear in Z_1, Z_2, and u. It then decomposes changes in the mean of W into a change due to a shift in the marginal distribution of Z_1 that leaves the mean of Z_2 unchanged as well as the corresponding shift of the marginal distribution of Z_2 that keeps the mean of Z_1 fixed. The Oaxaca-Blinder decomposition is unique in the sense that any such shifts in the marginal distributions will deliver the same contribution for Z_1 and Z_2 respectively. The Oaxaca-Blinder decomposition is not unique if we consider decomposing the change in $f(Z_1, Z_2|t') - f(Z_1, Z_2|t')$ into changes that do not keep the means of the respective variables unchanged. Furthermore, an analog for the Oaxaca-Blinder decomposition does not exist for statistics that are not linear in Z_1, Z_2 or if $W(Z,u)$ is not itself linear.

[11] We do not utilize the panel-weights that are designed to account for (conditionally random) attrition but instead estimate our own weights, as discussed below.

market outcomes in the age range that we use[12]. In the NLSY79 there are 3,650 in the supplemental sample and 6,111 in the cross-section. In NLSY97 the supplemental and cross-section samples contain 2,236 and 6,712 respondents respectively.

We match the distribution of characteristics of individuals who are observed at age 22. That is, in both surveys we construct our skill measures in a similar manner using the waves up to the survey year when these individuals were 22. We retain the observation that is closest to 22 years and 6 month old and then measure variables such as highest grade completed and early work experience by reference to this observation[13]. While all base year respondents of the NLSY79 cohort have reached age 22, only 5,406 out of a total of 8,984 NLSY97 respondents had turned 22 by the 2005 survey (all individuals born in 1981 or 1982 and some born in 1983). We thus have a base-line sample of 9,761 respondents for the NLSY79 cohorts and 5,406 respondents for the NLSY97.

3.2. **Representativeness of the Base Year Samples.** In this section we discuss whether the NLSY97 base year sample is representative and then turn to the problems of attrition and of nonresponse to the AFQT in Section 3.3. MaCurdy and Vytlacil (2003) have raised concerns about the representativeness of NLSY97. In particular, they document and consider the implications of the fact that the NORC's screening procedures for the NLSY97 found less than two-thirds of the young adults one would have expected to be present based on the 1997 Current Population Survey (CPS) in the age range that the screener interviews sought to identify (12-23) but found the expected amount in other ages. Apparently, families were "hiding" children in the 12-23 age range, perhaps to avoid participating in the survey. MaCurdy and Vytlacil analyze ETP97, a related sample drawn on 18-23 year olds from the same screening survey, and find that those responding to the ETP97 are more educated than comparable CPS respondents. They also have more educated mothers[14]. Moore et al's (2000) technical sampling report on NLSY97 also concludes that many parents failed to report children in the NLSY97 age range. They also discuss the substantial shortfall in the completion rate for the ASVAB. However they conclude that the distribution of respondents in the screening interviews and the CPS is similar in the dimensions of youth education, parental income and parental education.

We do not fully understand the sources of the differences between the two studies. One difference may arise from the fact that in the CPS mother's education is only available

[12]Furthermore, the military sample receives a very sample weight and would be of limited utility in estimating propensity weights.

[13]The interviews of a given individual are not exactly one year apart. Consequently, some individuals respond twice at age 22 and some do not respond at age 22 but instead are surveyed twice at age 21 or twice age 23. We retain the observation that is closed to 22 years and 6 month old and then measure variables such as highest grade completed and early work experience as of this age=22 observation.

[14]Their comparisons of the PAY80 and PAY97, which are also drawn from the same screening surveys as the NSLY79 and NLSY97 (respectively) show that the fraction of the youths who completed the ASVAB tests and for whom we therefore have an AFQT test score is significantly lower in the PAY97 than in the PAY80. This evidence for PAY80 and PAY97 is consistent with the evidence for NSLY79 and NLSY97 in Table 2.

for 18-23 years olds who are still living with their mothers. These youths may not be representative of 18-23-year-olds as a whole. In this case, McCurdy and Vytlacil's comparison of ETP97 to the CPS may not be directly relevant for the NLSY97 sample of 12-16 year olds.

We proceed under the assumption that the available data, after use of survey weights and adjustments for attrition prior to age 22 and for missing data on AFQT, are representative of the 1997 and 1979 populations, with the obvious caveat that our results will be affected if they are not. In a future draft we will draw on other data, including NAEP, SIPP and the CPS to confirm any patterns in skill characteristics found in the NLSY.

3.3. **Attrition and Missing Data on the AFQT.** Table 3.0 summarizes the retention patterns for our sample in the NLSY97 and NLSY79. There are 9,761 (5,406) individuals in the NLSY79 (NLSY97) who should have been observed at age 22 and thus fall with in the scope of our study. 435 (702) of these individuals left the sample by age 22. A negligible 29 (50) respondents are lost due to missing information on highest grade completed. This leaves 9,297 (4,654) observations. Non-participation in the ASVAB eliminates an additional 397 (879) respondents. We retain a total of 8,929 NLSY79 respondents and 3,775 NLSY97 respondents, which constitute 91.21% of the total eligible sample in the NLSY79 but only 69.83% for the NLSY97.

The NLSY97 has a lower retention rate than the NLSY79 at each step of the construction of our sample. In the case of attrition by age 22 this is partly due to the fact that NLSY97 respondents are first interviewed during age 12-16 whereas those in the NLSY79 are first interviewed aged 14-21. The respondents in the NLSY97 had more time to attrit. In the NLSY97 we loose the largest share of respondents when we consider the availability of the AFQT score. If we do not condition on observing the AFQT score, we retain more than 86% of the base sample. As a robustness check, we analyze a number of specifications that do not require the AFQT score on both our main sample (sample 1) and on a sample that includes those with missing AFQT scores (sample 2).

Table 3.1 provides information about differences in observable characteristics across attriters and stayers by age 22. Several of the characteristics are related to whether an individual remains in the survey until age 22. For instance, attrition prior to age 22 is related to race, especially in the 1997 sample. However, the attrition rates are not always negatively associated with characteristics that are favorable for wages. For example, whites are more likely to leave the sample prior to age 22 than are blacks. There are other important differences by observable characteristics in attrition patterns. In NLSY97 there is a substantially higher rate of attrition among the cross-sectional sample than among the supplemental sample. The difference in the fraction of individuals who live with neither their biological mother nor father at age 14 is 6.2%. Attriters from NLSY97 have less educated mothers, although the magnitude of the difference is not very large.

The average characteristics of those who remain in the sample to age 22 are very close to the averages for full population represented by NLSY79, in part because we lose only 4.55% of the sample. We also find relatively small differences between the full sample and the stayers in the 1997 cohort in spite of the higher 1997 attrition rate. For instance, the differences between the full population and stayers in the means of mother's education and father's education are only .04 and .01 respectively. Nevertheless, we adjust for attrition based on observables using weights based on a probit model relating attrition to parental education, parental presence at age 14, a polynomial in the birth-cohort, and indicator variables for race and sex.

ASVAB is large enough to have the potential to result in significant biases, especially in the NLSY97. Table 3.2 presents summary statistics for the distribution of skill characteristics conditional on cohort and on whether an individual has a valid observation for the AFQT or not. The structure of the table is similar to table 3.1. Since Table 3.2 is meant to inform us about the additional error introduced by non-response to the ASVAB, the figures in table account for attrition by age 22 using the weighting procedure described in the previous paragraph.

The differences in the mean characteristics by availability in the NLSY79 and NLSY97 are not uniformly larger in the NLSY97, but the fraction with missing AFQT scores is 3 times larger in the NLSY97. Some of the differences between those with or without an AFQT-score are sizable. The difference in racial composition is particularly striking: Whites are substantially overrepresented among those with valid AFQT scores. Consequently the share of Blacks and Hispanics is about 50% larger among the population missing an AFQT score. Furthermore, those who have AFQT scores have higher education levels by age 22 and have better educated parents. Overall, those with AFQT scores are more advantaged in both the NLSY79 and the NLSY97.

However, the difference in characteristics between those with and without the AFQT dramatically overstates the bias in mean characteristics obtained when estimating population means using those with valid scores only. For instance, those with valid AFQT scores in 1997 have $\frac{3}{4}$ of a year more education by age 22 than those without valid scores but only 1/6th of a year more education than the full population. We judge these differences to be sizeable, but not forbidding.

Our main approach to the problem of attrition and non-response to the AFQT is to generate two sets of weights. The first adjusts for attrition by age 22 whereas the second adjusts for both attrition and the possibility of missing responses to the AFQT. These weights are estimated using probit specifications based on race, sex, parental presence at age 14, parental education, and a quadratic in the birth year. We estimate these attrition models for the NLSY79 and the NLSY97 separately and apply these weights thoughout the analysis as applicable[15]. We also check whether our results for the models outlined in Section 4 that do not require an AFQT score are sensitive to excluding individuals with missing data on the AFQT score. The fact that they are not sensitive is reassuring.

[15]In fact the results in table 3.2 are generated using the attrition weights and thus display the attrition corrected differences across those with and without the AFQT-score among those who do not attrite by age 22 ("sample 1" and "sample 2", respectively).

It is important to acknowledge the limitations of our attempts to account for attrition and missing AFQT-scores. We can only adjust for attrition and non-response to the extent that it is predicted by observable characteristics[16]. This will not correct for possible correlation between attrition and unobservables that affect wages or employment conditional on other observable skill measures in the model. And we can only check for robustness for specifications that do not include the AFQT-score.

A final problem arises because 859 out of 9,326 respondents at age 22 do not respond at any time during the time-period 1998 and 2004. We use these individuals to estimate the propensity weights, but we cannot use them for generating the counterfactual wage distributions. The results presented below assume that attrition from NLSY79 after age 22 is random. It would be possible to construct an additional weight to adjust for this but we do not do so. Since attrition after age 22 in the NLY79 affects both the actual wage distribution and the counterfactual one, it probably has only a second order affect on the difference between the two, which is our main interest.

3.4. **Wages.** Our main metric to measure skills are wages of the NLSY79 cohort in the years 1998-2004. This period spans 4 survey years, since the NLSY79 moved to a biannual format in 1994. We use a regression specification to standardize log real wages between 1998-2004 to the year 2002 and 23 years of potential experience[17]. For each individual we have between 1 and 4 wage observations. We weight each wage observation by the reciprocal of the number of wage observations for an individual. Each individual with at least one valid wage observation has equal weight in generating the counterfactual wage distributions. This implies that our wage statistics reflect the wage distribution of the population "while working"[18]. There is no need to reweight the statistics for employment rates, because employment status is observed for each valid interview.

3.5. **Comparability of the Paper and Pencil (P&P) and Computer Assisted Test (CAT) of the ASVAB.** Our measure of cognitive ability, the AFQT-score is a composite score derived from the ASVAB. For the NLSY79 the test was administered in 1981 when respondents were between 16-21 years old, whereas for the NLSY97 the test was administered at the beginning, when individuals were between 12-16 years old. We exploit the overlap in the test-taking age

[16]For the attrition and missing AFQT adjustments we could make use of a much larger set of base year characteristics than we are currently using, because we are not restricted to variables that are also available and consistently measured in both cohorts. We will explore this in a future draft.

[17]For this purpose we estimate a log wage equation separately for high school drop-outs, high school graduates and individuals with more than a high school degree. We include a quartic in experience and year-effects.

[18]If we do not weight, then the statistics are representative of those who are working at a point in time. Consequently, they give more weight to those who work more frequently. Most studies use a cross section of workers with wages in a give year thus are representative of those actually working at a point in time. We weight person year observations rather than use time averages for each individual because in the presence of transitory wage variation, the distribution of the time averages will depend on the distribution of the number of observations per person.

across both cohorts by applying a equipercentile procedure on each cohort with the population of test takers who were 16 year olds when taking the test.

We must also account for the fact that the test-format differed between surveys. The NLSY79 cohort was administered a pencil and paper (P&P) version of the ASVAB while the NLSY97 took a computer assisted test (CAT) format. To achieve comparability between the two test formats we rely on a mapping between the P&P and the CAT test format[19].

In Appendix B, we present evidence on the comparability of the upgraded, age-adjusted test scores by presenting simple regressions of a set of personal characteristics measured at age 22 on the test scores, by survey and by age at which the ASVAB was taken. We find that the relation between AFQT scores and various key independent variables is the same across test taking ages. This suggests that the equipercentile procedure mapping of AFQT scores for each age cohort to the cohort of 16 year olds does not introduce systematic biases into the analysis. Appendix B says little about the quality of the mapping from the P&P to the CAT version of the test, which is considered in Segall (1997).

3.6. **School to Work Transition Variables.** We construct 2 sets of variables that measure whether an individual's schooling career was continuous and whether the individual had a smooth transition into the working population. These variables are generated for those within the sample of 22 year olds who did not attend school between ages 20-22. For these individuals we compare the observed age of graduation with the expected age of graduation using the formula age+highest grade completed+6. We thus generate a measure of whether an individual's school leaving age is older or younger than expected given the highest grade they attended. We also construct a measure of whether they have worked at least 14 weeks in either the first or second year after leaving school.

3.7. **Summary Statistics.** Table 3.3 presents summary statistics for the key variables used in this study. Details of the construction of these variables are available in the data appendix.

Most characteristics show improvement between NLSY79 and NLSY97. For example, father's education rises from 11.9 to 12.99. The mean of AFQT rises from 42.02 to 45.28 and highest grade completed as of age 22 increases from 12.61 to 13.17. On the other hand, the percentage of individuals living with both parents at age 14 drops from 75.07% in 1979 to only 48.6% in 1997.

Panel B of the table summarizes the main characteristics for whites, blacks, and Hispanics. The changes for the subgroups generally parallel those for the sample as a whole. However, the size of the increase in AFQT scores is substantially larger for blacks and Hispanics than for whites. Parental education improves by more than two years for Hispanics and by about 1.5 years for blacks.

[19]The mapping was constructed using test results from a sample that was randomly assigned to take either P&P or the CAT test. (See Segall 1997) The mapping assigns scores to equalizes percentiles on the various subtests of the P&P and the CAT. By definition this amounts to transforming the P&P subtest scores with a monotone function that matches the distribution among the CAT. We thank Daniel Segall for providing us with the P&P equivalents of the CAT scores for the NLY97 sample

Table 3.4 takes a closer at education and reports statistics by race and gender. The results show increases in education for most measures across all race and demographic groups. Hispanics females close the gap with blacks and Hispanic males even overtake Black males for some education. Black males are the only group that does not show uniform increases across education measures (with the exception of the GED).

Overall, the changes in characteristics suggest that the NLSY 97 cohort is more skilled than the previous cohort.

4. Estimation of Propensity Scores

We now turn to the estimation of the propensity score conditional on our available skill measures. Equation (2.3) shows the relation of the propensity weights to the estimated propensity scores. We estimate the propensity score using Probits based on various sets of skill measures. We use flexible functional forms for the latent index of the Probit model so as not to restrict the changes in the skill distributions across cohorts unduly.

We consider various specifications for the skill vector Z. We organize the variables used to estimate the propensity score in a hierachical structure according to the degree of predetermination. Our most basic skill vector consists of variables that are outside the individual's control (race and sex). We then sequentially add additional variables describing individuals skills. Each set of additional variables is fully interacted with race and gender. First we add parental education and the indicators for presence of mother and father at age 14 as measures that influence skill development and economic decision-making across generations, but are predetermined relative to the skill characteristics referring to the individual itself. Since changing social norms regarding childbearing out of wedlock may alter the relationship between the parental presence indicators and unobserved characteristics of family background, we experiment with excluding the parental presence indicators. Next we add a quadratic in the AFQT score. If cognitive skills are fully determined by inherited factors, environmental factors and primary schooling and are not amenable to individual investments after the early teens, then this variable will be predetermined relative to variables referring to educational attainment and the transition to the workplace. We add education next as measured by highest grade completed at age 22 as well as indicator variables for whether individuals are enrolled at age 22. To the extent that cognitive tests scores are influenced by high school and college education, as suggested by a number of studies, one might want to reverse the order of AFQT and Education[20]. We experiment with the order in which we include these variables and find that our results are typically robust to switching the order or including both the AFQT and individual schooling at the same time. Finally we add the measures referring to the early work experience and the coherence of the school career. These variables are meant to measure how quickly individuals transitioned from school to the work-place and whether the school career itself was uninterrupted. We conjecture that spending time neither at work nor at school is a negative indicator for future employment and wage rates.

[20] See for example, Hansen, Heckman, Mullen (2004), Neal and Johnson (1996) and Korenman and Winship (2000)

Table 4.1

Summary of Skill Vectors used in the Analysis

Specification	Components	Description
Race/Sex (Equation 4.1)	Race and sex fully interacted.	3 race categories: white, black and hispanic.
Race/Sex, Par_ED, Par_Present (Equation 4.2)	(4.1) + Parental Education and Mother and Father indicators. (fully interacted with race and sex)	HGC dummies for mother, HGC dummies for father. Missing values treated as separate category. Dummies for mother present at age 14, Father present at age 14, and Father and Father present at age 14.
Race/Sex, Par. Ed, Par. Present AFQT (Equation 4.3)	(4.2) + quadratic in AFQT score (fully interacted with race and sex)	Quadratic in standardized AFQT score. AFQT scores are made comparable across cohorst using the equipercentile mappings discussed in Section 3.5
Race/Sex, Par. Ed, Par. Present HGC (Equation 4.4)	(4.3) + Individual Schooling (fully interacted with race and sex)	1. Dummies for HGC (grades less than 8 are aggregated together with grade 8) 2. Enrollment status at age 22.
Race/Sex, Par. Ed, Par. Present, AFQT, ED (Equation 4.5)	(4.2) + Individual schooling and quadratic in AFQT (fully interacted with race and sex)	see controls from equation (4.3) and (4.4)
Race/Sex, Par. Ed, Par Present, AFQT, ED, School/Work (Equation 4.6)	(4.5) + School/Work Transition (fully interacted with race and sex)	For those who leave school before age ≤ 20 and do not return to by age 22: 1. Indicators for whether the individual graduate at an age less than, equal to, or older than "age+6+hgc"; 2. Indicator for whether respondent works at least 14 weeks in one year during the first two years after leaving school.

Equations (4.1) - (4.6) present the exact specifications of the latent index we estimate to obtain the propensity weights. We specify the distribution of the error in the latent index function to be normal and estimate the models by MLE-Probit after pooling the data from both samples. Sample weights provided by the NLSY for both surveys are used in the estimation. The sample weights are adjusted for attrition by age 22 based on observable characteristics (race, sex, parental presence at age 14, parental education, and a quadratic in the birth year) in both cohorts for specifications estimated on sample 2. All specifications estimated on sample 1, which excludes individuals without valid AFQT scores, are estimated with weights adjusting for both attrition by age 22 and AFQT non-response using the same variables. All specifications estimated on sample 2 account for attrition by age 22. These sample and attrition weights are omitted from the presentation here to keep the notation simpler.

We specify the probit models as follows. Denote the latent variable indicating whether an individual is observed in the NLSY97 rather than the NLSY79 as . Equation (4.1) shows the most basic formulation, which is based on race and gender only.

$$\theta_i = \sum_{s,r} Sex_{s,i} Race_{r,i} \beta_{s,r} + \epsilon_i \tag{4.1}$$

Equation (4.2) shows the specification including parental education. The variables MHGCe,i and FHGCe,i are dummy variables for highest grade completed, indexed by e. Missing data on parental education are treated as a category and thus observations with missing parental education are included in the analysis.

$$\begin{aligned} \theta_i &= \sum_{s,r} Sex_{s,i} Race_{r,i} (\sum_e FHGC_{e,i} \beta_{fe,s,r} + MHGC_{e,i} \beta_{me,s,r}) \\ &+ \sum_{s,r} Sex_{s,i} Race_{r,i} (\textit{Mother only} \beta_{mo,s,r} + \textit{Father only} \beta_{fo,s,r}) \\ &+ \sum_{s,r} Sex_{s,i} Race_{r,i} (\textit{Mother \& Father} \beta_{fm,s,r}) \end{aligned} \tag{4.2}$$

Specification (4.3) adds an interaction of a quadratic in the AFQT with a full set of race and sex to specification (4.2).

$$\begin{aligned} \theta_i &= \sum_{s,r} Sex_{s,i} Race_{r,i} (\sum_e FHGC_{e,i} \beta_{fe,s,r} + MHGC_{e,i} \beta_{me,s,r}) \\ &+ \sum_{s,r} Sex_{s,i} Race_{r,i} (\textit{Mother only} \beta_{mo,s,r} + \textit{Father only} \beta_{fo,s,r}) \\ &+ \sum_{s,r} Sex_{s,i} Race_{r,i} (\textit{Mother \& Father} \beta_{fm,s,r}) \\ &+ \sum_{s,r} Sex_{s,i} Race_{r,i} (afqt_i \beta_{1afqt,s,r} + afqt^2 \beta_{2afqt,s,r}) + \epsilon_i \end{aligned} \tag{4.3}$$

The alternative is to first augment (4.2) with education measures as indicated in (4.4). The dummy variables $hgc_{e,i}$ refers to highest grade completed of an individual as of age 22 and are indexed by e. The indicator variable *enroll* designates individuals still enrolled in school at age 22.

$$\begin{aligned}
\theta_i &= \sum_{s,r} Sex_{s,i} Race_{r,i} (\sum_e FHGC_{e,i}\beta_{fe,s,r} + MHGC_{e,i}\beta_{me,s,r}) \\
&+ \sum_{s,r} Sex_{s,i} Race_{r,i} (\textit{Mother only}\beta_{mo,s,r} + \textit{Father only}\beta_{fo,s,r}) \\
&+ \sum_{s,r} Sex_{s,i} Race_{r,i} (\textit{Mother \& Father}\beta_{fm,s,r}) \\
&+ \sum_{s,r} Sex_{s,i} Race_{r,i} (HGC_i\beta_{hgc,s,r} + enroll_i\beta_{enr,s,r}) + \epsilon_i
\end{aligned} \tag{4.4}$$

Specification (4.5) then contains both the quadratic in the AFQT and the individual schooling variable. One of the surprising results of our decomposition exercise below is that the order of AFQT and hgc is in fact relatively unimportant once parental education has been accounted for.

$$\begin{aligned}
\theta_i &= \sum_{s,r} Sex_{s,i} Race_{r,i} (\sum_e FHGC_{e,i}\beta_{fe,s,r} + MHGC_{e,i}\beta_{me,s,r}) \\
&+ \sum_{s,r} Sex_{s,i} Race_{r,i} (\textit{Mother only}\beta_{mo,s,r} + \textit{Father only}\beta_{fo,s,r}) \\
&+ \sum_{s,r} Sex_{s,i} Race_{r,i} (\textit{Mother \& Father}\beta_{fm,s,r}) \\
&+ \sum_{s,r} Sex_{s,i} Race_{r,i} (HGC_i\beta_{hgc,s,r} + enroll_i\beta_{enr,s,r}) \\
&+ \sum_{s,r} Sex_{s,i} Race_{r,i} (afqt_i\beta_{1afqt,s,r} + afqt^2\beta_{2afqt,s,r}) + \epsilon_i
\end{aligned} \tag{4.5}$$

Finally, specification (4.6) adds variables describing the transition from schooling to work. The variable *workuniv* is an indicator variable that takes the value 1 if an individual has not attended school for at least 2 years prior to age 22 and 0 otherwise. The variables (early, on time, late) are indicator variables generated by comparing school leaving age for these individuals with age+6+highest grade completed. The variable $work$ indicates whether individuals have worked for at least 14 weeks in one of the first two years following graduation:

$$
\begin{aligned}
(4.6) \qquad \theta_i &= \sum_{s,r} Sex_{s,i}Race_{r,i}(\sum_{e} FHGC_{e,i}\beta_{fe,s,r} + MHGC_{e,i}\beta_{me,s,r}) \\
&+ \sum_{s,r} Sex_{s,i}Race_{r,i}(\textit{Mother only}\beta_{mo,s,r} + \textit{Father only}\beta_{fo,s,r}) \\
&+ \sum_{s,r} Sex_{s,i}Race_{r,i}(\textit{Mother \& Father}\beta_{fm,s,r}) \\
&+ \sum_{s,r} Sex_{s,i}Race_{r,i}(HGC_i\beta_{hgc,s,r} + enroll_i\beta_{enr,s,r}) \\
&+ \sum_{s,r} Sex_{s,i}Race_{r,i}(afqt_i\beta_{1afqt,s,r} + afqt^2\beta_{2afqt,s,r}) \\
&+ workuniv_i * \sum_{s,r} Sex_{s,i}Race_{r,i}(early_i\beta_{ear,s,r} + ontime_i\beta_{ot,s,}) \\
&+ workuniv_i * \sum_{s,r} Sex_{s,i}Race_{r,i}(late_i\beta_{lt,s,r} + work_i\beta_{w,s,r}) + \epsilon_i
\end{aligned}
$$

The estimates of the propensity score parameters are hard to interpret because of the large number of interaction terms and the fact that they correspond to the probit index rather than marginal effects on probabilities. The next Section describes our findings regarding the changing skill distributions using labor market outcomes in middle age as our metric.

5. Changes in the Skill Distribution between NLSY79 and NLSY97

In this section we (finally) present the overall changes in the skill distribution across cohorts as measured using the methodology and data presented in Section 2-4.

5.1. **Overall increase in skills.** The first result to note is that the 1997 cohort is more skilled than is the 1979. Table 5.1 and 5.2 show how employment and log wages are predicted to change due to the variation in skill endowments across the two cohorts. Columns 1 and 2 present the results for the observed outcomes in the 1979 cohorts. The remaining columns of the tables present the difference between counterfactual statistics and the actual 1979 values. In column 3 we match on the full set of variables including parental education, parental presence, schooling, the AFQT, and work transition as well as race and gender. Columns 4 and 5 refer to the specification without the work transition variables and without the AFQT score on sample 1 and 2 separately. Column 6 omits the work transition variables from the full specification. We show bootstrapped standard errors in parentheses[21]. Generally results are quite robust across specifications.

[21]We choose bootstrap samples by selecting individuals with replacement from subsamples stratified by race and ethnicity and gender so as to preserve the basic demographic composition of the samples. Each replication sample consists of a bootstrap sample from the NLSY 79 and NLSY 97 *after estimating weights for attrition and AFQT-nonresponse*. We then applied all of our procedures to the replication sample. We repeated this process 200 times.

Using the metric of employment rates we find little difference across cohorts. For men and woman combined, the shift in skill components would lead to an increase in the employment rate of about .003 when we use the full set of skill indicators (model 4.6, bottom row, col. 3). This is the net result of an increase of .006 for males and a decrease of .003 for women. These results are not very sensitive to choice of skill components. We find more substantial increases for Black men and for Hispanic men and women, particularly when we use the full set of skill indicators. For example, the shift in skill components for black men implies an increase in employment from .854 to .889.

Using the metric of wages we find that the 1997 cohort is stronger than the 1979 cohort. Table 5.2 indicates that the average increase in skills between 1979 and 1997 is about 7-8%. Using the full specification (column 3) we find that skills have increase by on average 0.079 log points (bottom row). This finding of an increase in overall skills depends to some extent on the skill components on which the cohorts are matched. It is bit larger when AFQT and the work transition variables are excluded from Z (Model 4.4). It is somewhat smaller when we use Model 4.4 with sample 2, which includes person with missing AFQT scores. However, the results at all of the quantiles shown are much less sensitive to choice of sample for model 4.4 than the mean is. We are currently investigating whether the discrepancy is due to a small number of observations that get a very large weight when we use sample 1 to estimate (4.4).

5.2. **The Effects of Skill Shifts over the Wage Distribution.** Table 5.2 and Figure 1 show the shift in skills across the wage distribution[22]. The increase in skills implies a wage increase in the 10% range towards the top of the distribution, while the increases are only about 5% below the 10th percentile. Between the 20th and 85th percentile there is a large region with gains of about 7-9%. The very top 10% of the distribution gains more. The increase at the top of the distribution and the smaller increase at the bottom imply a widening in the skill distribution. This will, all else equal, result in increased economic inequality over the next decades.

Figure 1 also shows the gains across the log wage distribution for various specifications[23]. One can see that the changing racial composition of the work-force and the associated changes in the other characteristics that it implies generates only a small, fairly uniform, predicted decline in our skill metric. Adding parental education and presence indicators (model 4.2) implies a shift in the log wage of about .05 over most of its distribution. When all of the other components (model 4.6) are added, the shift is typically between 7 and 9 percent. The results with only AFQT and HGC (model 4.5) are similar to the results with AFQT, HGC, and

[22] We present results from the 5th to the 95th percentiles. Results from the tails are consistent with our findings here, but noisy. Figure A-1 in Appendix A displays the observed and counterfactual pdf of log wages. The counterfactual is generated using the full skill specification (model 4.6). Note that the bulge on the left tail comes from the fact the wages are bound below at $3 in 2003 values. The regression adjustment for experience and year 'spreads' the censored values a little rather than leaving them as a mass point.

[23] In fact we show results for the 5-95th percentiles only. The tail behavior is noisy and uninformative.

the school to work transition variables (model 4.6). Thus, we find a increase of about 8 percent in skills between the 1979 and 1997 cohorts, about 5 percent of which is associated with shifts in parent background We discuss the contributions of the various skill components in more detail in section 5.4.

5.3. **Race and Gender Gaps.** Overall we find a modest widening in the skill distribution for the recent generation. At the same time we find gains in skill endowments for various groups that were significantly disadvantaged in the 1979 cohort.

Figure 2, panels 1 and 2 presents the expected changes in log wage distribution conditional on race and gender. The counterfactual distribution is obtained by reweighing to match the changing distributions of all our skill measures (parental education, parental presence, schooling, AFQT, and work transition). The y-axis within each panel is forced to be the same scale and thus results across race can easily be compared. Across gender the scale differs, however. 1.65 standard error bands (bootstrapped) are provided.

For males we find that Blacks and Hispanics gain significantly relative to their White counterparts over almost all of the wage distribution. For Hispanic males the point estimates imply truly dramatic increases in skills in the upper tail of the wage distribution. The top of the Hispanic wage distribution is projected to increase by more than 20%. The shift in characteristics implies a reduction in the mean log wage gap between white and black men from 0.38 to 0.27 and a reduction at the 90th percentiles for the two groups from 0.46 to 0.24. The corresponding reductions in the gap between white and Hispanic men are from 0.21 to 0.09 at the mean. At the 90th percentile the projected difference reverses from 0.23 to a (statistically insignificant) wage advantage of only 0.03.

We find significant widening in skills within both the Black and Hispanic male population. And, as already described, Blacks and Hispanic males skills overall are increasing relative to Whites. Based on these findings we can thus expect a significant proportion of the Black and especially Hispanic population to enter the middle class.

Figure 2.2 suggests that the wage gains of females in wages are likely to exceed gains among males. Again, Hispanics show the most dramatic gains. About two-third of Hispanic females are expected to gain around 20%. The gains for Black females are more varied but exceed 10% over most of the distribution. Gains for white women are small near the bottom of the distribution but exceed ten percent in the top half and exceed twenty percent above the 80th percentile. The results imply that changes in skill components will lead the average gap in the wages of men and women to fall from 0.33 to 0.24 log points. The male/female gap in the 10th, 50th, and 90th quantiles will decline by 0.01, 0.10, and 0.17 respectively.

Overall we find that if the conditional distribution of wages for the 1997 cohort turns out to be similar to that of the 1979 cohort, then an increase in the skill endowments of Blacks and Hispanics relative to Whites as well as of women relative to men will tend to contribute to a decline in economic inequality across groups as the 1997 cohort enters its prime. It will also lead to much more diversity in the economic conditions within minority groups.

6. Identifying the Contribution of Subsets of Variables to Differences between the 1979 and 1997 Wage Distributions

In this section we look in more detail at the importance of the different skill components. First, we present decompositions based on sequential applications of (4.7). Then we use multivariate regression with additive separability restrictions imposed.

Table 6.1 breaks up the difference between the actual wage distribution for 1979 and the counterfactual distribution based on the sequential addition of variables to skill component vector Z. All calculations in the table are based upon Sample 1.

The first column shows the actual distribution of NLSY79 wages and is the same as Table 5.2, column 1. The second column shows the effect of the shift in the race and gender on the mean and various percentiles of the wage distribution. As we have already noted, these shifts have a small negative effect.. One should be keep in mind that the overall effect in column 2 is the sum of the direct effect of race and sex on wages and an indirect effect due to an association between race and sex and other characteristics, such as parental education and AFQT. The next column reports the marginal effect of adding parental education and presence indicators. (The combined effect of race and sex and the parent variables may be calculated by summing columns 2 and 3.) The parent variables are quite important. They imply any increase in the mean wage of about .05. Column 4 reports the marginal effect of adding AFQT. The shift in the AFQT distribution implies only a small additional increase of 0.01 across the entire distribution. In column 5 we add HGC. Adding schooling has a fairly sizable effect of 0.025 log points at the mean and 0.02, 0.035, and 0.01 at the 10th, 50th, and 90th quantiles respectively. In column 6 we add the school to work transition variables. These variables have a relatively small marginal effect on the counterfactual wage distribution.

Tables 6.1A and 6.1B report separate break downs of the marginal effects for each race/sex group. The results show much larger effects of parental background for Black and Hispanics males than for white males.

As we discussed in Section 2.2, the marginal effects of particular variables are not independent of the order in which they are introduced. There is a reasonable case for introducing race and sex followed by parental background before adding AFQT or education outcomes. But performance on the AFQT test and school outcomes are jointly determined, so it is far from obvious that causal priority should be given to AFQT. In column (7) and (8) of table 6.1 we switch the order in which we introduce AFQT and HGC. Reversing the order however does not change the observation that the change in schooling has a relatively large marginal effect on the wage distribution while adding the AFQT has a small marginal effects. Indeed, from Table 6.1A, column 8 we can see that the marginal effect of the AFQT once schooling has been accounted is negative at the mean, although the decline is statistically insignificant. For Black males and Black females the improvement in AFQT seems to be as or more important than the increase in HGC.

Figure 3 provide a different take on the shifts in various skill indicators. Each datapoint in the figure refers to individuals in a percentile of the log wage distribution in 1979. The vertical axis displays weight of these individuals in the sample after reweighing the 1979 data to match the 1997 distribution. We smooth the information in figure 3 using a non-parameteric kernel regression. Here, we observe that matching the 1979 cohort to the 1997 distribution of parental education and parental presence means increasing the weights for those in the top half of the distribution at the expense of those in the bottom part. Accounting for schooling and AFQT scores leads to a further increase in the weigh on NLSY79 cohort members who had characteristics that place them in the upper range of the wage distribution.

6.1. **Regression decompositions.** In column 1 of Table 6.2, we report the OLS regression of the log wage on race, sex, father's education, mother's education, HGC, AFQT, and the school to work transition dummies. For ease of interpretation, we enter the education variables and AFQT in linear form. The second column reports the difference between the 1997 and 1979 cohorts in the means of each of the characteristics. The third column reports the implied effect of shifts on wages, with sum of the groups of characteristics.

Overall, the counterfactual analysis based on OLS regression implies a mean log wage increase of .059, which is somewhat smaller than the estimate of .079 that we obtain using the DFL approach. Presumably, the discrepancy is due to the functional form restrictions placed on the wage equation, but we need to be investigate it more carefully in the next draft of the paper. The decomposition is interesting. The shift in the race and sex contribution implies an increase in log wages of .005. Note that these results reflect only the direct effect of the shift and holding the other variables constant, including parental education, HGC, and AFQT. In contrast, the small negative effect of race and sex on the mean reported in Column 2 of Table 6.1 combines the direct effect of the shifts in race and gender with the indirect effect on the other variables[24].The partial effect of the increase in parental education is .022 (.003). On the other hand, the shift away from 2 parent families implies a reduction of .009 (0.004). This reflects the coefficients of -0.02 on "neither mother nor father" and -.04 on "mother only". The total effect of the shift in parental background is .013. Again, these estimates, in contrast to those in Tables 5.2 and 6.1, hold constant the effects of HGC, AFQT, and the school to work transition. Therefore, they do not incorporate the pathways of the change in family structure on these variables.

HGC has a coefficient of .038. The increase in HGC of 0.708 implies a shift in mean log wages of 0.027. The increase of 3.151 in AFQT implies a wage increase of 0.016. Thus, as in table 6.1 we find that schooling tends to have a larger effect on wages than the AFQT. The combined effect of changes in HGC, AFQT, and the school to work transition variables is 0.041. Comparing the results from the regression decomposition with those from the DFL type decomposition we find that individual characteristics are seen to matter relatively more

[24]One may use regression to estimate the "total effect" of the shift in race and gender by excluding all other variables from the wage regression and using the resulting coefficient estimates to the weight change in race and gender. The resulting estimate is 0.13%.

in the regression decomposition. This is due to the fact that parental schooling is correlated with individual skill measures such as schooling and the AFQT.

Panel A of Table 6.3 provides regression decompositions by race for males. The first row shows that the shift in characteristics will lead to wage growth of .05 for whites, .113 for blacks, and .144 for Hispanics. The gains for blacks come primarily from the shifted of 9.234 in the AFQT score, which implies a wage shift of .076. Education contributes an additional .030. For Hispanics, AFQT, HGC and the school to work transition variables all makes significant contributions, but a large increase in parental education also plays an important independent role. The results for females in panel B also show much larger average increases for Blacks and Hispanics, with HGC and AFQT contributing .039 (.006) and .122 (.0097) for Black women and .062 (.0088) and .069 (.0086) for Hispanic women. The direct effects of parental background are small. The results in Table 6.3 panel B indicate that parental background plays an important indirect role for both groups.

7. Conclusion

Changes in the level and distribution of skill will play an important role in determining economic growth as well in driving changes in the level and distribution of wages and employment. In this paper we examine changes the characteristics of American youth between the late 1970s and the late 1990s, with a focus on what matters for labor market success. Drawing on the work of DFL, we examine the employment rates and wages of the NLSY79 sample after reweighting it to look like the NLSY97 along a number of dimensions that are related to labor market success, including race, gender, parental background, education, test scores, and variables that capture whether individuals transition smoothy from school to work. We also use more standard regression methods to assess the labor market consequences of differences between the two cohorts. Overall, our results suggest that the current generation is more skilled than the previous one. If the wage process faced by the NLSY79 cohort in their prime age years persists, our findings imply that the white/black and white/Hispanics gaps in wages will fall. We also find a reduced gap between men and women.

We wish to stress that our empirical findings are still somewhat preliminary. First, more needs to be done to assess the issue of whether the NLSY97 base year sample is non-representative. Second, while we believe that our corrections for attrition and for bias from missing data on test scores are adequate, it will be possible to improve upon them by using a larger set of covariates from the base year sample. Third, our results for NLSY79 and NLSY97 need to be supplemented with information from other sources, including the NAEP and the CPS. A cursory look at the NAEP scores for the relevant years indicate that they move in the same direction as the AFQT scores, but we need ot asses this more carefully. We are currently extending our analysis to other outcomes, notably incarceration and fertility. We will also explore the role of immigration in shaping the labor market potential of the cohort represented by NLSY97.

In future work, we hope to implement two extensions to our methods. The first involves using vectors, Z_1 and Z_2 say, of variables for which the joint distribution is available in NLSY79 but only the marginal distributions of Z_1 and Z_2 are observed for NLY97. The second involves using variables that measure the same concepts but are based on different questions in the two data sets.

8. References

Bishop, John H., "Is the Test Score Decline Responsible for the Productivity Growth Decline?", *The American Economic Review*, Vol. 79, No. 1. (March): pp. 178-197.

Blau, Francine D. And Andrea H. Beller. 1992. "Black-White Earnings Over the 1970s and 1980s: Gender Differences in Trends." *Review of Economics and Statistics*. Vol 74:2, pp276-86.

Belley, Philippe and Lance Lochner, (2007) “The Changing Role of Family Income and Ability in Determining Educational Achievement”, University of Western Ontario.

Card, D. and Krueger, A. 1996. School resources and student outcomes: an overview of the literature and new evidence from North and South Carolina. *Journal of Economic Perspectives* 10(4), 31-50.

Chinloy, Peter, (1980)"Sources of Quality Change in Labor Input," *American Economic Review*, 70, March, 108-19.

Denison, E.F, (1974), Accounting for United States economic growth, 1929-1969, Brookings Institution, Washington.

DiNardo J, Fortin NM, Lemieux T (1996) "Labor Market Institutions and the Distribution of Wages, 1973-1992: A Semiparametric Approach", *Econometrica* 64(5): 1001-46

Eide, Eric, and Jeff Grogger. 1995. "Changes in College Skills and the Rise in the College Wage Premium". *Journal of Human Resources* v30 n2 Spring pp 280-310.

Firpo, Sergio, Nicole Fortin and Thomas Lemieux, "Unconditional Quantile Regressions", unpublished paper, University of British Columbia (2007)

Jorgensen, Dale, Frank Gollop, and Barbara Fraumeni, (1987), Productivity and U.S. Economic Growth. Cambridge, MA: Harvard University Press.

Korenman, S., & Winship, C. (2000). "A reanalysis of The Bell Curve: Intelligence, family background, and schooling." In K. Arrow & S. Bowles & S. Durlauf (Eds.), Meritocracy and Economic Inequality. Princeton: Princeton University Press.

MaCurdy, Thomas and Edward Vytlacil (2003) "Establishing New Norms for the AFQT Using Data from PAY97", Working Paper.

W. Moore, S. Pedlow, P. Krishnamurty, and K. Wolter (2000), "National Longitudinal Survey of Youth 1997 (NLSY97) Technical Sampling Report"

NCES (National Center for Education Statistics). 2005. Long-term trend major results. July 2005. Available at http://nces.ed.gov/nationsreportcard/ltt/results2004/, accessed September 2007.

Neal, D. 2006. Why has black-white skill convergence stopped? In The Handbook of Economics of Education, vol. 1, ed. E. Hanushek and F. Welch. Amsterdam: North-Holland.

Segall, D. O. (1997). Equating the CAT-ASVAB. In W. A. Sands, B. K. Waters, & J. R. McBride (Eds.), Computerized adaptive testing: From inquiry to operation (pp. 181-198). Washington, DC: American Psychological Association.

Smith, J. and Welch, F. 1989. Black economic progress after Myrdal. *Journal of Economic Literature* 27, 519-64.

Appendix A: Variables used and their construction

- Observations (1997): the entire data set was used. NLSY 1997 has a supplemental sample, consisting of 2236 people. Unlike NLSY79, there is no breakdown of the supplemental sample by race. In total there are 8984 observations in the data.
- Observations (1979): We use the entire general sample as well as parts of the supplemental sample. From the supplemental sample we only keep (male and female) Hispanics and Blacks. Hence we have 9761 observations in the dataset - 6111 from the normal cross section and 3650 from the supplemental sample.
- Weights (1997 & 1979): We use weights as they are generated in 1997 and 1979. There is the option of creating one set of weights for all years combined, but we do not choose that because creating weights in that manner places a weight of 0 for people that later attrite out of the data. The weights are implied 2 decimal places, so we divide the weights by 100 before using them.
- Work after graduation (1979 & 1997): We construct this variable in the following manner. We examine a person when he/she is 22 or 23 years of age at the time of the interview (hence some people have 'gaps' in their ages) and note their highest grade completed. If they had achieved the same highest grade completed by the age of 20 or less, we consider them in the universe of people who could have worked after "graduation". To be coded as having worked, they need to have reported more than or equal to 14 weeks of work in either of the first 2 years after graduation. Respondents who work less than 14 weeks in either of the 2 years and who could have worked after graduation are coded as not working after graduation.
- Timing of graduation (1979 & 1997): Again, the universe we consider are the people whose highest grade completed at age 22 or 23 is the same as the highest grade completed by age 20 or below. Among these, we consider those people to have graduated "on time" if the age at which they graduate is exactly equal to the highest grade completed by June plus six. Graduation is assumed to occur in June of given year. A person is thought to have graduated early if age is less than highest grade completed as of June plus six. Late graduation occurs when age at graduation is more than highest grade completed in June plus six.
- AFQT scores (1997): The NLSY 1997 does not have AFQT score as designated by the DOD. Instead they have a self created variable that mimics what the DOD does to various parts of the CAT-ASVAB. The 1997 cohort was administered the test in 1997-98. The test was computerized unlike the 1980 test which was a paper and pencil test. For comparability, Dan Segall created a comparison between the paper and pencil test and the computerized test. We use AFQT scores as computed under his methodology so as to be able to compare AFQT scores across the 2 samples. AFQT scores (1979): These are used directly; however, they are age standardized in the same way as the 1997 scores

- Age standardizing AFQT scores (1979 & 1997): the AFQT scores are not standardized by age. Age could play a large role in determining the score one gets on the test and hence we need the scores to reflect this. We age standardize the AFQT scores in the following manner. We take 16 year olds and compute percentiles for their AFQT scores. We can now associate a percentile with an AFQT score. We then compute percentiles for every other age group. Now we assume that the percentile a person is placed in at (say) age 14 is the same as the percentile the person would have been placed in at age 16. We then associate the percentile with the AFQT score. Basically we create a function applying percentiles to AFQT scores using 16 year olds. For every other age group, we compute percentiles within that age group and then apply the above mentioned function to get their AFQT scores. We used 16 year olds as the base group because that is the age group with the most overlap in the 2 s! amples. Recall that NLSY 79 interviewed people between the ages of 14-22, while the NLSY 97 interview people between the ages of 12-17.
- Parental Presence at Age 14 (1997): To construct the number of biological parents a respondent lived with at age 14, we use the question "Respondents relationship to household parental figure". This question was asked every year, hence for respondents that were actually of age 14 at the time of being asked this question, we have direct answers as to how many biological parents they lived with. For people that were already 15 by the time this question was asked in 1997, we impute parental presence. In 1997, all respondents were asked their relationship to the household parental figure at age 12. Hence this is retrospective information for some people. We use this information and assume that the respondent's living situation did not change between ages 12 and 14 to impute parental presence at age 14 for people 15 years and older by 1997.
  - Mother only: this variable implies the presence of the mother at age 14 and not the presence of the father. Mother along with some other man (not father) is accounted under "mother only" for example.
  - Father only: this is defined in the same way as mother only. If father is present and mother not present, then the person is listed as having lived with father only at age 14.
  - Mother and father: this implies the presence of both parents
  - Neither mother nor father: this means neither father nor mother were present. If the respondent lived with grandparents at age 14 then he would be considered as having lived with neither father nor mother.
- Parental Presence at Age 14 (1979): This question was asked directly to the 1979 sample. While the answers varied from 1997, we are still able to accurately compute number of biological parents present in the respondent's house at age 14. Mother only, Father only etc breakdown is constructed similar to the 1997 variables.

- Race (1997): The variable we use is the one created by the NLSY combining race and ethnicity. There are self reported measures of race, but we do not use those. Using this variable we have 2335 Black, 1901 Hispanic, 83 Mixed Race and 4665 Non Black/Non-Hispanic persons in the dataset.
- Race (1979): In 1979 we use the screener definition of race. The variable was created by the NLSY, and according to the codebook by Dennis Gray in 1980. Out of 9761 persons, we have 4916 Non-Black/Non-Hispanic, 2921 Black, and 1924 Hispanic people.
- Highest grade completed by parents: this is available individually for mother and father. In both cohorts, we use the same strategy to identify father and mother's highest grade completed. If the screener question in 1979 and 1997 was left as missing for these variables, we fill those by using the demographic roster information collected each year. Hence, the final numbers are a combination of the answer to the screener question in the first survey year for each cohort, and if that is left blank, it has been filled in using roster information from each year.
- Wage: the actual wage variable used for the 1979 cohort is the hourly wage variable. This variable denotes the hourly wage in cents and has been CPI adjusted for 2003. Moreover we set the floor at 300 cents and the ceiling at 20000 cents. Hence if a person reports a wage, but is less than 300 then that person gets assigned an hourly wage of 300. The same applied for someone reporting a wage over 20000 cents an hour.
  – Residualized wages: for the 1979 cohort we compute experience and education adjusted wages in the following manner: regress the log of hourly wage on a cubic of experience (defined as age minus highest grade completed at age 22 minus 6) by education group. Education groups are less than 12 years of education, exactly 12 years of education and more than 12 years of education. From this regression we compute the residual and add in the constant experience of 23 years.

Appendix B:Comparability of AFQT-scores across NLSY79 and NLSY97

Two major problems arise in making the AFQT-scores comparable across the NLSY79 and NLSY97 cohort. First, the ASVAB changed from a paper and pencil (P&P) format in 1980 to a computer adminstered (CAT) format in 1997. Second, NLSY79 sample members were between 15 and 23 years old when they took the test. Test takers in the NLSY97 were between 12 and 18 years olds and thus typically were younger than their NLSY79 counterparts.

To make the AFQT scores comparable we perform two "equipercentile" procedures. The first method is based on the work of Daniel Segall (1997), who matches test scores of individuals across percentiles based on a study of individuals who were randomly adminstered either the P&P or the CAT. As noted above, Segall kindly provided us with the results of mapping within age P&P (1979) scores for the NLSY79 sample into equivalent CAT (1997) scores The second equipercentile procedure adjusts for the variation in age at test taking. For this purpose we use the overlap between the age ranges of NLSY79 and NLSY97 test takers. The most overlap exists for age 16 with 1329 respondents in 1997 taking the test at age 16 and 1324 respondents in 1980 taking the test at age 16. For each sample, we perform an equipercentile mapping to age 16 of the scores of respondents who took the test age other ages. Specifically, in the case of the NLSY79 sample, persons who took the test at age a who scored in the q'th percentile among age a test takers were assigned the qth percentile value for NLSY79 sample members who took the test at age 16. A corresponding set of assignments were made for the NLSY97 sample. This procedure assumes that the relative rankings of individuals in the AFQT-distribution on average do not depend on when they took the test. It also assumes that the level of cognitive skills in adulthood associated with the qth percentile in the age 16 test taker distribution is the same as that for the qth percentile in the age a distribution.

Table B.1 provides evidence that the joint distribution of observables and the AFQT score is indeed similar across ages in both surveys. We estimate regressions of the standardized AFQT-scores on interaction of the birth years with various observables used in the anlysis. If the joint distribution of observables and percentile score conditional on age varies across age, then we would expect that interacting age (or equivalently birth-year) with observables would predict the age standardized AFQT scores. Table B.1 shows the F-statistic for excluding various sets of interactions between observables and birth years for various specifications and both the NLSY79 and NLSY97. Table B.1 strongly suggests that there is no variation in the relationship between observables and the AFQT score across cohorts within either the NLSY79 and NLSY97.

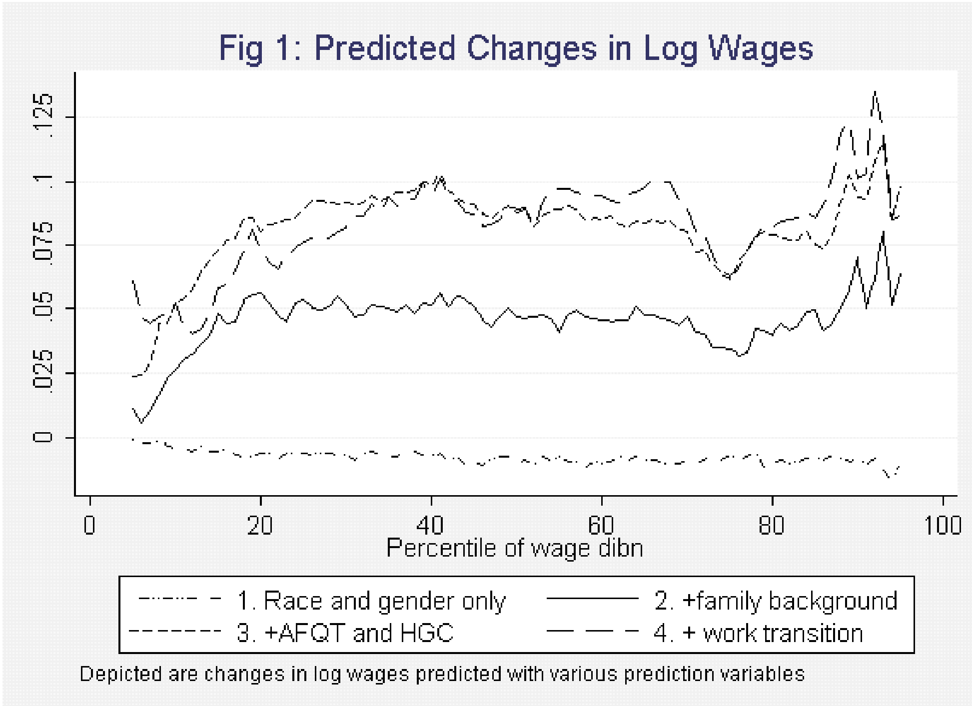


Fig 1: Predicted Changes in Log Wages
0
.025
.05
.075
.1
.125
0
20
40
60
80
100
Percentile of wage dibn
1. Race and gender only
2. +family background
3. +AFQT and HGC
4. + work transition
Depicted are changes in log wages predicted with various prediction variables

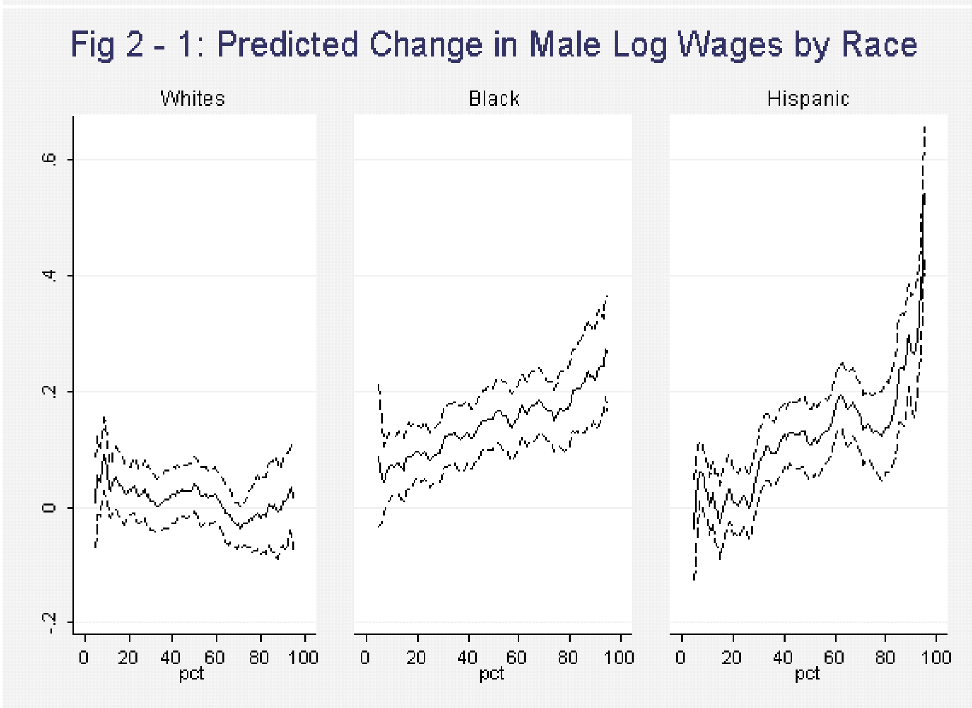


Fig 2 - 1: Predicted Change in Male Log Wages by Race
Whites
Black
Hispanic
-.2
0
.2
.4
.6
0 20 40 60 80 100
pct
0 20 40 60 80 100
pct
0 20 40 60 80 100
pct

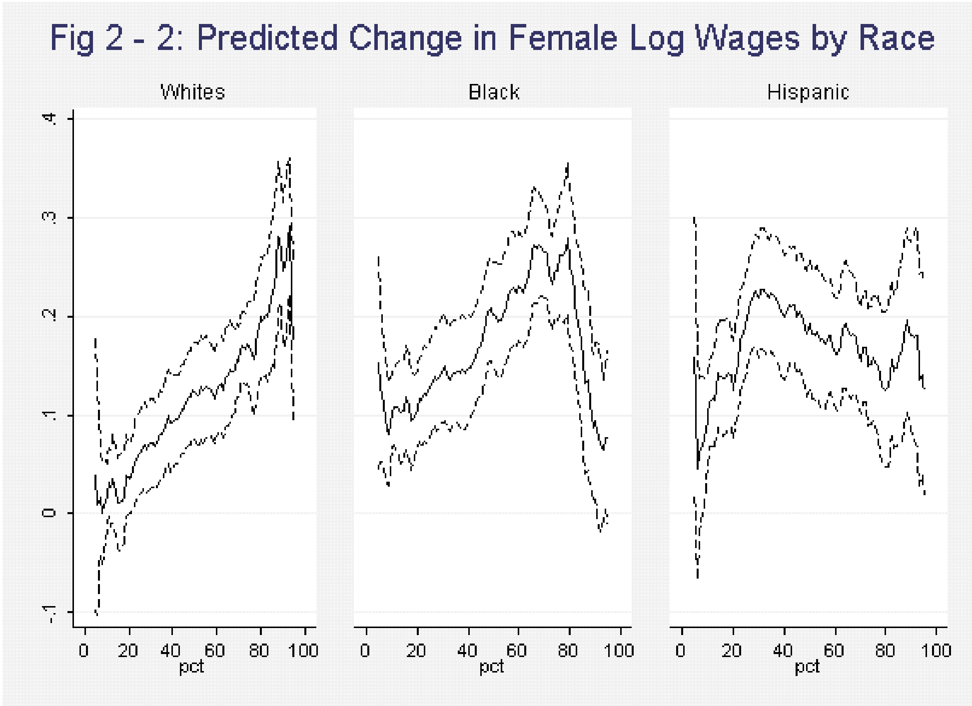


Fig 2 - 2: Predicted Change in Female Log Wages by Race
Whites
Black
Hispanic
-.1
0
.1
.2
.3
.4
0 20 40 60 80 100
pct

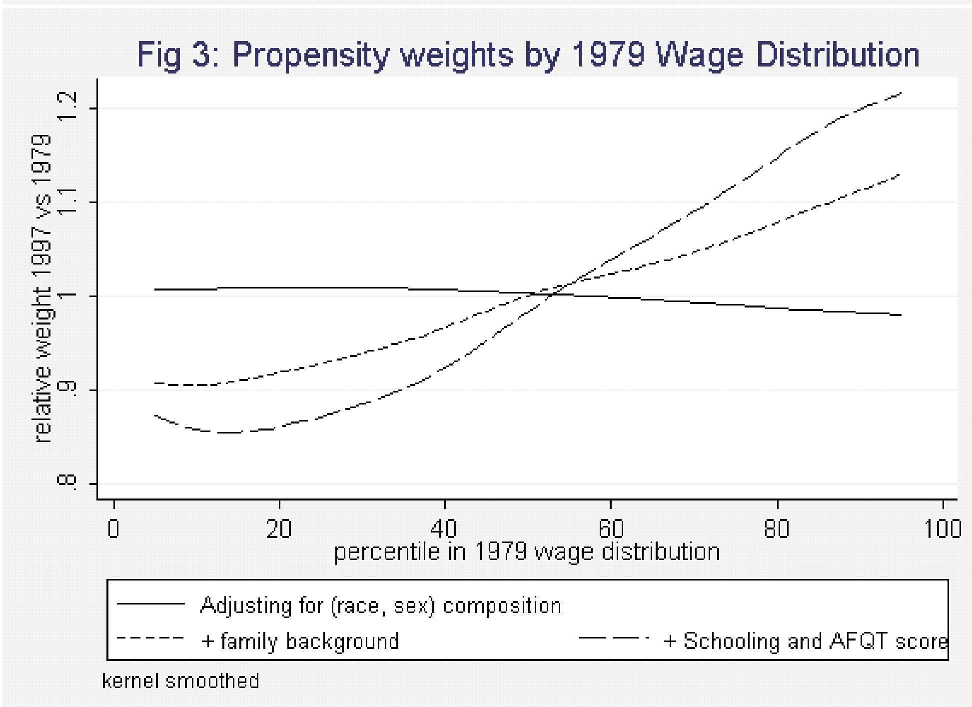
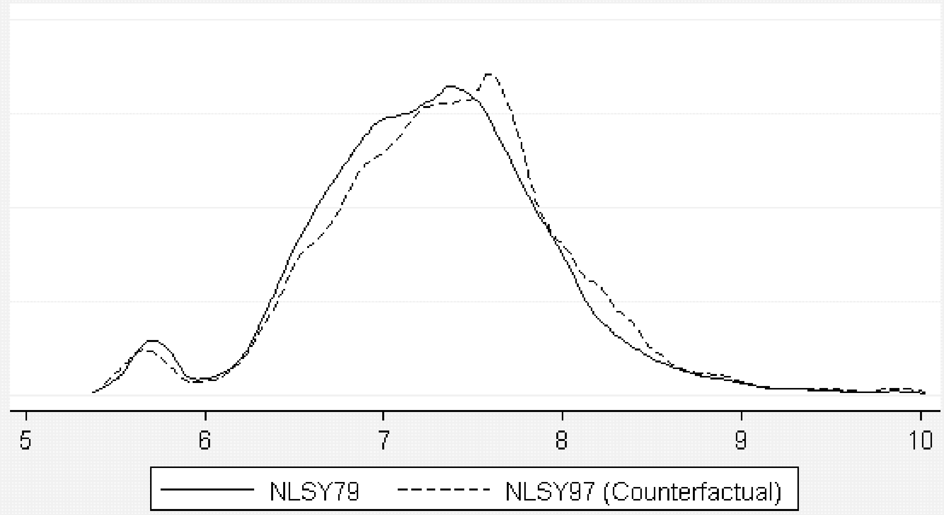


Fig 3: Propensity weights by 1979 Wage Distribution
relative weight 1997 vs 1979
.8
.9
1
1.1
1.2
0 20 40 60 80 100
percentile in 1979 wage distribution
Adjusting for (race, sex) composition
+ family background
+ Schooling and AFQT score
kernel smoothed

Fig A-1: Observed and Counterfactual Log Wage Density for NLSY79 and NLSY97



Counterfactual based on full prediction set as in specification 4.6, Sample 1
Density of log wages 'while working'. Log wages regression adjusted to year=2002, experience=23
Densities smoothed with Epanechnikov kernel (bw=0.075)

Table 3.0 SAMPLE ATTRITION		
Reason for exclusion	**NLSY 1979**	**NLSY 1997**
No excluded cases	12,686	8,984
Excluded oversampled White male and female	9,761	8,984
Ought to be present at age 22	**9,761**	**5,406**
	100.00%	***100.00%***
Not present at age 22	9,326	4,704
	95.54%	*87.01%*
Highest grade completed missing	9,297	4,654
	95.25%	*86.09%*
AFQT missing	8,903	3,775
	91.21%	*69.83%*

Notes: Ought to be present at age 22 is calculated using birth year information of respondents. In the 1979 cohort we expect to observe everyone at age 22. in the 1997 cohort, since the last year of interview is 2004, we only expect people born on or before 1982 to reach the age of 22 in the data. AFQT here means age standardized AFQT. Note that a small number of cases in both cohorts are lost due to a death prior to age 22.

Table 3.1 Characteristics by Attrition Status at Age 22										
	NLSY 1979					*NLSY 1997*				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	N	Pooled	Attriters	Stayers	Attriters-Stayers	N	Pooled	Attriters	Stayers	Attriters-Stayers
Race										
White	4,916	78.91%	80.10%	78.85%	1.25 (1.37)	2,783	70.43%	74.21%	69.71%	4.5 (1.16)***
Black	2,921	14.47%	11.59%	14.60%	-2.87 (1.10)**	1,431	15.33%	13.67%	15.65%	-1.98 (0.91)**
Hispanic	1,924	6.62%	8.31%	6.54%	1.77 (0.94)*	1,141	13.01%	11.13%	13.36%	-2.23 (0.84)**
Other						51	1.23%	0.99%	1.28%	-0.29 (0.27)
Sample										
Cross-Sectional Sample	6,111	84.62%	83.89%	84.65%	-0.76 (1.26)	4,028	87%	89.97%	86.44%	3.53 (0.81)***
Supplemental Sample	3,650	15.38%	16.11%	15.35%	0.76 (1.26)	1,378	13%	10.03%	13.56%	-3.53 (0.81)***
Parental Years of Schooling										
Father										
Years completed (average)	8,293	11.9	12	11.89	0.11 (0.48)	4,590	13	12.9	13.03	-0.12 (0.95)
Missing	1468	10.23%	13.60%	10.07%	3.53 (1.17)***	816	11.25%	12.64%	10.99%	1.65 (0.87)*
Mother										
Years completed (average)	9,121	11.64	11.58	11.64	-0.06 (0.37)	5,198	12.9	12.64	12.94	-0.30 (2.78)***
Missing	640	5.31%	7.64%	5.20%	2.44 (0.91)**	208	3.19%	3.20%	3.19%	0.01 (0.46)
Parental Presence at age 14										
Mother only	2,385	18.68%	18.55%	18.69%	-0.14 (1.33)	2,077	34.65%	28.44%	35.83%	-7.39 (1.2)***
Father only	281	3.01%	3.95%	2.96%	0.99 (0.66)	236	4.74%	5.24%	4.65%	0.59 (0.58)
Mother and Father	6,617	75.07%	73.66%	75.14%	-1.48 (1.51)	2,181	44.89%	45.39%	44.79%	0.6 (1.31)
Neither Mother nor Father	478	3.24%	3.84%	3.21%	0.63 (0.66)	912	15.72%	20.93%	14.73%	6.2 (1.05)***
Total	9,761		4.55%	95.45%		5406		15.80%	84.72%	

Reported statistics are generated by attrition status at age 22 and weighted using the the base year sample weights for NLSY79 and NLSY97 respectively. For each statistic the difference between the attriter and stayers is reported with standard errors. Difference statistically significant at the .01 level (***), .05 level (**) or .10 level (*). Std errors reported in parenthesis.

Table 3.2: Skill indicators/early outcomes by AFQT Missing status *Sample: persons observed at age 22*										
	NLSY 1979					NLSY 1997				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	N	Pooled	AFQT Missing	AFQT Not Missing	Missing - Not Missing	N	Pooled	AFQT Missing	AFQT Not Missing	Missing - Not Missing
Race										
White	4,690	78.95%	78.31%	78.98%	-0.67 (0.012)	2,373	70.19%	60.73%	72.17%	-11.44 (1.25)***
Black	2,818	14.47%	11.59%	14.60%	-3.01 (0.95)***	1,264	15.33%	19.04%	14.55%	4.49 (1.00)***
Hispanic	1,818	6.58%	10.10%	6.42%	3.68 (0.88)***	1,022	13.21%	18.93%	12.01%	6.93 (0.99)***
Other	NA					45	1.27%	1.30%	1.27%	0.03 (0.29)
Sample										
Cross Sectional Sample	5,847	84.73%	84.01%	84.76%	-0.75 (1.08)	3,471	86.64%	81.26%	87.76%	-6.5 (0.98)***
Supplemental Sample	3,479	15.27%	15.99%	15.24%	0.75 (1.08)	1,233	13.36%	18.74%	12.24%	6.5 (0.98)***
Highest grade completed at age 22										
Years completed (average)	9,297	12.61	11.73	12.64	-0.91 (0.12)***	4,654	13.16	12.6	13.29	-0.68 (0.08)***
Missing	29	0.30%	0.90%	0.27%	0.63 (0.27)**	50	0.69%	0.84%	0.66%	0.18 (0.23)
Parental Years of Schooling										
Father										
Years completed (average)	7,935	11.88	11.29	11.91	-0.62 (0.23)***	4,003	13.04	12.42	13.15	-0.73 (0.15)***
Missing	1,391	10.17%	13.50%	10.02%	3.48 (1.00)***	701	11.23%	17.17%	9.99%	7.18 (0.95)***
Mother										
Years completed (average)	8,721	11.62	11.04	11.64	-0.60 (0.17)***	4,521	12.92	12.23	13.06	-0.83 (0.12)***
Missing	605	5.27%	8.75%	5.12%	3.63 (0.82)***	183	3.15%	3.38%	3.10%	0.28 (0.46)
Parental presence at age 14										
Mother only	2,275	18.71%	16.07%	18.82%	-2.75 (1.09)**	1,853	35.10%	35.99%	34.92%	1.07 (1.25)
Father only	266	2.99%	4.33%	2.93%	1.4 (0.60)**	200	4.56%	6.10%	4.23%	1.87 (0.61)***
Mother and Father	6,330	75.06%	74.66%	75.09%	-0.43 (1.28)	1,913	44.61%	30.78%	47.51%	-16.73 (1.22)***
Neither Mother nor Father	455	3.24%	4.95%	3.16%	1.79 (0.63)**	738	15.73%	27.13%	13.34%	13.79 (1.11)***
Total	**9,326**		**6.56%**	**93.44%**		**4,704**		**19.41%**	**80.59%**	

Reported statistics are generated for groups defined by observed AFQT test score and weighted using the sample weights provided by the NLSY and the attrition weights generated by the authors to account for attrition by age 22 in both years. For each statistic the difference between the attriter and stayers is reported. *** denote rejecting the difference to be 0 at a 1%, ** at a 5% significance level, *** at a 1% significance level. Std error reported in parenthesis.

Table 3.3 Summary Statistics

Panel A: Statistics for Full Sample

Year	1979	Std Deviation	1997	Std Deviation
AFQT	42.02	27.01	45.28	27.26
HGC at age 22	12.61	2.06	13.17	1.94
Father's HGC	11.9	3.59	12.99	3.19
Mother's HGC	11.64	2.78	12.95	2.92
Enrolled at age 22	20.35%		30.61%	
Mother only	18.68%		34.82%	
Father only	3.01%		5.21%	
Mother and Father	75.07%		48.57%	
Neither Mother nor Father	3.24%		11.40%	
Work after leave school	82.96%		82.23%	

Panel B: Statistics by race

Year	1979			1997		
Race	White	Black	Hispanic	White	Black	Hispanic
AFQT	47.75	18.52	25.28	50.88	26.51	33.78
HGC at age 22	12.79	12.13	11.51	13.41	12.59	12.58
Father's HGC	12.37	10.37	8.46	13.45	12.28	10.81
Mother's HGC	12.05	10.93	8.09	13.4	12.55	10.84
Enrolled at age 22	21.73%	15.1%	15.39%	33.02%	24.24%	24.9%
Mother only	14.5%	38.07%	26.1%	30.39%	54.18%	36.14%
Father only	3.13%	2.66%	2.38%	5.52%	4.38%	4.18%
Mother and Father	80.37%	50.43%	65.7%	54.68%	22.37%	47.23%
Neither Mother nor Father	1.99%	8.83%	5.83%	9.4%	19.06%	12.46%
Work after leave school	86.57%	68.6%	79.3%	84.54%	75.5%	80.93%

Table 3.3 Summary Statistics (contd)

Panel C: Stats by race and sex

Year	1979					
Race	White		Black		Hispanic	
Sex	M	F	M	F	M	F
AFQT	48.22	47.28	18.06	18.96	26.83	23.75
HGC at age 22	12.72	12.87	11.91	12.35	11.52	11.52
Father's HGC	12.45	12.3	10.46	10.29	8.58	8.33
Mother's HGC	12.09	11.99	11.04	10.83	8.19	7.99
Enrolled at age 22	24.36%	19.07%	13.61%	16.53%	15.38%	15.39%
Mother only	13.74%	15.28%	37.26%	38.86%	27.49%	24.71%
Father only	3.72%	2.54%	2.99%	2.34%	2.76%	2%
Mother and Father	80.62%	80.12%	51.39%	49.51%	65.6%	66.09%
Neither Mother nor Father	1.92%	2.07%	8.36%	9.29%	4.44%	7.2%
Work after leave school	84.5%	88.66%	72.79%	64.35%	84.65%	73.51%
Year	1997					
Race	White		Black		Hispanic	
Sex	M	F	M	F	M	F
AFQT	50.68	51.1	24.7	28.3	33.71	33.86
HGC at age 22	13.17	13.65	12.32	12.86	12.46	12.73
Father's HGC	13.48	13.42	12.16	12.1	10.8	10.81
Mother's HGC	13.41	13.39	12.62	12.47	10.89	10.77
Enrolled at age 22	30.13%	35.97%	19.03%	29.58%	23.36%	26.71%
Mother only	28.43%	32.44%	54.09%	54.29%	34.75%	37.74%
Father only	5.49%	5.55%	5.05%	3.69%	4.64%	3.64%
Mother and Father	55.99%	53.32%	21.82%	22.95%	48.23%	46.07%
Neither Mother nor Father	10.09%	8.69%	19.04%	19.07%	12.38%	12.55%
Work after leave school	84.06%	85.14%	75.71%	75.27%	77.39%	85.93%

Table 3.4: Education at Age 22 by cohort, race and sex						
Year	1979					
Race	White		Black		Hispanic	
Sex	M	F	M	F	M	F
GED	8.83%	8.55%	15.28%	10.46%	14.55%	12.63%
HS Diploma	80.1%	83.72%	66.9%	74.89%	60.62%	64.3%
HGC>=14 at age 22	33.2%	34.81%	18.74%	25.62%	18.22%	18.44%
Enrolled at age 22	24.36%	19.07%	13.61%	16.53%	15.38%	15.39%
Year	1997					
Race	White		Black		Hispanic	
Sex	M	F	M	F	M	F
GED	7.16%	4.96%	10.77%	7.28%	6.25%	6.29%
HS Diploma	82.9%	86.25%	66.89%	77.88%	74.75%	77.12%
HGC>=14 at age 22	41.28%	53.23%	23.24%	34.4%	24.41%	32.51%
Enrolled at age 22	30.13%	35.97%	19.03%	29.58%	23.36%	26.71%

Table 5.1 Comparison of Actual Employment Rates of 1979 Cohort with Counterfactual Rates based on characteristics of 1997 cohort.[1]						
	Observed LFP in NLSY 79		Counterfactual minus observed LFP-Rates[2]			
			Model 4.6[3]	Model 4.4[3]		Model 4.5[3]
Percentile	Sample 1	Sample 2	Sample 1	Sample 1	Sample 2	Sample 1
Male Total	0.918	0.916	0.006	0.000	0.007	0.001
	(0.0036)	(0.0037)	(0.0096)	(0.0070)	(0.0072)	(0.0085)
White	0.93	0.929	-0.005	-0.003	0.000	-0.006
	(0.0044)	(0.0045)	(0.0128)	(0.0092)	(0.0093)	(0.0110)
Black	0.853	0.848	0.036	0.006	0.021	0.018
	(0.0086)	(0.0082)	(0.0113)***	(0.0119)	(0.0114)*	(0.0121)
Hispanic	0.911	0.901	0.038	0.027	0.043	0.032
	(0.0094)	(0.0090)	(0.0102)***	(0.0083)**	(0.0088)***	(0.0090)***
Female Total	0.836	0.834	-0.003	-0.006	-0.001	0.001
	(0.0056)	(0.0057)	(0.0114)	(0.0111)	(0.0103)	(0.0090)
White	0.838	0.836	-0.011	-0.007	-0.003	-0.008
	(0.0068)	(0.0068)	(0.0130)	(0.0091)	(0.0094)	(0.0098)
Black	0.831	0.829	0.009	-0.015	-0.012	0.023
	(0.0086)	(0.0085)	(0.0331)	(0.0472)	(0.0430)	(0.0281)
Hispanic	0.821	0.817	0.044	0.021	0.038	0.035
	(0.0128)	(0.0129)	(0.0149)***	(0.0135)	(0.0130)***	(0.0135)***
Total	0.876	0.874	0.003	-0.001	0.004	0.002

1) Employment Rate is measured by reference to a valid wage observation. An individual is coded to have a valid wage observation is the required average hourly rate of pay lies between $3 and $200 (in 2003 real values) in a given year. Reported percentages refer to shares with valid wages in years with positive responses between 1998-2004. Sample 1 includes respondents with observed AFQT scores. Sample 2 includes those with missing AFQT scores. All statistics are weighted by the cross-sectional weights. Specifications estimated on sample 1 are in addition weighted to account for both attrition by age 22 and AFQT-non response. Specifications estimated on Sample 2 are in addition weighted to account for attrition by age 22. Standard errors are bootstrapped with 200 repetitions. Bootstrap stratified on NLSY cohort, race and gender. Units are sampled at the individual level. * refers to significance at 10%, ** at 5%, and *** at 1 % level.

2) Measured against corresponding sample reported in columns 1 and 2.

3) All specifications match on race and gender. Model 4.4 (eq 4.4) refers to the specification matching on schooling, parental education and family structure. Model 4.5 (eq. 4.5) matches schooling, parental education, family structre, and the AFQT-scores. Model 4.6 refers to the full specification (eq 4.6) matching on schooling, AFQT scores, parental education, family structure and the school-work transition variables.

Table 5.2 Comparison of Actual Wages of 1979 Cohort with Counterfactual Wage distributions based on characteristics of 1997 cohort.[1]						
	Observed Wage distribution in NLSY 1979		Counterfactual minus Actual Wages[2]			
			Model 4.6[3]	Model 4.4[3]		Model 4.5[3]
Percentile	Sample 1	Sample 2	Sample 1	Sample 1	Sample 2	Sample 1
5%	6.232 (0.0287)	6.231 (0.0282)	0.06 (0.0363)*	0.03 (0.0395)	0.033 (0.0376)	0.037 (0.0389)
10%	6.490 (0.0105)	6.487 (0.0103)	0.05 (0.0195)***	0.051 (0.0190)***	0.053 (0.0181)***	0.05 (0.0190)***
25%	6.842 (0.0086)	6.839 (0.0085)	0.075 (0.0236)***	0.094 (0.0157)***	0.087 (0.0139)***	0.085 (0.0175)***
50%	7.267 (0.0090)	7.263 (0.0091)	0.087 (0.0272)***	0.092 (0.0159)***	0.082 (0.0170)***	0.094 (0.0197)***
75%	7.664 (0.0085)	7.661 (0.0080)	0.061 (0.0261)***	0.066 (0.0214)***	0.055 (0.0181)***	0.055 (0.0214)***
90%	8.040 (0.0152)	8.034 (0.0141)	0.097 (0.0394)***	0.09 (0.0250)***	0.096 (0.0227)***	0.090 (0.0279)***
95%	8.328 (0.0237)	8.322 (0.0226)	0.083 (0.0429)*	0.109 (0.0332)***	0.085 (0.0328)***	0.092 (0.0323)***
Mean	7.264	7.259	0.079	0.102	0.075	0.79

1) Sample 1 includes only respondents with observed AFQT scores. Sample 2 includes those with missing AFQT scores. Reported wage distributions are conditional on reporting positive wages. Wages are regression standardized to year=2002 and experience=23. Wages are inflation adjusted to 1990 using the CPI-U. All statistics are weighted by the cross-sectional weights. Specifications estimated on sample 1 are in addition weighted to account for attrition by age 22 and AFQT-non response. Sample 2 are in addition weighted to account for attrition by age 22. Standard errors: bootstrapped with 200 repetitions. Bootstrap stratified on NLSY cohort, race and gender. Units are sampled at the individual level. * refers to significance at 10%, ** at 5%, and *** at 1 % level.

2) Measured against corresponding sample reported in columns 1 and 2.

3) All Specifications match on race and gender. Model 4.4 (eq 4.4) refers to the specification matching on schooling, parental education and family structure. Model 4.5 (eq. 4.5) matches schooling, parental education, family structre, and the AFQT-scores. Model 4.6 refers to the full specification (eq 4.6) matching on schooling, AFQT scores, parental education, family structure and the school-work transition variables.

Table 6.1A Identifying the Contribution of Subsets of Variables to Differences between the 1979 and 1997 Wage Distributions								
Percentile	1979 Log Wage Distribution	Marginal Effects of Additional Variables						
		(1) + Race, Sex	(2) + Family Background	(3) + AFQT	(4) + Highest Grade	(5) + Work Transition	(3) + Highest Grade	(7) + AFQT
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
5%	6.232 (0.0287)	0.002 (0.0057)	0.009 (0.0497)	0.008 (0.0118)	0.018 (0.0454)	0.023 (0.0243)	0.019 (0.0198)	0.007 (0.0469)
10%	6.49 (0.0105)	-0.003 (0.0027)	0.026 (0.0217)	0.008 (0.0055)	0.019 (0.0203)	0.000 (0.0135)	0.028 (0.0097)	-0.001 (0.0207)
25%	6.842 (0.0086)	-0.003 (0.0025)	0.058 (0.0163)***	0.004 (0.0062)	0.026 (0.0183)	-0.010 (0.0158)	0.039 (0.0096)	-0.013 (0.0178)
50%	7.267 (0.0090)	-0.007 (0.0027)***	0.056 (0.0171)***	0.009 (0.0064)	0.036 (0.0209)	-0.007 (0.0160)	0.043 (0.0082)	0.002 (0.0209)
75%	7.664 (0.0085)	-0.007 (0.023)***	0.043 (0.0165)***	0.006 (0.0060)	0.013 (0.0239)	0.006 (0.0129)	0.030 (0.0119)**	-0.011 (0.0167)
90%	8.04 (0.0152)	-0.009 (0.0035)***	0.080 (0.0284)***	0.008 (0.0091)	0.011 (0.0335)	0.007 (0.0224)	0.019 (0.0129)	0.000 (0.0308)
95%	8.328 (0.0237)	-0.008 (0.0054)	0.072 (0.0423)*	0.009 (0.0144)	0.019 (0.0427)	-0.009 (0.0335)	0.045 (0.0162)***	-0.017 (0.0418)
Mean	7.264	-0.008	0.073	0.010	0.020	-0.010	0.046	-0.018

1. Estimated on respondents with non-missing AFQT scores (Sample 1). Reported wage distributions are conditional on reporting positive wages. Wages are regression standardized to year=2002 and experience=23. Wages are inflation adjusted to 1990 using the CPI-U. Standard errors: bootstrapped with 200 repetitions. Bootstrap stratified on NLSY cohort, race and gender. Units are sampled at the individual level. All statistics are weighted by NLSY cross-sectional weights adjusted for attrition by age 22 and non-response to the AFQT variable.

2. Columns 2-8 show the incremental contribution of relevant variables in the title of each column.

Table 6.1B: Identifying the Contribution of Subsets of Variables to Differences between the 1979 and 1997 Wage Distributions by Race and Sex

Percentile	1979 Log Wage Distribution	Marginal Effects of Additional Variables					
		(1) + Family Background	(3) + AFQT	(4) + Highest Grade	(5) + Work Transition	(3) + Highest Grade	(7) + AFQT
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
PANEL A: Males							
White Male							
5%	6.465	0.033	-0.004	0.000	-0.026	0.010	-0.014
	(0.0287)	(0.0426)	(0.0099)	(0.0225)	(0.0207)	(0.0192)	(0.0118)
10%	6.745	0.064	0.000	0.002	-0.002	0.013	-0.011
	(0.0105)	(0.0191)***	(0.0049)	(0.0112)	(0.0135)	(0.0093)	(0.0061)*
25%	7.115	0.016	-0.015	0.039	-0.024	0.038	-0.014
	(0.0086)	(0.0135)	(0.0056)***	(0.0105)***	(0.0144)*	(0.0090)***	(0.0079)*
50%	7.487	0.008	-0.004	0.032	-0.002	0.033	-0.005
	(0.0090)	(0.0148)	(0.0057)	(0.0106)***	(0.0153)	(0.0076)***	(0.0099)
75%	7.863	0.012	-0.013	0.015	-0.040	0.023	-0.021
	(0.0085)	(0.0144)	(0.0059)**	(0.0130)	(0.0121)***	(0.0116)**	(0.0076)***
90%	8.250	0.036	-0.004	0.011	-0.040	0.025	-0.018
	(0.0152)	(0.0251)	(0.0084)	(0.0150)	(0.0209)*	(0.0127)*	(0.0099)*
95%	8.558	0.035	0.000	0.031	-0.058	0.045	-0.014
	(0.0237)	(0.0308)	(0.0135)	(0.0129)**	(0.0288)**	(0.0150)***	(0.0126)
Mean	7.494	0.02	-0.006	0.018	-0.017	0.026	-0.014
Black Male							
5%	6.201	-0.045	0.015	0.013	0.076	0.028	0.000
	(0.0287)	(0.0426)	(0.0099)	(0.0225)	(0.0207)***	(0.0192)	(0.0118)
10%	6.436	0.005	0.026	0.012	0.014	0.028	0.010
	(0.0105)	(0.0191)	(0.0049)***	(0.0112)	(0.0135)	(0.0093)***	(0.0061)*
25%	6.733	0.060	0.014	0.001	0.007	0.012	0.003
	(0.0086)	(0.0135)***	(0.0056)**	(0.0105)	(0.0144)	(0.0090)	(0.0079)
50%	7.084	0.098	0.034	0.013	0.008	0.025	0.022
	(0.0090)	(0.0148)***	(0.0057)***	(0.0106)	(0.0153)	(0.0076)***	(0.0099)**
75%	7.474	0.116	0.042	0.006	-0.009	0.011	0.037
	(0.0085)	(0.0144)***	(0.0059)***	(0.0130)	(0.0121)	(0.0116)	(0.0076)***
90%	7.794	0.165	0.054	-0.014	0.014	0.000	0.040
	(0.0152)	(0.0251)***	(0.0084)***	(0.0150)	(0.0209)	(0.0127)	(0.0099)***
95%	8.064	0.203	0.053	-0.029	0.037	-0.003	0.027
	(0.0237)	(0.0308)***	(0.0135)***	(0.0129)**	(0.0288)	(0.0150)	(0.0126)**
Mean	7.107	0.084	0.034	0.006	0.013	0.017	0.023
Hispanic Male							
5%	6.350	0.004	0.002	0.008	-0.052	0.016	-0.006
	(0.0287)	(0.0426)	(0.0099)	(0.0225)	(0.0207)**	(0.0192)	(0.0118)
10%	6.540	0.070	0.005	0.000	-0.058	0.011	-0.006
	(0.0105)	(0.0191)***	(0.0049)	(0.0112)	(0.0135)***	(0.0093)	(0.0061)
25%	6.919	0.062	0.012	0.010	-0.086	0.048	-0.026
	(0.0086)	(0.0135)***	(0.0056)**	(0.0105)	(0.0144)***	(0.0090)***	(0.0079)***
50%	7.287	0.092	0.000	0.000	0.007	0.018	-0.018
	(0.0090)	(0.0148)***	(0.0057)	(0.0106)	(0.0153)	(0.0076)**	(0.0099)**
75%	7.666	0.086	0.012	-0.012	0.031	0.019	-0.019
	(0.0085)	(0.0144)***	(0.0059)**	(0.0130)	(0.0121)**	(0.0116)	(0.0076)**
90%	8.021	0.222	0.000	0.012	0.027	0.015	-0.003
	(0.0152)	(0.0251)***	(0.0084)	(0.0150)	(0.0209)	(0.0127)	(0.0099)
95%	8.285	0.206	0.000	0.100	0.236	0.100	0.000
	(0.0237)	(0.0308)***	(0.0135)	(0.0129)***	(0.0288)***	(0.0150)***	(0.0126)
Mean	7.29	0.107	0.009	0.009	0.014	0.031	-0.013

Table 6.1B (contd): Identifying the Contribution of Subsets of Variables to Differences between the 1979 and 1997 Wage Distributions by Race and Sex

Percentile	1979 Log Wage Distribution	Marginal Effects of Additional Variables: (1) + Family Background	(3) + AFQT	(4) + Highest Grade	(5) + Work Transition	(3) + Highest Grade	(7) + AFQT
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
PANEL B: Females							
White Female							
5%	5.869 (0.0287)	-0.032 (0.0426)	-0.004 (0.0099)	-0.029 (0.0225)	0.144 (0.0207)***	-0.062 (0.0192)***	0.029 (0.0118)**
10%	6.407 (0.0105)	0.013 (0.0191)	-0.003 (0.0049)	0.024 (0.0112)**	-0.019 (0.0135)	0.021 (0.0093)**	0.000 (0.0061)
25%	6.731 (0.0086)	0.051 (0.0135)***	0.001 (0.0056)	0.023 (0.0105)**	-0.018 (0.0144)	0.033 (0.0090)***	-0.009 (0.0079)
50%	7.115 (0.0090)	0.076 (0.0148)***	0.006 (0.0057)	0.046 (0.0106)***	-0.009 (0.0153)	0.058 (0.0076)***	-0.006 (0.0099)
75%	7.529 (0.0085)	0.093 (0.0144)***	0.006 (0.0059)	0.055 (0.0130)***	0.014 (0.0121)	0.057 (0.0116)***	0.004 (0.0076)
90%	7.885 (0.0152)	0.104 (0.0251)***	0.000 (0.0084)	0.141 (0.0150)***	0.000 (0.0209)	0.118 (0.0127)***	0.023 (0.0099)**
95%	8.138 (0.0237)	0.165 (0.0308)***	0.000 (0.0135)	0.000 (0.0129)	0.005 (0.0288)	0.000 (0.0150)	0.000 (0.0126)
Mean	7.127	0.062	0.003	0.04	0.003	0.045	-0.002
Black Female							
5%	6.094 (0.0287)	-0.019 (0.0426)	0.061 (0.0099)***	0.076 (0.0225)***	0.033 (0.0207)	0.043 (0.0192)**	0.094 (0.0118)***
10%	6.359 (0.0105)	0.026 (0.0191)	0.021 (0.0049)***	0.047 (0.0112)***	0.006 (0.0135)	0.029 (0.0093)***	0.039 (0.0061)***
25%	6.600 (0.0086)	0.056 (0.0135)***	0.005 (0.0056)	0.054 (0.0105)***	0.006 (0.0144)	0.033 (0.0090)***	0.026 (0.0079)***
50%	6.938 (0.0090)	0.056 (0.0148)***	0.044 (0.0057)***	0.095 (0.0106)***	-0.002 (0.0153)	0.058 (0.0076)***	0.081 (0.0099)***
75%	7.316 (0.0085)	0.104 (0.0144)***	0.085 (0.0059)***	0.069 (0.0130)***	-0.014 (0.0121)	0.044 (0.0116)***	0.110 (0.0076)***
90%	7.664 (0.0152)	0.057 (0.0251)**	0.043 (0.0084)***	-0.017 (0.0150)	0.008 (0.0209)	0.038 (0.0127)***	-0.012 (0.0099)
95%	7.854 (0.0237)	0.025 (0.0308)	0.065 (0.0135)***	-0.031 (0.0129)**	0.016 (0.0288)	0.027 (0.0150)*	0.007 (0.0126)
Mean	6.966	0.055	0.048	0.053	0	0.043	0.058
Hispanic Female							
5%	6.088 (0.0287)	-0.060 (0.0426)	0.024 (0.0099)**	0.069 (0.0225)***	0.105 (0.0207)***	0.094 (0.0192)***	-0.001 (0.0118)
10%	6.389 (0.0105)	-0.022 (0.0191)	0.012 (0.0049)**	0.041 (0.0112)***	0.060 (0.0135)***	0.051 (0.0093)***	0.002 (0.0061)
25%	6.681 (0.0086)	0.034 (0.0135)**	0.037 (0.0056)***	0.072 (0.0105)***	0.042 (0.0144)***	0.109 (0.0090)***	0.000 (0.0079)
50%	7.065 (0.0090)	0.092 (0.0148)***	0.048 (0.0057)***	0.029 (0.0106)***	0.022 (0.0153)	0.064 (0.0076)***	0.013 (0.0099)
75%	7.463 (0.0085)	0.071 (0.0144)***	0.075 (0.0059)***	0.010 (0.0130)	0.001 (0.0121)	0.064 (0.0116)***	0.021 (0.0076)**
90%	7.814 (0.0152)	0.141 (0.0251)***	0.041 (0.0084)***	0.010 (0.0150)	-0.015 (0.0209)	0.024 (0.0127)**	0.027 (0.0099)***
95%	8.058 (0.0237)	0.110 (0.0308)***	0.021 (0.0135)	0.000 (0.0129)	-0.006 (0.0288)	0.015 (0.0150)	0.006 (0.0126)
Mean	7.076	0.068	0.045	0.027	0.019	0.063	0.009

Table 6.2 Regression Decompositions[1]	OLS Regression[2]	Difference in mean characteristics (1997-1979)	Implied effect of shift in wages
	(1)	(2)	(3)
Overall Change			**0.059 (0.0063)*****
Highest Grade Completed	0.038 (0.0036)***	0.708	0.027 (0.0026)***
AFQT	0.005 (0.0003)***	3.151	0.016 (0.0009)***
Parental Years of Schooling			
Mother			
Dummy for missing	0.063 (0.0364)*		
Years of schooling	0.009 (0.0027)***		
Father		0.708	0.022 (0.0034)***
Dummy for missing	0.118 (0.0277)***		
Years of schooling	0.009 (0.0022)***		
Parental presence at age 14			
Mother only	-0.043 (0.0132)***	0.154	
Father only	0.008 (0.0293)	0.012	-0.009 (0.0043)***
Neither mother nor Father	-0.021 (0.0257)	0.129	
Work Transition			
Work after graduation	0.119 (0.0204)***	-0.002	
Graduate early	-0.135 (0.0235)***	0.104	-0.002 (0.0025)
Graduate on time	-0.150 (0.0215)***	-0.165	
Graduate late	-0.206 (0.0248)***	0.061	
Race and Sex dummies			
Black Male	-0.163 (0.0160)***	0.001	
Hispanic Male	0.000 (0.0178)	0.041	
White Female	-0.367 (0.0127)***	-0.023	0.005 (0.0009)***
Black Female	-0.324 (0.0151)***	-0.004	
Hispanic Female	-0.195 (0.0177)***	0.0223	
Constant	6.563 (0.0446)***		

1) The sample excludes respondents without valid AFQT scores and attriters by age 22. The excluded category in the regression specification are white males, with both mother and father present at age 193.37, N = 25300.

Table 6.3 Regression Decompositions by Race [1]

PANEL A: Males	White			Black			Hispanic		
	OLS Regression	Difference in mean characteristics (1997-1979)	Implied effect of shift in wages	OLS Regression	Difference in mean characteristics (1997-1979)	Implied effect of shift in wages	OLS Regression	Difference in mean characteristics (1997-1979)	Implied effect of shift in wages
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Overall Change			**0.051 (0.0139)*****			**0.113 (0.0159)*****			**0.144 (0.0185)*****
Highest Grade Completed	0.0455 (0.006)***	0.61	0.028 (0.0039)***	0.0584 (0.008)***	0.506	0.030 (0.0039)***	0.0213 (0.009)***	0.976	0.021 (0.0092)***
AFQT	0.0049 (0.000)***	2.464	0.012 (0.0011)***	0.0082 (0.001)***	9.234	0.076 (0.0068)***	0.0055 (0.001)***	6.167	0.034 (0.0063)***
Parental Years of Schooling									
Mother									
Dummy for missing	0.0785 (0.072)	-0.020		0.0962 (0.082)	-0.048		0.1191 (0.070)*	-0.078	
Years of schooling	0.0121 (0.005)***	1.608		0.0041 (0.006)	2.013		0.0219 (0.006)***	2.956	
Father			0.031 (0.0064)***			0.010 (0.0093)***			0.057 (0.0133)***
Dummy for missing	0.1795 (0.057)***	0.029		0.0004 (0.059)	-0.031		0.0093 (0.060)	-0.059	
Years of schooling	0.0128 (0.004)***	0.647		0.0036 (0.005)	1.871		0.0015 (0.005)	1.726	
Parental presence at age 14									
Mother only	-0.0706 (0.024)***	0.134		-0.0421 (0.024)*	0.139		-0.0331 (0.035)	0.066	
Father only	-0.0319 (0.050)	0.003	-0.014 (0.0097)	-0.1269 (0.074)*	0.021	-0.014 (0.0097)	0.1172 (0.108)	0.025	0.001 (0.0084)
Neither mother nor Father	-0.0288 (0.058)	0.151		-0.0335 (0.051)	0.155		0.0024 (0.060)	0.115	
Work Transition									
Work after graduation	0.1417 (0.041)***	0.039		0.0688 (0.037)*	-0.022		0.1566 (0.072)**	0.042	
Graduate early	-0.2287 (0.047)***	0.203	-0.007 (0.0076)	0.0092 (0.054)	0.169	0.010 (0.0080)	-0.1023 (0.081)	0.179	0.031 (0.0100)***
Graduate on time	-0.1709 (0.044)***	-0.208		-0.0598 (0.038)	-0.157		-0.3056 (0.079)***	-0.027	
Graduate late	-0.2747 (0.046)***	0.006		-0.0822 (0.040)**	-0.012		-0.2267 (0.068)***	-0.152	
Constant	6.4127 (0.079)***			6.1937 (0.104)***			6.7355 (0.104)***		
R-sq	0.166			0.1636			0.137		
Observations	6615			3589			2311		

Table 6.3 (contd) Regression Decompositions by Race[1]									
PANEL B: Females									
	White			**Black**			**Hispanic**		
	OLS Regression	Difference in mean characteristics (1997-1979)	Implied effect of shift in wages	OLS Regression	Difference in mean characteristics (1997-1979)	Implied effect of shift in wages	OLS Regression	Difference in mean characteristics (1997-1979)	Implied effect of shift in wages
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Overall Change			**0.044 (0.0155)*** **			**0.165 (0.0120)*** **			**0.161 (0.0132)*** **
Highest Grade Completed	0.0237 (0.007)***	0.883	0.021 (0.0059)***	0.0449 (0.007)***	0.858	0.039 (0.0058)***	0.0590 (0.008)***	1.048	0.062 (0.0088)***
AFQT	0.0044 (0.000)***	2.8	0.012 (0.0013)**	0.0096 (0.001)***	12.668	0.122 (0.0097)***	0.0074 (0.001)***	9.383	0.069 (0.0086)***
Parental Years of Schooling									
Mother									
Dummy for missing	0.0673 (0.073)	-0.004		-0.0084 (0.060)	-0.022		0.0536 (0.082)	-0.016	
Years of schooling	0.0067 (0.005)	1.493		0.0043 (0.005)	1.711		0.0075 (0.007)	1.968	
Father			0.024 (0.0075)***			0.009 (0.0068)			-0.001 (0.0111)
Dummy for missing	0.2027 (0.058)***	-0.013		0.0065 (0.047)	0.047		-0.1438 (0.060)**	0.031	
Years of schooling	0.0111 (0.004)***	1.488		0.0026 (0.004)	0.605		-0.0109 (0.006)*	0.972	
Parental presence at age 14									
Mother only	-0.0576 (0.026)**	0.184		-0.0055 (0.021)	0.138		0.1293 (0.032)***	0.116	
Father only	0.0678 (0.043)	0.021	-0.008 (0.0088)	0.0191 (0.054)	0.004	-0.007 (0.0071)	0.1483 (0.099)	0.001	0.011 (0.0066)*
Neither mother nor Father	0.0114 (0.067)	0.096		-0.0360 (0.030)	0.185		-0.0423 (0.055)	0.091	
Work Transition									
Work after graduation	0.1133 (0.045)**	-0.064		0.1387 (0.027)***	0.01		0.0905 (0.040)**	0.11	
Graduate early	-0.1264 (0.047)**	-0.018	-0.005 (0.0099)	-0.0936 (0.036)***	0.124	0.002 (0.0051)	-0.0353 (0.052)	0.104	0.020 (0.0060)***
Graduate on time	-0.1442 (0.046)	-0.163		-0.1447 (0.028)***	-0.194		-0.1293 (0.042)***	-0.071	
Graduate late	-0.1283 (0.065)**	0.181		-0.2194 (0.032)***	0.071		-0.1381 (0.047)***	-0.034	
Constant	6.4112 (0.078)***			6.1802 (0.083)***			6.2276 (0.091)***		
R-sq	0.076			0.194			0.174		
Observations	6472			3947			2366		

1) The sample excludes respondents without valid AFQT scores and attriters by age 22. The excluded category in the regression specification are with both mother and father present at age 14 and who did not graduate by age 20. Observations are weighted using the cross-section weights provided by the NLSY adjusted to account for attrition by age 22 and AFQT non-response. Standard errors in parenthesis.
*** significant at 1%, ** significant at 5%, * significant 10%

Appendix B Table 1: Testing Age Standardization of AFQT Scores						
	NLSY 1979			NLSY 1997		
Specification 1	F-stat	Degrees of Freedom	P value	F-stat	Degrees of Freedom	P value
Cohort	0.62	8, 8885	0.77	0.64	4, 3728	0.64
Cohort X HGC	0.71	8, 8885	0.69	0.58	4, 3728	0.68
All combined	0.76	16, 8885	0.73	0.84	8, 3728	0.56
Specification 2						
Cohort	0.52	8, 8867	0.85	0.72	4, 3718	0.58
Cohort X HGC	0.6	8, 8867	0.78	0.54	4, 3718	0.71
Cohort X Race	0.96	16, 8867	0.5	1.26	8, 3718	0.26
All combined	1.09	32, 8867	0.33	1.05	16, 3718	0.4
Specification 3						
Cohort	0.93	8, 7287	0.49	0.21	4, 3170	0.93
Cohort X HGC	0.99	8, 7287	0.44	0.21	4, 3170	0.93
Cohort X Race	0.95	16, 7287	0.51	0.37	8, 3170	0.93
Cohort X Father's HGC	0.73	16, 7287	0.76	0.5	8, 3170	0.86
All combined	0.88	48, 7287	0.71	0.62	24, 3170	0.93

Notes: F tests are for the null that all coefficients are equal to zero.

Specification 1: regression of standardized afqt on cohort dummies, cohort dummies interacted with hgc, and hgc

Specification 2: regression of standardized afqt on cohort dummies, cohort dummies interacted with hgc, cohort dummies interacted with race, and hgc

Specification 3: regression of standardized afqt on cohort dummies, cohort dummies interacted with hgc, cohort dummies interacted with race, cohort dummies interacted with father's hgc, hgc and mother's hgc

Cohort: dummy for each cohort

Cohort X HGC: interaction of highest grade completed and cohort dummy

Cohort X Race: interaction of race and cohort dummy

Cohort X Father's HGC: interaction of Father's highest grade completed and cohort dummy

All combined: Testing the restriction that all coeffcicients in the regression are equal to zero